Graduate Texts in Mathematics

Béla Bollobás

Modern Graph Theory

Springer

Graduate Texts in Mathematics 184

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces. 2nd ed.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 J.-P. SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 J.-P. SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.
61 WHITEHEAD. Elements of Homotopy Theory.
62 KARGAPOLOV/MERLZJAKOV. Fundamentals of the Theory of Groups.
63 BOLLOBAS. Graph Theory.

(continued after index)

Béla Bollobás

Modern Graph Theory

With 118 Figures

Springer

Béla Bollobás
Department of Mathematical Sciences
University of Memphis
3725 Norriswood
Memphis, TN 38152
USA
bollobas@msci.memphis.edu

Trinity College
University of Cambridge
Cambridge CB2 1TQ
United Kingdom
b.bollobas@dpmms.cam.ac.uk

Mathematics Subject Classification (2000): 05-01, 05Cxx

Library of Congress Cataloging-in-Publication Data
Bollobás, Béla.
Modern graph theory / Béla Bollobás.
p. cm. — (Graduate texts in mathematics ; 184)
Includes bibliographical references (p. -) and index.
ISBN 978-0-387-98488-9 ISBN 978-1-4612-0619-4 (eBook)
DOI 10.1007/978-1-4612-0619-4
acid-free paper)
1. Graph Theory. I. Title. II. Series.
QA166.B663 1998
511′.5—dc21 98-11960

ISBN 978-0-387-98488-9 Printed on acid-free paper.

9 8 7 6 5

springeronline.com

To Gabriella

As long as a branch of science offers an abundance of problems, so long is it alive; a lack of problems foreshadows extinction or the cessation of independent development. Just as any human undertaking pursues certain objects, so also mathematical research requires its problems. It is by the solution of problems that the investigator tests the temper of his steel; he finds new methods and new outlooks, and gains a wider and freer horizon.

David Hilbert, *Mathematical Problems*,
International Congress of Mathematicians,
Paris, 1900.

Apologia

This book has grown out of *Graph Theory – An Introductory Course* (GT), a book I wrote about twenty years ago. Although I am still happy to recommend GT for a fairly fast-paced introduction to the basic results of graph theory, in the light of the developments in the past twenty years it seemed desirable to write a more substantial introduction to graph theory, rather than just a slightly changed new edition.

In addition to the classical results of the subject from GT, amounting to about 40% of the material, this book contains many beautiful recent results, and also explores some of the exciting connections with other branches of mathematics that have come to the fore over the last two decades. Among the new results we discuss in detail are: Szemerédi's Regularity Lemma and its use, Shelah's extension of the Hales-Jewett Theorem, the results of Galvin and Thomassen on list colourings, the Perfect Graph Theorem of Lovász and Fulkerson, and the precise description of the phase transition in the random graph process, extending the classical theorems of Erdős and Rényi. One whole field that has been brought into the light in recent years concerns the interplay between electrical networks, random walks on graphs, and the rapid mixing of Markov chains. Another important connection we present is between the Tutte polynomial of a graph, the partition functions of theoretical physics, and the powerful new knot polynomials.

The deepening and broadening of the subject indicated by all the developments mentioned above is evidence that graph theory has reached a point where it should be treated on a par with all the well-established disciplines of pure mathematics. The time has surely now arrived when a rigorous and challenging course on the subject should be taught in every mathematics department. Another reason why graph theory demands prominence in a mathematics curriculum is its status as that branch of pure mathematics which is closest to computer science. This proximity enriches both disciplines: not only is graph theory fundamental to theoretical computer science, but problems arising in computer science and other areas of application greatly influence the direction taken by graph theory. In this book we shall not stress applications: our treatment of graph theory will be as an exciting branch of pure mathematics, full of elegant and innovative ideas.

Graph theory, more than any other branch of mathematics, feeds on problems. There are a great many significant open problems which arise naturally in the subject: many of these are simple to state and look innocent but are proving to be surprisingly hard to resolve. It is no coincidence that Paul Erdős, the greatest problem-poser the world has ever seen, devoted much of his time to graph theory. This amazing wealth of open problems is mostly a blessing, but also, to some extent, a curse. A blessing, because there is a constant flow of exciting problems stimulating the development of the subject: a curse, because people can be misled into working on shallow or dead-end problems which, while bearing a superficial resemblence to important problems, do not really advance the subject.

In contrast to most traditional branches of mathematics, for a thorough grounding in graph theory, absorbing the results and proofs is only half of the battle. It is rare that a genuine problem in graph theory can be solved by simply applying an existing theorem, either from graph theory or from outside. More typically, solving a problem requires a "bare hands" argument together with a known result with a new twist. More often than not, it turns out that none of the existing high-powered machinery of mathematics is of any help to us, and nevertheless a solution emerges. The reader of this book will be exposed to many examples of this phenomenon, both in the proofs presented in the text and in the exercises. Needless to say, in graph theory we are just as happy to have powerful tools at our disposal as in any other branch of mathematics, but our main aim is to solve the substantial problems of the subject, rather than to build machinery for its own sake.

Hopefully, the reader will appreciate the beauty and significance of the major results and their proofs in this book. However, tackling and solving a great many challenging exercises is an equally vital part of the process of becoming a graph theorist. To this end, the book contains an unusually large number of exercises: well over 600 in total. No reader is expected to attempt them all, but in order to really benefit from the book, the reader is strongly advised to think about a fair proportion of them. Although some of the exercises are straightforward, most of them are substantial, and some will stretch even the most able reader.

Outside pure mathematics, problems that arise tend to lack a clear structure and an obvious line of attack. As such, they are akin to many a problem in graph theory: their solution is likely to require ingenuity and original thought. Thus the expertise gained in solving the exercises in this book is likely to pay dividends not only in graph theory and other branches of mathematics, but also in other scientific disciplines.

"As long as a branch of science offers an abundance of problems, so long is it alive", said David Hilbert in his address to the Congress in Paris in 1900. Judged by this criterion, graph theory could hardly be more alive.

B. B.
Memphis
March 15, 1998

Preface

Graph theory is a young but rapidly maturing subject. Even during the quarter of a century that I lectured on it in Cambridge, it changed considerably, and I have found that there is a clear need for a text which introduces the reader not only to the well-established results, but to many of the newer developments as well. It is hoped that this volume will go some way towards satisfying that need.

There is too much here for a single course. However, there are many ways of using the book for a single-semester course: after a little preparation any chapter can be included in the material to be covered. Although strictly speaking there are almost no mathematical prerequisites, the subject matter and the pace of the book demand mathematical maturity from the student.

Each of the ten chapters consists of about five sections, together with a selection of exercises, and some bibliographical notes. In the opening sections of a chapter the material is introduced gently: much of the time results are rather simple, and the proofs are presented in detail. The later sections are more specialized and proceed at a brisker pace: the theorems tend to be deeper and their proofs, which are not always simple, are given rapidly. These sections are for the reader whose interest in the topic has been excited.

We do not attempt to give an exhaustive list of theorems, but hope to show how the results come together to form a cohesive theory. In order to preserve the freshness and elegance of the material, the presentation is not over-pedantic: occasionally the reader is expected to formalize some details of the argument. Throughout the book the reader will discover connections with various other branches of mathematics, like optimization theory, group theory, matrix algebra, probability theory, logic, and knot theory. Although the reader is not expected to have intimate knowledge of these fields, a modest acquaintance with them would enhance the enjoyment of this book.

The bibliographical notes are far from exhaustive: we are careful in our attributions of the major results, but beyond that we do little more than give suggestions for further readings.

A vital feature of the book is that it contains hundreds of exercises. Some are very simple, and test only the understanding of the concepts, but many go way beyond that, demanding mathematical ingenuity. We have shunned routine drills: even in the simplest questions the overriding criterion for inclusion was beauty. An attempt has been made to grade the exercises: those marked by $^{-}$ signs are five-finger exercises, while the ones with $^{+}$ signs need some inventiveness. Solving an exercise marked with $^{++}$ should give the reader a sense of accomplishment. Needless to say, this grading is subjective: a reader who has some problems with a standard exercise may well find a $^{+}$ exercise easy.

The conventions adopted in the book are standard. Thus, Theorem 8 of Chapter IV is referred to as Theorem 8 within the chapter, and as Theorem IV.8 elsewhere. Also, the symbol, $\square$, denotes the end of a proof; we also use it to indicate the absence of one.

The quality of the book would not have been the same without the valuable contributions of a host of people, and I thank them all sincerely. The hundreds of talented and enthusiastic Cambridge students I have lectured and supervised in graph theory; my past research students and others who taught the subject and provided useful feedback; my son, Márk, who typed and retyped the manuscript a number of times. Several of my past research students were also generous enough to give the early manuscript a critical reading: I am particularly grateful to Graham Brightwell, Yoshiharu Kohayakawa, Imre Leader, Oliver Riordan, Amites Sarkar, Alexander Scott and Andrew Thomason for their astute comments and perceptive suggestions. The deficiencies that remain are entirely my fault.

Finally, I would like to thank Springer-Verlag and especially Ina Lindemann, Anne Fossella and Anthony Guardiola for their care and efficiency in producing this book.

B. B.
Memphis
March 15, 1998

For help with preparation of the third printing, I would like to thank Richard Arratia, Peter Magyar, and Oliver Riordan. I am especially grateful to Don Knuth for sending me lists of misprints. For the many that undoubtedly remain, I apologize. Please refer to the website for this book, where I will maintain a list of further misprints that come to my attention; I'd be grateful for any assistance in making this list as complete as possible. The url for this book is http://www.msci.memphis.edu/faculty/bollobasb.html

B. B.
Memphis
April 16, 2002

Contents

Neque ingenium sine disciplina,
aut disciplina sine ingenio
perfectum artificem potest efficere.

Vitruvius

I

Fundamentals

The basic concepts of graph theory are extraordinarily simple and can be used to express problems from many different subjects. The purpose of this chapter is to familiarize the reader with the terminology and notation that we shall use in the book. In order to give the reader practice with the definitions, we prove some simple results as soon as possible. With the exception of those in Section 5, all the proofs in this chapter are straightforward and could have safely been left to the reader. Indeed, the adventurous reader may wish to find his own proofs before reading those we have given, to check that he is on the right track.

The reader is not expected to have complete mastery of this chapter before sampling the rest of the book; indeed, he is encouraged to skip ahead, since most of the terminology is self-explanatory. We should add at this stage that the terminology of graph theory is still not standard, though the one used in this book is well accepted.

I.1 Definitions

A *graph* G is an ordered pair of disjoint sets (V, E) such that E is a subset of the set $V^{(2)}$ of unordered pairs of V. Unless it is explicitly stated otherwise, we consider only finite graphs, that is, V and E are always finite. The set V is the set of *vertices* and E is the set of *edges*. If G is a graph, then $V = V(G)$ is the vertex set of G, and $E = E(G)$ is the edge set. An edge $\{x, y\}$ is said to *join* the vertices x and y and is denoted by xy. Thus xy and yx mean exactly the same edge; the vertices x and y are the *endvertices* of this edge. If $xy \in E(G)$, then x and y are *adjacent*, or *neighbouring*, vertices of G, and the vertices x and y are *incident* with the edge xy. Two edges are *adjacent* if they have exactly one common endvertex.

As the terminology suggests, we do not usually think of a graph as an ordered pair, but as a collection of vertices some of which are joined by edges. It is then a natural step to draw a picture of the graph. In fact, sometimes the easiest way to describe a small graph is to draw it; the graph with vertices $1, 2, \ldots, 9$ and edges 12, 23, 34, 45, 56, 61, 17, 72, 29, 95, 57, 74, 48, 83, 39, 96, 68, and 81 is immediately comprehended by looking at Fig. I.1.

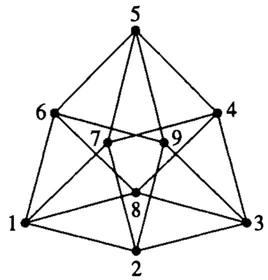


FIGURE I.1. A graph.

We say that $G' = (V', E')$ is a *subgraph* of $G = (V, E)$ if $V' \subset V$ and $E' \subset E$. In this case we write $G' \subset G$. If G' contains *all edges* of G that join two vertices in V' then G' is said to be the subgraph *induced* or *spanned by* V' and is denoted by $G[V']$. Thus, a subgraph G' of G is an induced subgraph if $G' = G[V(G')]$. If $V' = V$, then G' is said to be a *spanning* subgraph of G. These concepts are illustrated in Fig. I.2.

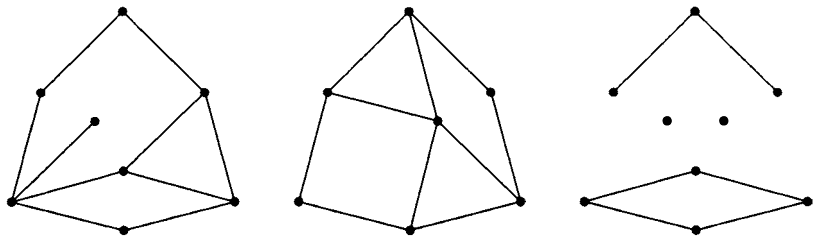


FIGURE I.2. A subgraph, an induced subgraph and a spanning subgraph of the graph in Fig. I.1.

We shall often construct new graphs from old ones by deleting or adding some vertices and edges. If $W \subset V(G)$, then $G - W = G[V \setminus W]$ is the subgraph of G obtained by deleting the vertices in W *and all edges incident with them.* Similarly, if $E' \subset E(G)$, then $G - E' = (V(G), E(G) \setminus E')$. If $W = \{w\}$ and $E' = \{xy\}$, then this notation is simplified to $G - w$ and $G - xy$. Similarly, if x and y are nonadjacent vertices of G, then $G + xy$ is obtained from G by joining x to y.

If x is a vertex of a graph G, then occasionally we write $x \in G$ instead of $x \in V(G)$. The *order* of G is the number of vertices in G; it is denoted by $|G|$. The same notation is used for the number of elements (cardinality) of a *set*: $|X|$ denotes the number of elements of the set X. Thus $|G| = |V(G)|$. The *size* of G is the number of edges in G; it is denoted by $e(G)$. We write G^n for an *arbitrary graph of order n*. Similarly, $G(n, m)$ denotes an *arbitrary graph of order n and size m*.

Given disjoint subsets U and W of the vertex set of a graph, we write $E(U, W)$ for the set of $U - W$ edges, that is, for the set of edges joining a vertex in U to a vertex in W. Also, $e(U, W) = |E(U, W)|$ is the number of $U - W$ edges. If we wish to emphasize that our underlying graph is G, then we put $E_G(U, W)$ and $e_G(U, W)$.

Two graphs are *isomorphic* if there is a correspondence between their vertex sets that preserves adjacency. Thus $G = (V, E)$ is isomorphic to $G' = (V', E')$ if there is a bijection $\phi : V \to V'$ such that $xy \in E$ iff $\phi(x)\phi(y) \in E'$. Clearly, isomorphic graphs have the same order and size. Usually we do not distinguish between isomorphic graphs, unless we consider graphs with a distinguished or labelled set of vertices (for example, subgraphs of a given graph). In accordance with this convention, if G and H are isomorphic graphs, then we write either $G \cong H$ or simply $G = H$. In Fig. I.3 we show all graphs (up to isomorphism) that have order at most 4 and size 3.

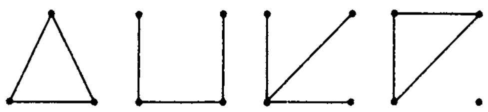


FIGURE I.3. Graphs of order at most 4 and size 3.

The size of a graph of order n is at least 0 and at most $\binom{n}{2}$. Clearly, for every m, $0 \le m \le \binom{n}{2}$, there is a graph $G(n, m)$. A graph of order n and size $\binom{n}{2}$ is called a *complete n-graph* and is denoted by K_n; an *empty n-graph* E_n has order n and no edges. In K_n every two vertices are adjacent, while in E_n no two vertices are adjacent. The graph $K_1 = E_1$ is said to be *trivial*.

As E_n is rather close to the notation for the edge set of a graph, we frequently use $\overline{K}_n$ for the empty graph of order n, signifying that it is the complement of the complete graph. In general, for a graph $G = (V, E)$ the *complement* of G is $\overline{G} = (V, V^{(2)} - E)$; thus, two vertices are adjacent in $\overline{G}$ if and only if they are *not* adjacent in G.

The set of vertices adjacent to a vertex $x \in G$, the *neighbourhood* of x, is denoted by $\Gamma(x)$. Occasionally one calls $\Gamma(x)$ the *open* neighbourhood of x, and $\Gamma \cup \{x\}$ the *closed* neighbourhood of x. Also, $x \sim y$ means that the vertex x is adjacent to the vertex y. Thus $y \in \Gamma(x)$, $x \in \Gamma(y)$, $x \sim y$, and $y \sim x$ are all equivalent: each of them means that xy is an edge. The *degree* of x is $d(x) = |\Gamma(x)|$. If we want to emphasize that the underlying graph is G, then we write $\Gamma_G(x)$ and $d_G(x)$; a similar convention will be adopted for other functions depending on an underlying graph. Thus if $x \in H = G[W]$, then

$$\Gamma_H(x) = \{y \in H : xy \in E(H)\} = \Gamma_G(x) \cap W.$$

The *minimal degree* of the vertices of a graph G is denoted by $\delta(G)$ and the *maximal degree* by $\Delta(G)$. A vertex of degree 0 is said to be an *isolated* vertex. If $\delta(G) = \Delta(G) = k$, that is, every vertex of G has degree k, then G is said to be *k-regular* or *regular of degree k*. A graph is *regular* if it is k-regular for some k. A 3-regular graph is said to be *cubic*.

If $V(G) = \{x_1, x_2, \ldots, x_n\}$, then $(d(x_i))_1^n$ is a *degree sequence* of G. Usually we order the vertices in such a way that the degree sequence obtained in this way is monotone increasing or monotone decreasing, for example, $\delta(G) = d(x_1) \leq \cdots \leq d(x_n) = \Delta(G)$. Since each edge has two endvertices, the sum of the degrees is exactly twice the number of edges:

$$\sum_1^n d(x_i) = 2e(G). \tag{1}$$

In particular, the sum of degrees is even:

$$\sum_1^n d(x_i) \equiv 0 \pmod 2. \tag{2}$$

This last observation is sometimes called the *handshaking lemma*, since it expresses the fact that in any party the total number of hands shaken is even. Equivalently, (2) states that the number of vertices of *odd* degree is *even*. We see also from (1) that $\delta(G) \leq \lfloor 2e(G)/n \rfloor$ and $\Delta(G) \geq \lceil 2e(G)/n \rceil$. Here $\lfloor x \rfloor$ denotes the greatest integer not greater than x and $\lceil x \rceil = -\lfloor -x \rfloor$ is the smallest integer not less than x.

A *path* is a graph P of the form

$$V(P) = \{x_0, x_1, \ldots, x_l\}, \qquad E(P) = \{x_0x_1, x_1x_2, \ldots, x_{l-1}x_l\}.$$

This path P is usually denoted by $x_0x_1 \cdots x_l$. The vertices x_0 and x_l are the *endvertices* of P and $l = e(P)$ is the *length* of P. We say that P is a path *from x_0 to x_l*, or an x_0–x_l path. Of course, P is also a path from x_l to x_0, or an x_l–x_0 path. Sometimes we wish to emphasize that P is considered to go from x_0 to x_l, and we then call x_0 the *initial* and x_l the *terminal* vertex of P. A path with initial vertex x is an *x-path.*

The term *independent* will be used in connection with vertices, edges, and paths of a graph. A set of *vertices* (*edges*) is *independent* if no two elements of it are adjacent; also, $W \subset V(G)$ consists of *independent vertices* iff $G[W]$ is an empty graph. A set of *paths* is *independent* if for any two paths each vertex belonging to both paths is an endvertex of both. Thus $P_1, P_2, \ldots, P_k$ are independent x–y paths iff $V(P_i) \cap V(P_j) = \{x, y\}$ whenever $i \neq j$. The paths P_i are also said to be *internally disjoint*. There are several notions closely related to that of a path in a graph. A *walk* W in a graph is an alternating sequence of vertices and edges, say $x_0, e_1, x_1, e_2, \ldots, e_l, x_l$ where $e_i = x_{i-1}x_i$, $0 < i \leq l$. In accordance with the terminology above, W is an x_0–x_l *walk* and is denoted by $x_0x_1 \cdots x_l$; the *length* of W is l. This walk W is called a *trail* if all its edges are distinct. Note that a path is a walk with distinct vertices. A trail whose endvertices coincide (a *closed* trail) is called a *circuit*. To be precise, a circuit is a closed trail without distinguished endvertices and direction, so that, for example, two triangles sharing a single vertex give rise to precisely *two* circuits with six edges. If a walk $W = x_0x_1 \cdots x_l$ is such that $l \geq 3$, $x_0 = x_l$, and the vertices x_i, $0 < i < l$, are distinct from each other and x_0, then W is said to be a *cycle*. For simplicity this cycle is denoted by $x_1x_2 \cdots x_l$. Note that the notation differs from that of a path since x_1x_l is also an edge of this cycle. A cycle has neither a starting vertex nor a direction, so that $x_1x_2 \cdots x_l$, $x_lx_{l-1} \cdots x_1$, $x_2x_3 \cdots x_lx_1$, and $x_ix_{i-1} \cdots x_1x_lx_{l-1} \cdots x_{i+1}$ all denote the same cycle.

We frequently use the symbol P_ℓ to denote an arbitrary *path of length* ℓ and C_ℓ to denote a *cycle of length* ℓ. We call C_3 a *triangle*, C_4 a *quadrilateral*, C_5 a *pentagon*, and so on; also, C_ℓ is called an *ℓ-cycle* (see Fig. I.4). A cycle is *even* (*odd*) if its length is even (odd).

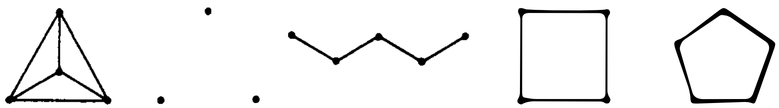


FIGURE I.4. The graphs K_4, E_3, P_4, C_4 and C_5.

It would be less confusing to use P^ℓ and C^ℓ for *generic* paths and cycles, and to reserve $P_1, P_2, \ldots, C_1, C_2, \ldots$ for *particular* paths and cycles. However, in order to conform to the widely accepted usage of subscripts, we also opt for subscripts, although with some reluctance. It is to be hoped that this convention will not lead to any misunderstanding.

Before continuing with our definitions, let us present two results concerning cycles. The first was noted by Veblen in 1912.

Theorem 1 *The edge set of a graph can be partitioned into cycles if, and only if, every vertex has even degree.*

Proof. The condition is clearly necessary, since if a graph is the union of some edge disjoint cycles and isolated vertices, then a vertex contained in k cycles has degree $2k$.

Suppose that every vertex of a graph G has even degree and $e(G) > 0$. How can we find a single cycle in G? Let $x_0x_1 \cdots x_\ell$ be a path of maximal length ℓ in G. Since $x_0x_1 \in E(G)$, we have $d(x_0) \geq 2$. But then x_0 has another neighbour y in addition to x_1; furthermore, we must have $y = x_i$ for some i, $2 \leq i \leq \ell$, since otherwise $yx_0x_1 \cdots x_\ell$ would be a path of length $\ell + 1$. Therefore, we have found our cycle: $x_0x_1 \cdots x_i$.

Having found one cycle, C_1, say, all we have to do is to repeat the procedure over and over again. To formalize this, set $G_1 = G$, so that C_1 is a cycle in G_1, and define $G_2 = G_1 - E(C_1)$. Every vertex of G_2 has even degree, so either

$E(G_2) = \emptyset$ or else G_2 contains a cycle C_2. Continuing in this way, we find vertex disjoint cycles $C_1, C_2, \ldots, C_s$ such that $E(G) = \bigcup_{i=1}^{s} E(C_i)$. $\square$

To prove the second result, a beautiful theorem of Mantel from 1907, we shall use observation (1) and Cauchy's inequality.

Theorem 2 *Every graph of order n and size greater than* $\lfloor n^2/4 \rfloor$ *contains a triangle.*

Proof. Let G be a triangle-free graph of order n. Then $\Gamma(x) \cap \Gamma(y) = \emptyset$ for every edge $xy \in E(G)$, so

$$d(x) + d(y) \leq n.$$

Summing these inequalities for all $e(G)$ edges xy, we find that

$$\sum_{x \in G} d(x)^2 \leq ne(G). \tag{3}$$

Now by (1) and Cauchy's inequality,

$$(2e(G))^2 = \left(\sum_{x \in G} d(x) \right)^2 \leq n \left(\sum_{x \in G} d(x)^2 \right).$$

Hence, by (3),

$$(2e(G))^2 \leq n^2 e(G),$$

implying that $e(G) \leq n^2/4$. $\square$

The bound in this result is easily seen to be best possible (see Exercise 4). Mantel's theorem was greatly extended by Turán in 1941: as we shall see in Chapter IV, this theorem of Turán is the starting point of extremal graph theory.

Given vertices x and y, their *distance* $d(x, y)$ is the minimal length of an x–y path. If there is no x–y path then $d(x, y) = \infty$.

A graph is *connected* if for every pair $\{x, y\}$ of distinct vertices there is a path from x to y. Note that a connected graph of order at least 2 cannot contain an isolated vertex. A *maximal connected subgraph* is a *component* of the graph. A *cutvertex* is a vertex whose deletion increases the number of components. Similarly, an edge is a *bridge* if its deletion increases the number of components. Thus an edge of a connected graph is a bridge if its deletion disconnects the graph. A graph without any cycles is a *forest*, or an *acyclic* graph; a *tree* is a connected forest. (See Fig. I.5.) The relation of a tree to a forest sounds less absurd if we note that a forest is a disjoint union of trees; in other words, a forest is a graph whose every component is a tree.

A graph G is a *bipartite graph with vertex classes* V_1 and V_2 if $V(G) = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$ and every edge joins a vertex of V_1 to a vertex of V_2. One also says that G has *bipartition* (V_1, V_2). Similarly G is *r-partite with vertex classes* $V_1, V_2, \ldots, V_r$ (or r-partition $(V_1, \ldots, V_r)$) if $V(G) = V_1 \cup V_2 \cup \cdots \cup V_r$, $V_i \cap V_j = \emptyset$ whenever $1 \leq i < j \leq r$, and no edge joins two vertices in the same class. The graphs in Fig. I.1 and Fig. I.5 are bipartite. The symbol $K(n_1, \ldots, n_r)$

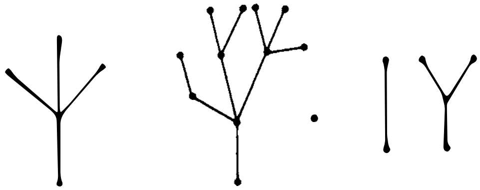


FIGURE I.5. A forest.

denotes a *complete r-partite graph*: it has n_i vertices in the i th class and contains all edges joining vertices in distinct classes. For simplicity, we often write $K_{p,q}$ instead of $K(p, q)$ and $K_r(t)$ instead of $K(t, \dots, t)$.

We shall write $G \cup H = (V(G) \cup V(H), E(G) \cup E(H))$ and kG for the union of k disjoint copies of G. We obtain the *join* $G + H$ from $G \cup H$ by adding all edges between G and H. Thus, for example, $K_{2,3} = E_2 + E_3 = \overline{K}_2 + \overline{K}_3$ and $K_r(t) = E_t + \dots E_t = \overline{K}_t + \dots + \overline{K}_t$.

There are several notions closely related to that of a graph. A *hypergraph* is a pair (V, E) such that $V \cap E = \emptyset$ and E is a subset of $\mathcal{P}(V)$, the power set of V, that is the set of all subsets of V (see Fig. I.6). In fact, there is a simple 1-to-1 correspondence between the class of hypergraphs and the class of certain bipartite graphs. Given a hypergraph $H = (V, E)$, the *incidence graph* of H is the bipartite

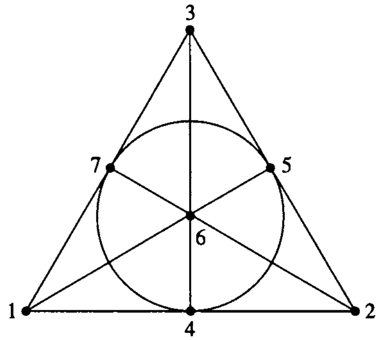


FIGURE I.6. The hypergraph of the Fano plane, the projective plane $PG(2, 2)$ of seven points and seven lines: the lines are 124, 235, 346, 457, 561, 672, and 713.

graph with vertex classes V and E in which $x \in V$ is joined to a hyperedge $S \in E$ iff $x \in S$ (see Fig. I.7).

By definition a graph does not contain a *loop*, an "edge" joining a vertex to itself; neither does it contain *multiple edges*, that is, several "edges" joining the same two vertices. In a *multigraph* both multiple edges and multiple loops are allowed; a loop is a special edge. When there is any danger of confusion, graphs are called *simple graphs*. In this book the emphasis will be on graphs rather than multigraphs. However, sometimes multigraphs are the natural context for our results, and it is artificial to restrict ourselves to (simple) graphs. For example, Theorem 1 is valid

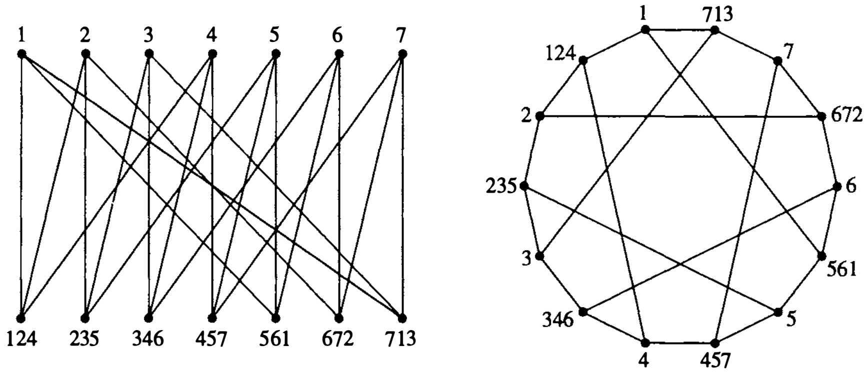


FIGURE I.7. The drawings of the Heawood graph, the incidence graph of the Fano plane in Fig. I.6.

for multigraphs, provided that a loop is taken to contribute 2 to the degree of a vertex, and we allow cycles of length 1 (loops) and length 2 (formed by two edges joining the same vertices.

If the edges are *ordered* pairs of vertices, then we get the notions of a *directed graph* and *directed multigraph*. An ordered pair (a, b) is said to be an edge *directed from a to b*, or an edge *beginning at a and ending at b*, and is denoted by $\overrightarrow{ab}$ or simply ab. The notions defined for graphs are easily carried over to multigraphs, directed graphs, and directed multigraphs, mutatis mutandis. Thus a *(directed) trail* in a directed multigraph is an alternating sequence of vertices and edges: $x_0, e_1, x_1, e_2, \ldots, e_l, x_l$, such that e_i begins at x_{i-1} and ends at x_i. Also, a vertex x of a directed graph has an outdegree and an indegree: the *outdegree* $d^+(x)$ is the number of edges starting at x, and the *indegree* $d^-(x)$ is the number of edges ending at x.

An *oriented graph* is a directed graph obtained by orienting the edges of a graph, that is, by giving the edge ab an orientation $\overrightarrow{ab}$ or $\overrightarrow{ba}$. Thus an oriented graph is a directed graph in which at most one of $\overrightarrow{ab}$ and $\overrightarrow{ba}$ occurs.

Note that Theorem 1 has a natural variant for directed multigraphs as well: the edge set of a directed multigraph can be partitioned into (directed) cycles if and only if each vertex has the same outdegree as indegree, that is, $d^+(x) = d^-(x)$ for every vertex x. To see the sufficiency of the condition, all we have to notice is that, as before, if our graph has an edge, then it has a (directed) cycle as well.

I.2 Paths, Cycles, and Trees

With the concepts defined so far we can start proving some results about graphs. Though these results are hardly more than simple observations, in keeping with the style of the other chapters we shall call them theorems.

Theorem 3 *Let x be a vertex of a graph G and let W be the vertex set of a component containing x. Then the following assertions hold.*

i. $W = \{y \in G : G \textit{ contains an } x\text{–}y \textit{ path}\}$.
ii. $W = \{y \in G : G \textit{ contains an } x\text{–}y \textit{ trail}\}$.
iii. $W = \{y \in G : d(x, y) < \infty\}$.
iv. *For $u, v \in V = V(G)$ put uRv iff $uv \in E(G)$, and let $\tilde{R}$ be the smallest equivalence relation on V containing R. Then W is the equivalence class of x.* □

This little result implies that every graph is the vertex disjoint union of its components (equivalently, every vertex is contained in a unique component), and that an edge is a bridge iff it is not contained in a cycle.

Theorem 4 *A graph is bipartite iff it does not contain an odd cycle.*

Proof. Suppose G is bipartite with vertex classes V_1 and V_2. Let $x_1x_2 \cdots x_l$ be a cycle in G. We may assume that $x_1 \in V_1$. Then $x_2 \in V_2$, $x_3 \in V_1$, and so on: $x_i \in V_1$ iff i is odd. Since $x_l \in V_2$, we find that l is even.

Suppose now that G does not contain an odd cycle. Since a graph is bipartite iff each component of it is, we may assume that G is connected. Pick a vertex $x \in V(G)$ and put $V_1 = \{y : d(x, y) \text{ is odd}\}$, $V_2 = V \setminus V_1$. There is no edge joining two vertices of the same class V_i, since otherwise G would contain an odd cycle. Hence G is bipartite. □

A bipartite graph with bipartition (V_1, V_2) has at most $|V_1||V_2|$ edges, so a bipartite graph of order n has at most $\max_k k(n-k) = \lfloor n^2/4 \rfloor$ edges, with the maximum attained at the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. By Theorem 4, $\lfloor n^2/4 \rfloor$ is also the maximal size of a graph of order n containing no odd cycles. In fact, as we saw in Theorem 2, forbidding a single odd cycle, the triangle, restricts the size just as much.

Theorem 5 *A graph is a forest iff for every pair $\{x, y\}$ of distinct vertices it contains at most one x–y path.*

Proof. If $x_1x_2 \cdots x_l$ is a cycle in a graph G, then $x_1x_2 \cdots x_l$ and x_1x_l are two x_1–x_l paths in G.

Conversely, let $P_1 = x_0x_1 \cdots x_l$ and $P_2 = x_0y_1y_2 \cdots y_kx_l$ be two distinct x_0–x_l paths in a graph G. Let $i+1$ be the minimal index for which $x_{i+1} \neq y_{i+1}$ and let j be the minimal index for which $j \geq i$ and y_{j+1} is a vertex of P_1, say $y_{j+1} = x_h$. Then $x_ix_{i+1} \cdots x_hy_jy_{j-1} \cdots y_{i+1}$ is a cycle in G. □

Theorem 6 *The following assertions are equivalent for a graph G.*

i. *G is a tree.*
ii. *G is a minimal connected graph, that is, G is connected and if $xy \in E(G)$, then $G - xy$ is disconnected. [In other words, G is connected and every edge is a bridge.]*
iii. *G is a maximal acyclic graph; that is, G is acyclic and if x and y are nonadjacent vertices of G, then $G + xy$ contains a cycle.*

Proof. Suppose G is a tree. For an edge $xy \in E(G)$, the graph $G - xy$ cannot contain an x–y *path* $xz_1z_2 \cdots z_ky$, since otherwise G contains the *cycle* $xz_1z_2 \cdots z_ky$. Hence $G - xy$ is disconnected; and so G is a minimal connected graph. Similarly, if x and y are nonadjacent vertices of the tree G then G, contains a *path* $xz_1z_2 \cdots z_ky$, and so $G + xy$ contains the *cycle* $xz_1z_2 \cdots z_ky$. Hence $G + xy$ contains a cycle, and so G is a maximal acyclic graph.

Suppose next that G is a minimal connected graph. If G contains a *cycle* $xz_1z_2 \cdots z_ky$, then $G - xy$ is still connected, since in any u–v walk in G the edge xy can be replaced by the *path* $xz_1z_2 \cdots z_ky$. As this contradicts the minimality of G, we conclude that G is acyclic and so it is a tree.

Suppose, finally, that G is a maximal acyclic graph. Is G connected? Yes, since if x and y belong to different components, the addition of xy to G cannot create a *cycle* $xz_1z_2 \cdots z_ky$, since otherwise the path $xz_1z_2 \cdots z_ky$ is in G. Thus G is a tree. □

Corollary 7 *Every connected graph contains a spanning tree, that is, a tree containing every vertex of the graph.*

Proof. Take a minimal connected spanning subgraph. □

There are several simple constructions of a spanning tree of a graph G; we present two of them. Pick a vertex x and put $V_i = \{y \in G : d(x, y) = i\}$, $i = 0, 1, \ldots$. Note that if $y_i \in V_i$, $i > 0$, and $xz_1z_2 \cdots z_{i-1}y_i$ is an x–y_i path (whose existence is guaranteed by the definition of V_i), then $d(x, z_j) = j$ for every j, $0 < j < i$. In particular, $V_j \neq \emptyset$, and for every $y \in V_i$, $i > 0$, there is a vertex $y' \in V_{i-1}$ joined to y. (Of course, this vertex y' is usually not unique, but for each $y \neq x$ we *pick* only one y'.) Let T be the subgraph of G with vertex set V and edge set $E(T) = \{yy' : y \neq x\}$. Then T is connected, since every $y \in V - \{x\}$ is joined to x by a path $yy'y'' \cdots x$. Furthermore, T is acyclic, since if W is any subset of V and w is a vertex in W furthest from x, then w is joined to at most one vertex in W. Thus T is a spanning tree.

The argument above shows that with $k = \max_y d(x, y)$, we have $V_i \neq \emptyset$ for $0 \leq i \leq k$ and $V = V(G) = \bigcup_0^k V_i$. At this point it is difficult to resist the remark that $\mathrm{diam} G = \max_{x,y} d(x, y)$ is called the *diameter* of G and $\mathrm{rad} G = \min_x \max_y d(x, y)$ is the *radius* of G.

If we choose $x \in G$ with $k = \max_y d(x, y) = \mathrm{rad} G$, then the spanning tree T also has radius k.

A slight variant of the above construction of T goes as follows. Pick $x \in G$ and let T_1 be the subgraph of G with this single vertex x. Then T_1 is a tree. Suppose we have constructed trees $T_1 \subset T_2 \subset \cdots \subset T_k \subset G$, where T_i has order i. If $k < n = |G|$ then by the connectedness of G there is a vertex $y \in V(G) \setminus V(T_k)$ that is adjacent (in G) to a vertex $z \in T_k$. Let T_{k+1} be obtained from T_k by adding to it the vertex y and the edge yz. Then T_{k+1} is connected and as yz cannot be an edge of a cycle in T_{k+1}, it is also acyclic. Thus T_{k+1} is also a tree, so the sequence $T_0 \subset T_1 \subset \cdots$ can be continued to T_n. This tree T_n is then a spanning tree of G.

The spanning trees constructed by either of the methods above have order n (of course!) and size $n-1$. In the first construction there is a 1-to-1 correspondence between $V-\{x\}$ and $E(T)$, given by $y \mapsto yy'$, and in the second construction $e(T_k)=k-1$ for each k, since $e(T_1)=0$ and T_{k+1} has one more edge than T_k. Since by Theorem 6 every tree has a *unique* spanning tree, namely itself, we have arrived at the following result, observed by Listing in 1862.

Corollary 8 *A tree of order n has size $n-1$; a forest of order n with k components has size $n-k$.*

The first part of this corollary can be incorporated into several other characterizations of trees. In particular, a graph of order n is a tree iff it is connected and has size $n-1$. The reader is invited to prove these characterizations (Exercises 5 and 6).

Corollary 9 *A tree of order at least 2 contains at least 2 vertices of degree 1.*

Proof. Let $d_1 \le d_2 \le \cdots \le d_n$ be the degree sequence of a tree T of order $n \ge 2$. Since T is connected, $\delta(T)=d_1 \ge 1$. Hence if T had at most one vertex of degree 1, by (1) and Corollary 8 we would have

$$2n-2=2e(T)=\sum_1^n d_i \ge 1+2(n-1). \qquad \square$$

A well-known problem in optimization theory asks for a relatively easy way of finding a spanning subgraph with a special property. Given a graph $G=(V,E)$ and a positive valued cost function f defined on the edges, $f: E \to \mathbb{R}^+$, find a connected spanning subgraph $T=(V,E')$ of G for which

$$f(T)=\sum_{xy\in E'} f(xy)$$

is minimal. We call such a spanning subgraph T an *economical* spanning subgraph. One does not need much imagination to translate this into a "real life" problem. Suppose certain villages in an area are to be joined to a water supply situated in one of the villages. The system of pipes is to consist of pipelines connecting the water towers of two villages. For any two villages we know how much it would cost to build a pipeline connecting them, provided such a pipeline can be built at all. How can we find an economical system of pipes?

In order to reduce the second question to the above problem about graphs, let G be the graph whose vertex set is the set of villages and in which xy is an edge iff it is possible to build a pipeline joining x to y; denote the cost of such a pipeline by $f(xy)$ (see Fig. I.8). Then a system of pipes corresponds to a connected spanning subgraph T of G. Since the system has to be economical, T is a minimal connected spanning subgraph of G, that is, a spanning tree of G.

The connected spanning subgraph T we look for has to be a *minimal* connected subgraph, since otherwise we could find an edge α whose deletion would leave T connected, and then $T-\alpha$ would be a more economical spanning subgraph.

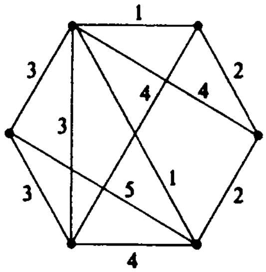


FIGURE I.8. A graph with a function $f : E \to \mathbb{R}^+$; the number next to an edge xy is the cost $f(xy)$ of the edge.

Thus T is a spanning tree of G. Corresponding to the various characterizations and constructions of a spanning tree, we have several easy ways of finding an economical spanning tree; we shall describe four of these methods.

(1) Given G and $f : E(G) \to \mathbb{R}^+$, we choose one of the cheapest edges of G, that is, an edge α for which $f(\alpha)$ is minimal. Each subsequent edge will be chosen from among the cheapest remaining edges of G with the only restriction that we must not select all edges of any cycle; that is, the subgraph of G formed by the selected edges is acyclic.

The process terminates when no edge can be added to the set E' of edges selected so far without creating a cycle. Then $T_1 = (V(G), E')$ is a maximal acyclic subgraph of G, so by Theorem 6(iii), it is a spanning tree of G.

(2) This method is based on the fact that it is foolish to use a costly edge unless it is needed to ensure the connectedness of the subgraph. Thus let us delete one by one a costliest edge whose deletion does not disconnect the graph. By Theorem 6(ii) the process ends in a spanning tree T_2.

(3) Pick a vertex x_1 of G and select one of the least costly edges incident with x_1, say x_1x_2. Then choose one of the least costly edges of the form x_ix, where $1 \leq i \leq 2$ and $x \notin \{x_1, x_2\}$. Having found vertices $x_1, x_2, \ldots, x_k$ and an edge x_ix_j, $i < j$, for each vertex x_j with $j \leq k$, select one of the least costly edges of the form x_ix, say x_ix_{k+1}, where $1 \leq i \leq k$ and $x_{k+1} \notin \{x_1, x_2, \ldots, x_k\}$. The process terminates after we have selected $n - 1$ edges. Denote by T_3 the spanning tree given by these edges (see Fig. I.9).

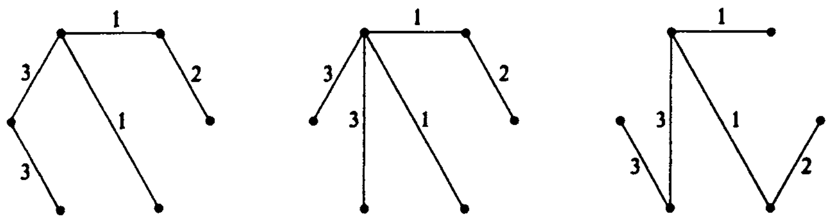


FIGURE I.9. Three of the six economical spanning trees of the graph shown in Fig. I.8.

(4) This method is applicable only if no two pipelines cost the same. The advantage of the method is that every village can make its own decision and start building a pipeline without bothering to find out what the other villages are going to do. Of course, each village will start building the cheapest pipeline ending in the village. It may happen that both village x and village y will build the pipeline xy; in this case they meet in the middle and end up with a single pipeline from x to y. Thus at the end of this stage some villages will be joined by pipelines but the whole system of pipes need not be connected. At the next stage each group of villages joined to each other by pipelines finds the cheapest pipeline going to a village not in the group and begins to build that single pipeline. The same procedure is repeated until a connected system is obtained. Clearly, the villages will never build all the pipes of a cycle, so the final system of pipes will be a spanning tree (see Fig. I.10).

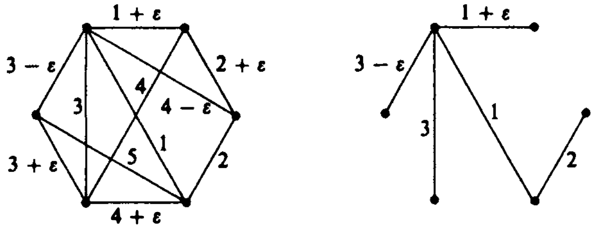


FIGURE I.10. The graph of Fig. I.8 with a slightly altered cost function ($0 < \varepsilon < \frac{1}{2}$) and its unique economical spanning tree.

Theorem 10 *Each of the four methods described above produces an economical spanning tree. If no two edges have the same cost, then there is a unique economical spanning tree.*

Proof. Choose an economical spanning tree T of G that has as many edges in common with T_1 as possible, where T_1 is a spanning tree constructed by the first method.

Suppose that $E(T_1) \neq E(T)$. The edges of T_1 have been selected one by one: let xy be the first edge of T_1 that is not an edge of T. Then T contains a unique $x - y$ path, say P. This path P has at least one edge, say uv, that does not belong to T_1, since otherwise T_1 would contain a cycle. When xy was selected as an edge of T_1, the edge uv was also a candidate. As xy was chosen and not uv, the edge xy cannot be costlier then uv; that is, $f(xy) \leq f(uv)$. Then $T' = T - uv + xy$ is a spanning tree, and since $f(T') = f(T) - f(uv) + f(xy) \leq f(T)$, the new tree T' is an economical spanning tree of G. (Of course, this inequality implies that $f(T') = f(T)$ and $f(xy) = f(uv)$.) This tree T' has more edges in common with T_1 than T, contradicting the choice of T. Hence $T = T_1$, so T_1 is indeed an economical spanning tree.

Slight variants of the proof above show that the spanning trees T_2 and T_3, constructed by the second and third methods, are also economical. We invite the reader to furnish the details (Exercise 44).

Suppose now that no two edges have the same cost; that is, $f(xy) \neq f(uv)$ whenever $xy \neq uv$. Let T_4 be the spanning tree constructed by the fourth method and let T be an economical spanning tree. Suppose that $T \neq T_4$, and let xy be the first edge not in T that we select for T_4. The edge xy was selected, since it is the *least costly* edge of G joining a vertex of a subtree F of T_4 to a vertex outside F. The $x - y$ path in T has an edge uv joining a vertex of F to a vertex outside F so $f(xy) < f(uv)$. However, this is impossible, since $T' = T - uv + xy$ is a spanning tree of G and $f(T') < f(T)$. Hence $T = T_4$. This shows that T_4 is indeed an economical spanning tree. Furthermore, since the spanning tree constructed by the fourth method is unique, the economical spanning tree is unique if no two edges have the same cost. □

I.3 Hamilton Cycles and Euler Circuits

The so-called *travelling salesman problem* greatly resembles the economical spanning tree problem discussed in the preceding section, but the similarity is only superficial. A salesman is to make a tour of n cities, at the end of which he has to return to the head office he starts from. The cost of the journey between any two cities is known. The problem asks for an efficient algorithm for finding a least expensive tour. (As we shall not deal with algorithmic problems, we leave the term "efficient" undefined; loosely speaking, an algorithm is efficient if the computing time is bounded by a polynomial in the number of vertices.) Though a considerable amount of work has been done on this problem, since its solution would have important practical applications, it is not known whether or not there is an efficient algorithm for finding a least expensive route.

In another version of the travelling salesman problem the route is required to be a cycle, that is, the salesman is not allowed to visit the same city twice (except the city of the head office, where he starts and ends his journey). A cycle containing all the vertices of a graph is said to be a *Hamilton cycle* of the graph. The origin of this term is a game invented in 1857 by Sir William Rowan Hamilton based on the construction of cycles containing all the vertices in the graph of the dodecahedron (see Fig. I.11). A *Hamilton path* of a graph is a path containing all the vertices of the graph. A graph containing a Hamilton cycle is said to be *Hamiltonian*.

In fact, Hamilton cycles and paths in special graphs had been studied well before Hamilton proposed his game. In particular, the puzzle of the knight's tour on a chess board, thoroughly analysed by Euler in 1759, asks for a Hamilton cycle in the graph whose vertices are the 64 squares of a chessboard and in which two vertices are adjacent if a knight can jump from one square to the other. Fig. I.12 shows two solutions of this puzzle.

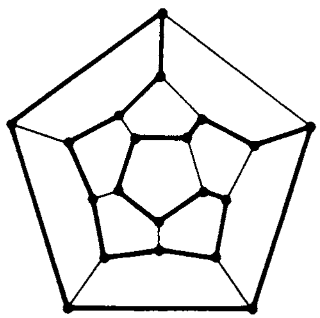


FIGURE I.11. A Hamilton cycle in the graph of the dodecahedron.

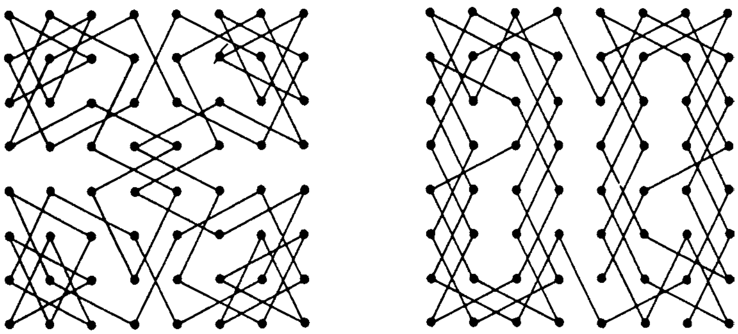


FIGURE I.12. Two tours of a knight on a chessboard.

If in the second, more restrictive, version of the travelling salesman problem there are only two travel costs, 1 and ∞ (expressing the impossibility of the journey), then the question is whether or not the graph formed by the edges with travel cost 1 contains a Hamilton cycle. Even this special case of the travelling salesman problem is unsolved: no efficient algorithm is known for constructing a Hamilton cycle, though neither is it known that there is no such algorithm.

If the travel cost between any two cities is the same, then our salesman has no difficulty in finding a least expensive tour: any permutation of the $n-1$ cities (the nth city is that of the head office) will do. Revelling in his new found freedom, our salesman decides to connect duty and pleasure, and promises not to take the same road xy again whilst there is a road uv he hasn't seen yet. Can he keep his promise? In order to plan a required sequence of journeys for our salesman, we have to decompose K_n into the union of some edge-disjoint Hamilton cycles. For which values of n is this possible? Since K_n is $(n-1)$-regular and a Hamilton cycle is 2-regular, a *necessary* condition is that $n-1$ should be even, that is, n should be odd. This necessary condition also follows from the fact that $e(K_n) = \frac{1}{2}n(n-1)$ and a Hamilton cycle contains n edges, so K_n has to be the union of $\frac{1}{2}(n-1)$ Hamilton cycles.

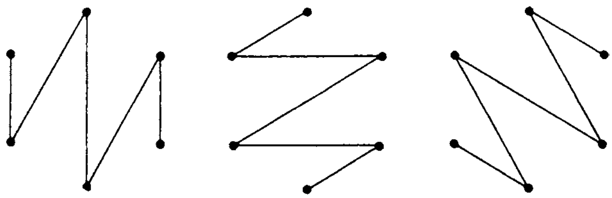


FIGURE I.13. Three edge disjoint Hamilton paths in K_6.

Let us assume now that n is odd, $n \geq 3$. Deleting a vertex of K_n we see that if K_n is the union of $\frac{1}{2}(n-1)$ Hamilton cycles then K_{n-1} is the union of $\frac{1}{2}(n-1)$ Hamilton paths. (In fact, $n-1$ has to be even if K_{n-1} is the union of some Hamilton paths, since $e(K_{n-1}) = \frac{1}{2}(n-1)(n-2)$ and a Hamilton path in K_{n-1} has $n-2$ edges.) With the hint shown in Fig. I.13 the reader can show that for odd values of n the graph K_{n-1} is indeed the union of $\frac{1}{2}(n-1)$ Hamilton paths. In this decomposition of K_{n-1} into $\frac{1}{2}(n-1)$ Hamilton paths each vertex is the endvertex of exactly one Hamilton path. (In fact, this holds for every decomposition of K_{n-1} into $\frac{1}{2}(n-1)$ edge-disjoint Hamilton paths, since each vertex x of K_{n-1} has odd degree, so at least one Hamilton path has to end in x.) Consequently, if we add a new vertex to K_{n-1} and extend each Hamilton path in K_{n-1} to a Hamilton cycle in K_n, then we obtain a decomposition of K_n into $\frac{1}{2}(n-1)$ edge-disjoint Hamilton cycles. Thus we have proved the following result.

Theorem 11 *For $n \geq 3$ the complete graph K_n is decomposable into edge disjoint Hamilton cycles iff n is odd. For $n \geq 2$ the complete graph K_n is decomposable into edge-disjoint Hamilton paths iff n is even.*

The result above shows that if $n \geq 3$ is odd, then we can string together $\frac{1}{2}(n-1)$ edge disjoint cycles in K_n to obtain a circuit containing all the edges of K_n. In general, a circuit in a graph G containing all the edges is said to be an *Euler circuit* of G. Similarly, a trail containing all edges is an *Euler trail.*

A graph is *Eulerian* if it has an Euler circuit. Euler circuits and trails are named after Leonhard Euler, who, in 1736, characterized those graphs that contain them. At the time Euler was a professor of mathematics in St. Petersburg, and was led to the problem by the puzzle of the seven bridges on the Pregel (see Fig. I.14) in the ancient Prussian city Königsberg (birthplace and home of Kant and seat of a great German university, which was taken over by the USSR and renamed Kaliningrad in 1946; since the collapse of the Soviet Union it has belonged to Russia). The good burghers of Königsberg wondered whether it was possible to plan a walk in such a way that each bridge would be crossed once and only once? It is clear that such a walk is possible iff the graph (or multigraph) in Fig. I.15 has an Euler trail. Here is then Euler's theorem inspired by the puzzle of the bridges of Königsberg.

Theorem 12 *A non-trivial connected graph has an Euler circuit iff each vertex has even degree.*

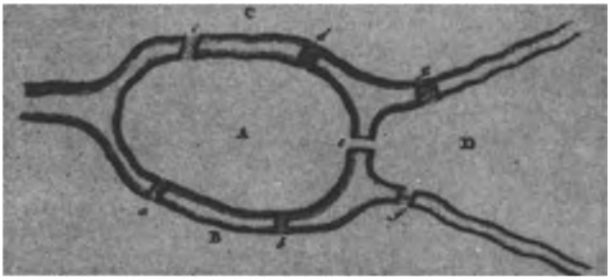


FIGURE I.14. The seven bridges on the Pregel in Königsberg.

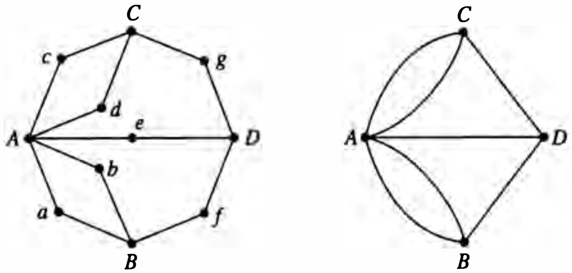


FIGURE I.15. A graph of the Königsberg bridges and its simpler representation by a multigraph.

A connected graph has an Euler trail from a vertex x to a vertex $y \neq x$ iff x and y are the only vertices of odd degree.

Proof. The conditions are clearly necessary. For example, if G has an Euler circuit $x_1x_2 \cdots x_m$, and x occurs k times in the sequence $x_1, x_2, \ldots, x_m$, then $d(x) = 2k$.

We prove the sufficiency of the first condition by induction on the number of edges. If there are no edges, there is nothing to prove, so we proceed to the induction step.

Let G be a non-trivial connected graph in which each vertex has even degree. Since $e(G) \geq 1$, we find that $\delta(G) \geq 2$, so by Corollary 9, G contains a cycle. Let C be a *circuit* in G with the maximal number of edges. Suppose C is not Eulerian. As G is connected, C contains a vertex x that is in a non-trivial component H of $G - E(C)$. Every vertex of H has even degree in H, so by the induction hypothesis, H contains an Euler circuit D. The circuits C and D (see Fig. I.16) are edge-disjoint and have a vertex in common, so they can be concatenated to form a circuit with more edges than C. As this contradicts the maximality of $e(C)$, the circuit C is Eulerian.

Suppose now that G is connected and x and y are the only vertices of odd degree. Let G^* be obtained from G by adding to it a vertex u together with the edges ux and uy. Then, by the first part, G^* has an Euler circuit C^*. Clearly, $C^* - u$ is an Euler trail from x to y. □

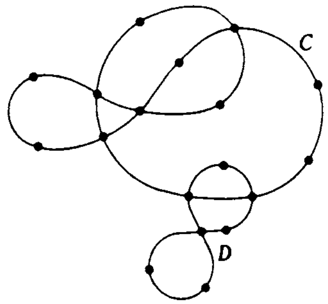


FIGURE I.16. The circuits C and D.

The alert reader has no doubt noticed that Theorem 12 is practically the same as Theorem 1: every Euler circuit is a union of edge-disjoint cycles, and if a connected graph is a union of edge-disjoint cycles, then these cycles can be concatenated to form an Euler circuit. Like Theorem 1, Theorem 12 holds for multigraphs as well: in fact, the natural models that arise (as in Fig. I.15) are frequently multigraphs.

It is also very easy to guess the variant of Theorem 12 for directed multigraphs: a directed multigraph has a (directed) Euler circuit if and only if the underlying multigraph is connected and each vertex has the same outdegree as indegree. To see this, we proceed as before, but take care to concatenate the circuits in the right (that is, permissible) direction.

There is a beautiful connection between the set of Euler circuits and certain sets of oriented spanning trees. In order to state this connection precisely, let G be a directed multigraph with vertex set $V(G) = \{v_1, \ldots, v_n\}$, such that $d^+(v_i) = d^-(v_i)$ for every i. We know that if G has a (directed) Euler circuit, then these conditions are satisfied. Let $\mathcal{E}$ be the set of (directed) Euler circuits, and let $\mathcal{E}_i$ be the set of (directed) Euler trails starting and ending in v_i. Since each Euler circuit passes through v_i exactly $d^+(v_i) = d^-(v_i)$ times, $|\mathcal{E}_i| = d^+(v_i)|\mathcal{E}| = d^-(v_i)|\mathcal{E}|$.

We say that a spanning tree is *oriented towards* v_i, its *root*, if for every $j \neq i$ the unique path from v_j to v_i is oriented towards v_i. Let $\mathcal{T}_i$ be the set of spanning trees oriented towards v_i.

Our aim is to define a map $\phi_i : \mathcal{E}_i \to \mathcal{T}_i$, but for notational simplicity we take $i = 1$. Given an Euler trail $S \in \mathcal{E}_1$, for $j = 2, \ldots, n$, let e_j be the edge through which S exits from v_j for the last time, never to return to v_j. In particular, e_j is not a loop but an edge from v_j to another vertex. Also, if e_i goes from v_i to v_j then on S the edge e_i precedes e_j.

Let T be the directed graph with vertices $v_1, \ldots, v_n$ and edges $e_2, \ldots, e_n$. We claim that $T \in \mathcal{T}_1$. To prove this, we have to show that (1) T is a tree, and (2) T is oriented towards v_1.

Suppose first that T contains a cycle C. Since $d_T^+(v_1) = 0$ and $d_T^+(v_j) = 1$ for $j > 1$, it follows that C is an oriented cycle that does not contain v_1. But if e_l is the last edge of S on C, going from v_l to v_m, say, then S gets back to v_m after having left it for the last time (through e_m). This contradiction shows that T is indeed a tree.

Is T oriented towards v_1? Suppose T contains the path $v_k v_{k-1} \cdots v_1$. Then the edge $v_2 v_1$ is e_2, since there is no e_1. What about $v_3 v_2$? It is either e_2 or e_3. But it is not e_2, so it is e_3. Continuing in this way, we find that our path $v_k v_{k-1} \cdots v_1$ is indeed oriented towards v_1. Hence $T \in \mathcal{T}_1$, as claimed.

To get our map $\phi_1 : \mathcal{E}_1 \to \mathcal{T}_1$, set $\phi_1(S) = T$. Now, for $T \in \mathcal{T}_1$, the set $\phi_1^{-1}(T)$ is easily described. Indeed, to construct an Euler trail $S \in \mathcal{E}_1$ with $\phi_1(S) = T$, one has to proceed as follows. Start at v_1 through any edge; also, having returned to v_1, leave it by an unused edge, if there is any; otherwise; terminate the trail. More importantly, having arrived in v_j, $j > 1$, leave v_j by an unused edge that is different from e_j, if there are any such edges; otherwise, leave v_j by e_j. Since $d^+(v_j) = d^-(v_j)$ for every j, this process does give us an Euler trail $S \in \mathcal{E}_1$ with $\phi_1(S) = T$. Consequently,

$$|\phi_1^{-1}(T)| = d^+(v_1)! \prod_{j=2}^{n} (d^+(v_j) - 1)!,$$

and so

$$|\mathcal{E}| = |\mathcal{T}_1| \prod_{j=1}^{n} (d^+(v_j) - 1)!.$$

With this, we have proved a theorem of de Bruijn, van Aardenne-Ehrenfest, Smith, and Tutte; the result is occasionally called the BEST theorem.

Theorem 13 *Let G be a directed multigraph with vertex set $V(G) = \{v_1, \ldots, v_n\}$, such that $d^+(v_i) = d^-(v_i)$ for every i. Denote by $s(G)$ the number of Euler circuits of G, and by $t_i(G)$ the number of spanning trees oriented towards i. Then*

$$s(G) = t_i(G) \prod_{j=1}^{n} (d^+(v_j) - 1)!$$

for every i, $1 \le i \le n$. In particular, $t_1(G) = \cdots = t_n(G)$.

Note that the conditions of Theorem 13 are satisfied if G is *Eulerian*, that is, has an Euler circuit.

Concerning the puzzle of the seven bridges on the Pregel, Theorem 12 tells us that there is no suitable tour, since the associated graph in Fig. I.15 has four vertices of odd degree (and, needless to say, so has the associated multigraph: each of its vertices has odd degree).

The plan of the corridors of an exhibition is also easily turned into a graph: an edge corresponds to a corridor and a vertex to the conjunction of several corridors. If the entrance and exit are the same, a visitor can walk along every corridor exactly once iff the corresponding graph has an Eulerian circuit. In general, a visitor must have a plan in order to achieve this: he cannot just walk through any new corridor he happens to come to. However, in a well planned (!) exhibition a visitor would be certain to see all the exhibits, provided that he avoided going along the same corridor twice and continued his walk until there were no new exhibits ahead of him. The graph of such an exhibition is said to be *randomly Eulerian* from the

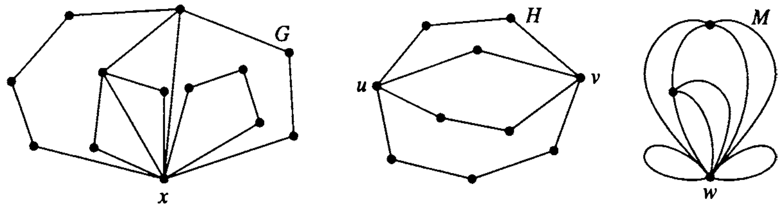


FIGURE I.17. The graph G is randomly Eulerian from x; H is randomly Eulerian from both u and v; the multigraph M is randomly Eulerian from w.

vertex corresponding to the entrance (which is also the exit). See Fig. I.17 for three examples. Randomly Eulerian graphs are also easily characterized (Exercises 50–52).

To conclude this section, let us note a result from the first half of this century, concerning two-way infinite Euler trails in infinite graphs. These are the natural analogues of Euler circuits in finite graphs: given an infinite graph $G = (V, E)$, a *two-way infinite Euler trail* in G is a two-way infinite sequence $\cdots x_{-2}x_{-1}x_0x_1x_2\cdots$ of vertices of G such that $x_i \sim x_{i+1}$ for all $i \in \mathbb{Z}$ and each edge of G occurs precisely once in the sequence $\ldots, x_{-2}x_{-1}, x_{-1}x_0, x_0x_1, x_1x_2, \cdots$. In 1936, Erdős, Grünwald and Weiszfeld proved the following analogue of Theorem 12.

Theorem 14 *Let $G = (V, E)$ be a connected multigraph with E infinite. Then G has a two-way infinite Euler trail if and only if the following conditions are satisfied:*

(i) *E is countable,*

(ii) *every degree is even or infinite,*

(iii) *for every subgraph $G' \subset G$, $G' = (V, E')$, with E' finite, the graph $G - E'$ has at most two infinite components; furthermore, if $d_{G'}(x)$ is even for every $x \in V$, then $G - E'$ has precisely one infinite component.*

Although the proof is not too difficult, we do not give it here. The reader is encouraged to do Exercises 54–56, which are related to this result.

I.4 Planar Graphs

The graph of the corridors of an exhibition is a *planar graph*: it can be drawn in the plane in such a way that no two edges intersect. Putting it a little more rigorously, it is possible to represent it by a drawing in the plane in which the vertices correspond to distinct points and the edges to simple Jordan curves connecting the points of its endvertices. In this drawing every two curves are either disjoint or meet only at a common endpoint. The above representation of a graph is said to be a *plane graph*.

There is a simple way of associating a topological space with a graph, which leads to another definition of planarity, trivially equivalent to the one given above.

Let $p_1, p_2, \ldots$ be distinct points in $\mathbb{R}^3$, the 3-dimensional Euclidean space, such that every plane in $\mathbb{R}^3$ contains at most 3 of these points. Write (p_i, p_j) for the straight line segment with endpoints p_i and p_j (open or closed, as you like). Given a graph $G = (V, E)$, $V = (x_1, x_2, \ldots, x_n)$, the topological space

$$R(G) = \bigcup\{(p_i, p_j) : x_i x_j \in E\} \cup \bigcup_1^n \{p_i\} \subset \mathbb{R}^3$$

is said to be a *realization* of G. A graph G is *planar* if $R(G)$ is homeomorphic to a subset of $\mathbb{R}^2$, the plane.

Let us make some more remarks in connection with $R(G)$. A graph H is said to be a *subdivision* of a graph G, or a *topological G graph* if H is obtained from G by subdividing some of the edges, that is, by replacing the edges by paths having at most their endvertices in common. We shall write TG for a topological G graph. Thus TG denotes *any* member of a rather large family of graphs; for example, TK_3 is an arbitrary cycle, and TC_8 is an arbitrary cycle of length at least 8. It is clear that for any graph G the spaces $R(G)$ and $R(TG)$ are homeomorphic. We shall say that a graph G is *homeomorphic* to a graph H if $R(G)$ is homeomorphic to $R(H)$ or, equivalently, G and H have isomorphic subdivisions.

At first sight one may think that in the study of planar graphs one might run into topological difficulties. This is certainly not the case. It is easily seen that the Jordan curves corresponding to the edges can be assumed to be polygons. More precisely, every plane graph is homotopic to a plane graph representing the same graph, in which the Jordan curves are piecewise linear. Indeed, given a plane graph, let $\delta > 0$ be less than half the minimal distance between two vertices. For each vertex a place a closed disc D_a of radius δ about a. Denote by J_α the curve corresponding to an edge $\alpha = ab$ and let a_α be the last point of J_α in D_α when going from a to b. Denote by J'_α the part of J_α from a_α to b_α. Let $\varepsilon > 0$ be such that if $\alpha \neq \beta$ then J'_α and J'_β are at a distance greater than 3ε. By the uniform continuity of a Jordan curve, each J'_α can be approximated within ε by a polygon J''_α from a_α to b_α. To get the required piecewise linear representation of the original graph simply replace each J_α by the polygon obtained from J''_α by extending it in both directions by the segments aa_α and $b_\alpha b$ (see Fig. I.18).

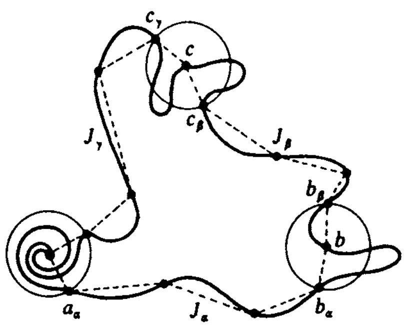


FIGURE I.18. Constructing a piecewise linear representation.

A less pedestrian argument shows that every planar graph has a *straight line representation*: it can be drawn in the plane in such a way that the edges are actually straight line segments (Exercise 63^+).

If we omit the vertices and edges of a plane graph G from the plane, the remainder falls into connected components, called *faces*. Clearly, each plane graph has exactly one unbounded face. The *boundary* of a face is the set of edges in its closure. Since a cycle (that is a simple closed polygon) separates the points of the plane into two components, each edge of a cycle is in the boundary of two faces. A plane graph together with the set of faces it determines is called a *plane map*. The faces of a plane map are usually called *countries*. Two countries are *neighbouring* if their boundaries have an edge in common.

If we draw the graph of a convex polyhedron in the plane, then the faces of the polyhedron clearly correspond to the faces of the plane graph. This leads us to another contribution of Leonhard Euler to graph theory, namely *Euler's polyhedron theorem* or simply *Euler's formula.*

Theorem 15 *If a connected plane graph G has n vertices, m edges, and f faces, then*

$$n - m + f = 2.$$

Proof. Let us apply induction on the number of faces. If $f = 1$, then G does not contain a cycle, so it is a tree, and the result holds by Corollary 8.

Suppose now that $f > 1$ and the result holds for smaller values of f. Let ab be an edge in a cycle of G. Since a cycle separates the plane, the edge ab is in the boundary of two faces, say S and T. Omitting ab, in the new plane graph G' the faces S and T join up to form a new face, while all other faces of G remain unchanged. Thus if n', m' and f' are the parameters of G', then $n' = n$, $m' = m-1$, and $f' = f - 1$. Hence $n - m + f = n' - m' + f' = 2$. □

Let G be a connected plane graph with n vertices, m edges, and f faces; furthermore, denote by f_i the number of faces having exactly i edges in their boundaries. Clearly,

$$\sum_i f_i = f, \tag{4}$$

and if G has no bridge, then

$$\sum_i i f_i = 2m, \tag{5}$$

since every edge is in the boundary of two faces. Relations (4), (5), and Euler's formula give an upper bound for the number of edges of a planar graph of order n. This bound can be improved if the *girth* of the graph, that is the number of edges in a shortest cycle, is large. (The girth of an acyclic graph is defined to be ∞.)

Theorem 16 *A planar graph of order $n \geq 3$ has at most $3n - 6$ edges. Furthermore, a planar graph of order n and girth at least g, $3 \leq g < \infty$, has size at* *most*

$$\max\left\{\frac{g}{g-2}(n-2),\ n-1\right\}.$$

Proof. The first assertion is the case $g = 3$ of the second, so it suffices to prove the second assertion. Let G be a planar graph of order n, size m, and girth at least g. If $n \leq g-1$, then G is acyclic, so $m \leq n-1$. Assume now that $n \geq g$ and the assertion holds for smaller values of n. We may assume without loss of generality that G is connected. If ab is a bridge then $G - ab$ is the union of two vertex disjoint subgraphs, say G_1 and G_2. Putting $n_i = |G_i|$, $m_i = e(G_i)$, $i = 1, 2$, by induction we find that

$$\begin{aligned} m = m_1 + m_2 + 1 &\leq \max\left\{\frac{g}{g-2}(n_1-2), n_1-1\right\} \\ &\quad + \max\left\{\frac{g}{g-2}(n_2-2), n_2-1\right\} + 1 \\ &\leq \max\left\{\frac{g}{g-2}(n-2), n-1\right\}. \end{aligned}$$

On the other hand, if G is bridgeless, (4) and (5) imply that

$$2m = \sum_i if_i = \sum_{i\geq g} if_i \geq g\sum_i f_i = gf.$$

Hence, by Euler's formula,

$$m + 2 = n + f \leq n + \frac{2}{g}m,$$

and so

$$m \leq \frac{g}{g-2}(n-2). \qquad \square$$

Theorem 16 can often be used to show that certain graphs are nonplanar. Thus K_5, the complete graph order 5, is nonplanar since $e(K_5) = 10 > 3(5-2)$. Another nonplanar graph is $K_{3,3}$, the complete 3 by 3 bipartite graph, also called the *Thomsen graph*, since its girth is 4 and $e(K_{3,3}) = 9 > (4/(4-2))(6-2)$. The nonplanarity of $K_{3,3}$ implies that it is impossible to join each of 3 houses to each of 3 wells by non-crossing paths, as demanded by a well-known puzzle (see Fig. I.19).

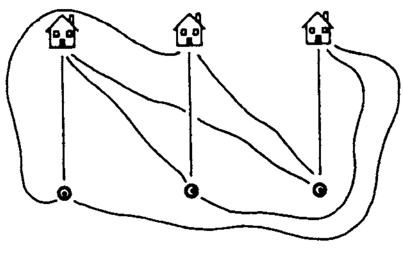


FIGURE I.19. The Thomsen graph: three houses and three wells.

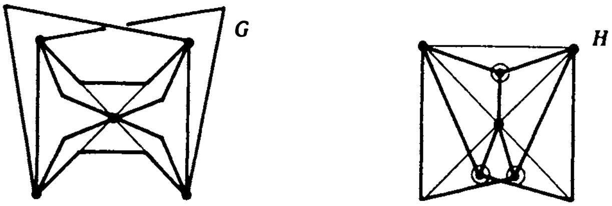


FIGURE I.20. G contains a TK_5 and H contains a $TK_{3,3}$.

If a graph G is nonplanar, then so is every topological G graph and every graph containing a topological G graph. Thus the graphs in Fig. I.20 are nonplanar, since they contain TK_5 and $TK_{3,3}$, respectively.

It is somewhat surprising that the converse of the trivial remarks above is also true: this beautiful result was proved by Kuratowski in 1930.

Theorem 17 *A graph is planar iff it does not contain a subdivision of K_5 or $K_{3,3}$.* □

A variant of Theorem 17 characterizes planar graphs in terms of *forbidden minors*, rather than forbidden topological subgraphs. At first sight, the concept of a minor may seem a little artificial, but it is, in fact, the right notion related to drawing graphs on surfaces.

Given an edge xy of a graph G, the graph G/xy is obtained from G by *contracting* the edge xy; that is, to get G/xy we identify the vertices x and y and remove all resulting loops and duplicate edges. A graph G obtained by a sequence of edge-contractions is said to be a *contraction* of G. A graph H is a *minor* of G, written $G \succ H$ or $H \prec G$, if it is a subgraph of a graph obtained from G by a sequence of edge-contractions (see Fig. I.21). It is easily checked that if $V(H) = \{y_1, y_2, \ldots, y_k\}$ then $H \prec G$ if and only if G has vertex-disjoint connected subgraphs $G_1, G_2, \ldots, G_k$ such that if $y_i y_j \in E(H)$, then G has an edge from G_i to G_j (see Exercise 88^-).

In 1937, Wagner proved the following analogue of Kuratowski's theorem.

Theorem 18 *A graph is planar iff it contains neither K_5 nor $K_{3,3}$ as a minor.* □

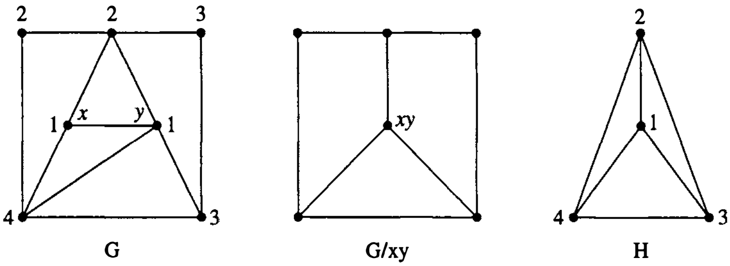


FIGURE I.21. A graph G, its contraction G/xy and a minor H.

It is easy to see that Theorems 17 and 18 are equivalent. Indeed, if $G \supset TH$, then, rather trivially, $G \succ H$. In fact, if H has maximal degree at most 3, then $G \supset TH$ iff $G \succ H$. In particular, $G \supset TK_{3,3}$ if and only if $G \succ H$. Also, if $G \succ K_5$ then either $G \supset TK_5$ or $G \supset TK_{3,3}$. The reader is encouraged to fill in the details (see Exercise 91).

I.5 An Application of Euler Trails to Algebra

To conclude this chapter we shall show that even simple notions like the ones presented so far may be of use in proving important results. The result we are going to prove is the fundamental theorem of Amitsur and Levitzki on polynomial identities. The *commutator* of two elements a and b of a ring S is $[a, b] = ab - ba$. Similarly, if $a_i \in S$, $1 \le i \le k$, we write

$$[a_1, a_2, \dots, a_k] = \sum_\sigma \operatorname{sgn}(\sigma) a_{\sigma 1} a_{\sigma 2} \cdots a_{\sigma k},$$

where the summation is over all permutations σ of the integers $1, 2, \dots, k$, and $\operatorname{sgn}(\sigma)$ is the sign of σ. For example, $[a_1, a_2, a_3] = a_1 a_2 a_3 - a_1 a_3 a_2 + a_3 a_1 a_2 - a_3 a_2 a_1 + a_2 a_3 a_1 - a_2 a_1 a_3$. If $[a_1, a_2, \dots, a_k] = 0$ for all $a_i \in S$, $1 \le i \le k$, then S is said to satisfy the kth *polynomial identity*. The theorem of Amitsur and Levitzki states that the ring $M_k(R)$ of k by k matrices with entries in a commutative ring R satisfies the $(2k)$th polynomial identity.

Theorem 19 *Let R be a commutative ring and let the matrices $A_1, A_2, \dots, A_{2k}$ be in $M_k(R)$. Then $[A_1, A_2, \dots, A_{2k}] = 0$.*

Proof. We shall deduce the result from a lemma about Euler trails in directed multigraphs. Let $\vec{G}$ be a directed multigraph of order n with edges $e_1, e_2, \dots, e_m$. Thus to each edge e_i we associate an ordered pair of not necessarily distinct vertices: the initial vertex of e_i and the terminal vertex of e_i. Every (directed) Euler trail P is readily identified with a permutation of $\{1, 2, \dots, m\}$; define $\varepsilon(P)$ to be the sign of this permutation. Given not necessarily distinct vertices x, y of $\vec{G}$, put $\varepsilon(\vec{G}; x, y) = \sum_P \varepsilon(P)$, where the summation is over all Euler trails from x to y.

Lemma 20 *If $m \ge 2n$ then $\varepsilon(\vec{G}; x, y) = 0$.*

Before proving this lemma, let us see how it implies Theorem 19. Write $E_{ij} \in M_n(R)$ for the matrix whose only non-zero entry is a 1 in the ith row and jth column. Since $[A_1, A_2, \dots, A_{2n}]$ is R-linear in each variable and $\{E_{ij} : 1 \le i, j \le n\}$ is a basis of $M_n(R)$ as an R-module, it suffices to prove Theorem 19 when $A_k = E_{i_k j_k}$ for each k. Assuming that this is the case, let $\vec{G}$ be the *directed multigraph* with vertex set $\{1, 2, \dots, n\}$ whose set of directed edges is $\{i_1 j_1, i_2 j_2, \dots, i_{2n} j_{2n}\}$. By the definition of matrix multiplication, a product $A_{\sigma 1} A_{\sigma 2} \cdots A_{\sigma 2n}$ is E_{ij} if the corresponding sequence of edges is a (directed) Euler trail from i to j and otherwise it is 0. Hence $[A_1, A_2, \dots, A_{2n}] = \sum_{i,j} \varepsilon(\vec{G}; i, j)E_{ij}$. By Lemma 20 each summand is 0, so the sum is also 0, and Theorem 19 is proved. □

Proof of Lemma 20. We may clearly assume that $\vec{G}$ has no isolated vertices. Let $\vec{G}'$ be obtained from $\vec{G}$ by adding to it a vertex x', a path of length $m+1-2n$ from x' to x, and an edge from y to x' (see Fig. I.22). Then $\vec{G}'$ has order $n+(m+1-2n) = m+1-n$ and size $m+m+1-2n+1 = 2(m+1-n)$. Furthermore, it is easily checked that $|\varepsilon(\vec{G}; x, y)| = |\varepsilon(\vec{G}'; x', x')|$. Hence it suffices to prove the theorem when $m = 2n$ and $x = y$.

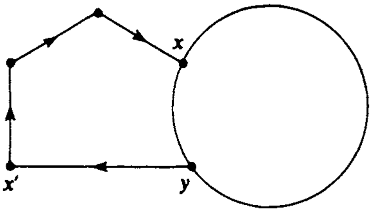


FIGURE I.22. The construction of $\vec{G}'$.

Given a vertex z, recall that $d^+(z)$ is the number of edges starting at z and recall that $d^-(z)$ is the number of edges ending at z. Call $d(z) = d^+(z) + d^-(z)$ the *degree* of z and $f(z) = d^+(z) - d^-(z)$ the *flux* at z. We may assume that $\vec{G}$ contains an Euler circuit (an Euler trail from x to x; otherwise, there is nothing to prove. In this case, each vertex has 0 flux, even degree, and the degree is at least 2. Furthermore, we may assume that there is no double edge (and so no double loop), for otherwise the assertion is trivial.

In order to prove the theorem in the case $m = 2n$ and $x = y$ we apply induction on n. The case $n = 1$ being trivial, we turn to the induction step. We shall distinguish three cases.

(i) *There is a vertex $b \neq x$ of degree* 2; say $e_{m-1} = ab$ ends at b and $e_m = bc$ starts at b. If $a = c$, the assertion follows by applying the induction hypothesis to $\vec{G} - b$. If $a \neq c$, then without loss of generality $x \neq c$. Let $e_1 = cc_1$, $e_2 = cc_2, \dots, e_t = cc_t$ be the edges starting at c. For each i, $1 \le i \le t$, construct a graph $\vec{G}_i$ from $\vec{G} - b$ by omitting e_i and adding $e_i' = ac_i$ (see Fig. I.23). Then $\varepsilon(\vec{G}; x, x) = \sum_{i=1}^t \varepsilon(\vec{G}_i; x, x) = 0$.

(ii) *There is a loop at a vertex $b \neq x$ of degree* 4. Let e_m be the loop at b and let $e_{m-2} = ab$ and $e_{m-1} = bc$ be the other edges at b. Let $\vec{G}_0$ be obtained from

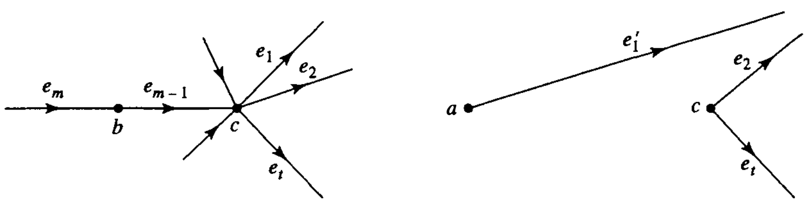


FIGURE I.23. The construction of $\vec{G}_1$.

$\vec{G} - b$ by adding to it an edge $e'_{m-2} = ac$. Then $\varepsilon(\vec{G}; x, x) = \varepsilon(\vec{G}_0; x, x)$, which is 0 by the induction hypothesis.

(iii) *The cases* (i) *and* (ii) *do not apply.* Since $m = 2n = \frac{1}{2}\sum_1^n d_i$ and each vertex distinct from x has degree at least 4, either each vertex has degree 4 or else $d(x) = 2$ and there is a vertex of degree 6 and all other vertices have degree 4. It is easily checked (Exercise 93) that there are two adjacent vertices of degree 4, say a and b, since otherwise (ii) holds. Now we shall apply our fourth and final graph transformation. This is more complicated than the previous ones, since we shall construct two pairs of essentially different graphs from $\vec{G}$: the graphs $\vec{G}_1$, $\vec{G}_2$, $\vec{H}_6$, and $\vec{H}_7$ shown in Fig. I.24. Each Euler trail from x to x in $\vec{G}$ is transformed to an Euler trail in exactly one of $\vec{G}_1$ and $\vec{G}_2$. However, the graphs $\vec{G}_1$ and $\vec{G}_2$ contain some spurious Euler trails: Euler trails that do not come from Euler trails in $\vec{G}$. As these spurious Euler trails are Euler trails in exactly one of $\vec{H}_6$ and $\vec{H}_7$ and they exhaust all the Euler trails of $\vec{H}_6$ and $\vec{H}_7$, we find that

$$\varepsilon\left(\vec{G}; x, x\right) = \sum_{i=1}^{2} \varepsilon\left(\vec{G}_i; x, x\right) - \sum_{i=6}^{7} \varepsilon\left(\vec{H}_i; x, x\right).$$

The first two terms are 0 because of (i), and the second two terms are 0 because of (ii), so $\varepsilon(\vec{G}; x, x) = 0$, completing the proof of Lemma 20. □

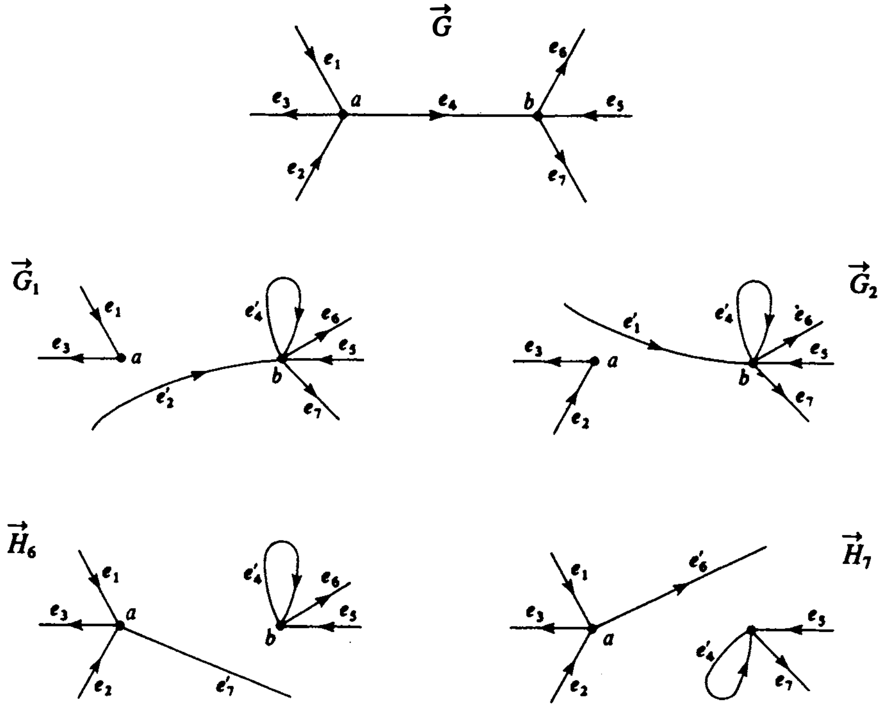


FIGURE I.24. The graphs $\vec{G}$, $\vec{G}_1$, $\vec{G}_2$, $\vec{H}_6$ and $\vec{H}_7$.

The operations $\vec{G} \mapsto \vec{G}_1, \vec{G}_2, \vec{H}_6, \vec{H}_7$ are somewhat similar to various graph operations used to construct graph polynomials; a simple example is that of the chromatic polynomial, to be studied in Chapter V.

I.6 Exercises

1. Prove that either a graph or its complement is connected.

2. (i) Show that every graph contains two vertices of equal degree.
   (ii) Determine all graphs with one pair of vertices of equal degree.

3.$^-$ Let a be a vertex of a connected graph G. Show that G is bipartite if and only if $d(a, b) \neq d(a, c)$ for every edge bc.

4. Prove that the bound in Mantel's theorem (Theorem 2) is best possible: even more, for every $n \geq 1$, the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ is the unique triangle-free graph of order n and maximal size.

5. Show that the following conditions are equivalent for a graph G of size at least 2:
   (i) G is connected and has no cutvertex,
   (ii) any two vertices are on a cycle,
   (iii) any two edges are on a cycle,
   (iv) for any three vertices x, y and z, there is an x-z path containing y.

6. Let G be a graph of order n. Prove the equivalence of the following assertions.
   (i) G is a tree.
   (ii) G is connected and has at most $n - 1$ edges.
   (iii) G is acyclic and has at least $n - 1$ edges.
   (iv) $G = K_n$ for $n = 1, 2$, and if $n \geq 3$, then $G \neq K_n$ and the addition of an edge to G produces exactly one new cycle.

7.$^-$ Show that every connected graph G of order at least two contains vertices x and y such that both $G - x$ and $G - y$ are connected.

8. In the puzzle of *jealous husbands*, three husbands and their wives wish to cross a river. They have only one small boat, which can take two persons at a time. No husband ever allows his wife to be in the company of other men unless he is also present. Draw the graph of permissible distributions of people and advise the travelers how they could cross the river.

9. In the puzzle of the *man* and his *dog*, *goat*, and (large) *cabbage*, a man wishes to cross the river with his dog, goat, and (large) cabbage, but the small boat he has access to can take only one of his possessions besides himself. To complicate matters, for obvious reasons, the goat cannot be left in the company of the dog or the cabbage, unless the man is also present. Draw the very simple bipartite graph of permissible situations, and advise the man how he should proceed.

10. Show that in an infinite graph G with countably many edges there exists a set of cycles and two-way infinite paths such that each edge of G belongs to

exactly one of these iff for every $X \subset V(G)$ either there are infinitely many edges joining X to $V(G) - X$, or else $e(X, V(G) - X)$ is even.

11. Show that every graph G has a bipartition $V(G) = U \cup W$ such that $e(U, W) \geq \frac{1}{2}e(G)$. Show also that if G is cubic of order n, then we may demand that $e(U, W) \geq n$.

12. Show that for every graph $G = (V, E)$ there is a partition $V = V_1 \cup V_2$ such that
$$e(G[V_1]) + e(G[V_2]) \leq \frac{1}{2}e(G).$$
Show also that one may also demand that each V_i span at most a third of the edges, that is, $e(G[V_i]) \leq \frac{1}{3}e(G)$ for $i = 1, 2$.

13. Show that every graph with average degree d contains a subgraph of minimal degree at least $d/2$.

14. Show that every graph with average degree d contains a bipartite subgraph of average degree at least $d/2$.

15. Show that every graph of order n and average degree d contains a subgraph of order greater than $n/2$ and maximal degree at most d.

16. Let G be a graph of average degree $d > 0$. Show that for some vertex x of G the average of the degrees of the neighbours of x is at least d. What if we replace "at least" by "at most"?

17. Show that $d_1 \leq d_2 \leq \cdots \leq d_n$ is the degree sequence of a tree iff $d_1 \geq 1$ and $\sum_1^n d_i = 2n - 2$.

18. Show that every integer sequence $d_1 \leq d_2 \leq \cdots \leq d_n$ with $d_1 \geq 1$ and $\sum_1^n d_i = 2n - 2k, k \geq 1$, is the degree sequence of a forest with k components.

19. Characterize the degree sequences of forests!

20.$^-$ Show that, up to isomorphism, there is a unique graph with degree sequence $2, 2, \ldots, 2, 1, 1$.

21. Show that for every degree sequence $(d_i)_1^n$, $1 \leq d_1 \leq \cdots \leq d_n \leq n - 1$, there are at most $(n-2)!$ trees on $\{x_1, \ldots, x_n\}$, with $d(x_i) = d_i$ for every i. Show also that, for every n there is a unique degree sequence on which this upper bound is attained.

22.$^-$ Show that there is a unique sequence $(d_i)_1^n$, $1 \leq d_1 \leq \cdots \leq d_n$, for which there is only one tree on $\{x_1, \ldots, x_n\}$ with $d(x_i) = d_i$ for every i.

23. Show that if n is large enough, then for every sequence $1 \leq d_1 \leq \cdots \leq d_n \leq n - 2$, with $\sum_{i=1}^n d_i = 2n - 2$, there are at least $n - 2$ trees on $\{x_1, \ldots, x_n\}$ with $d(x_i) = d_i$ for every i.

24. Prove that a regular bipartite graph of degree at least 2 does not have a bridge.

25. Let $V(G) = \bigcup_{i=1}^k V_i$ be a partition of the vertex set of a connected graph G into $k \geq 2$ nonempty subsets such that each $G[V_i]$ is connected. Prove

that there are indices $1 \le i < j \le k$ such that both $G - V_i$ and $G - V_j$ are connected.

26.$^-$ Let G be a connected graph of order n and let $1 \le k \le n$. Show that G contains a connected subgraph of order k.

27. Let $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ be a family of $n \ge 1$ distinct subsets of a set X with n elements. Define a graph G with vertex set $\mathcal{A}$ in which $A_i A_j$ is an edge iff there exists an $x \in X$ such that $A_i \Delta A_j = \{x\}$. Label the edge $A_i A_j$ with x. For $H \subset G$ let Lab(H) be the set of labels used for edges of H. Prove that there is a forest $F \subset G$ such that Lab(F) = Lab(G).

28. (Exercise 27 contd.) Deduce that there is an element $x \in X$ such that the sets $A_1 - \{x\}, A_2 - \{x\}, \ldots, A_n - \{x\}$ are all distinct. Show that this need not hold for any n if $|\mathcal{A}| = n + 1$.

29.$^+$ (Exercise 27 contd.) Describe *all* families $\mathcal{A} = \{A_1, A_2, \ldots, A_{n+1}\}$ of $n+1$ distinct subsets of X, $|X| = n$, such that for every $x \in X$ there are i, j, $1 \le i < j \le n + 1$, with $A_i - \{x\} = A_j - \{x\}$.

30. A *tournament* is a complete *oriented* graph, that is, a directed graph in which for any two distinct vertices x and y either there is an edge from x to y or there is an edge from y to x, but not both. Prove that every tournament contains a (directed) Hamilton path.

31.$^-$ Prove that the radius and diameter of a graph satisfy the inequalities

$$\mathrm{rad}G \le \mathrm{diam}G \le 2\mathrm{rad}G,$$

and both inequalities are best possible.

32.$^-$ Given $d \ge 1$, determine

$$\max_{\mathrm{diam}G=d} \min\{\mathrm{diam}T : T \text{ is a spanning tree of } G\}.$$

33. Let a and b be vertices of a tree T at maximal distance $d(a, b) = 2r$, and let c be the vertex on the unique $a - b$ path at distance r from a and b. Show that c is the unique vertex of T with $d(c, x) \le r$ for every $x \in T$.

34. Deduce from the proof of Theorem 1 the following strengthening of the assertion. Let G be a triangle-free graph of order n. Then $e(G) \le \lfloor n^2/4 \rfloor$, with equality iff G is a complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

35. Denote by $\alpha(G)$ the maximal cardinality of a set of independent vertices in G. Prove that if G does not contain a triangle, then $\Delta(G) \le \alpha(G)$ and deduce that $e(G) \le \frac{1}{2} n\alpha(G)$, where $n = |G|$.

36.$^-$ Show that if for every vertex z of a directed graph there is an edge starting at z (that is, $d^+(z) > 0$) then the graph contains a (directed) cycle.

37. A *grading* of a directed graph $\vec{G} = (V, \vec{E})$ is a partitioning of V into sets $V_1, V_2, \ldots, V_k$ such that if $\vec{xy} \in \vec{E}$, then $x \in V_i$ and $y \in V_{i+1}$ for some i.

Given a directed graph G and a (nondirected) path $P = x_0x_1 \cdots x_s$, denote by $v(x_0, x_s; P)$ the number of edges $\overrightarrow{x_ix_{i+1}}$ minus the number of edges $\overrightarrow{x_{j+1}x_j}$. Prove that G has a grading iff $v(x_0, x_s; P)$ is independent of P for every pair of vertices x_0, x_s.

38. Is it true that, for every $n \geq 2$, the complete graph K_n is the union of cycles $C_3, C_4, \ldots, C_{n-1}$, an edge, and a path of length 2?

39. Show that a complete graph K_n has a decomposition into edge-disjoint paths of length 2 if and only if $n \equiv 0$ or 1 (mod 4).

40. Show that for $n \geq 2$ the complete graph K_n is the union of paths of distinct lengths.

41. A *Steiner triple system of order n* is a decomposition of a complete graph K_n into edge disjoint triangles. Equivalently, a Steiner triple system on a set X is a set system $\mathcal{A} \subset X^{(3)}$ such that every pair $e \in X^{(2)}$ is contained in precisely one triple $A \in \mathcal{A}$; the number of elements of the ground set X is the order of $\mathcal{A}$. Show that if there is a Steiner triple system of order n then $n \equiv 1$ or 3 (mod 6).

42. Show that up to relabelling, there is a unique Steiner triple system of order 7, namely the Fano plane in Fig. I.6.

43. Let $\mathcal{A} \subset X^{(3)}$ and $\mathcal{B} \subset Y^{(3)}$ be Steiner triple systems. Let $\mathcal{C} \subset (X \times Y)^{(3)}$ consist of all triples of the form
(1) $\{(x_1, y), (x_2, y), (x_3, y)\}$ with $A = \{x_1, x_2, x_3\} \in \mathcal{A}$ and $y \in Y$,
(2) $\{(x, y_1), (x, y_2), (x, y_3)\}$ with $B = \{y_1, y_2, y_3\} \in \mathcal{B}$ and $x \in X$,
(3) $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ with $A = \{x_1, x_2, x_3\} \in \mathcal{A}$ and $B = \{y_1, y_2, y_3\} \in \mathcal{B}$.
[Note that in (3) each pair of triples (A, B) gives rise to precisely six different triples.] Show that $\mathcal{C}$, the *product* of $\mathcal{A}$ and $\mathcal{B}$, is a Steiner triple system on $X \times Y$. Deduce that there are infinitely many Steiner triple systems of order congruent to 1 (mod 6), and likewise for 3 (mod 6).

44. Complete the proof of Theorem 10 by showing in detail that both the second and third methods construct an economical spanning tree.

45. Show how the fourth method in Theorem 10 can be applied to find an economical spanning tree even if several edges have the same cost (cf. Fig. I.8).

46. Show that every economical spanning tree can be constructed by each of the first three methods.

47. Deduce from Theorem 12 that a graph contains an Euler circuit iff all but at most one of its components are isolated vertices and each vertex has even degree. State and prove an analogous statement about the existence of an Euler trail from x to y.

48. Show that every multigraph with $2\ell \geq 2$ vertices of odd degrees is the edge-disjoint union of ℓ trails.

49. Fleury gave the following algorithm for finding an Euler circuit $x_1x_2\cdots x_n$ in a graph G. Pick x_1 arbitrarily. Having chosen $x_1, x_2, \ldots, x_k$, put $G_k = G - \{x_1x_2, x_2x_3, \ldots, x_{k-1}x_k\}$. If every edge incident with x_k in G_k is a bridge (in particular, if x_k is an isolated vertex of G_k), then terminate the algorithm. Otherwise, let x_{k+1} be a neighbour of x_k in G_k such that x_kx_{k+1} is not a bridge of G_k.
Prove that if G has an Euler circuit, then the trail $x_1x_2\cdots x_\ell$ constructed by the algorithm is an Euler circuit.

50. Recall that a graph G is *randomly Eulerian from a vertex x* if any maximal trail starting at x is an Euler circuit. (If $T = xx_1\cdots x_\ell$, then T is a maximal trail starting at x iff x_ℓ is an isolated vertex in $G - E(T)$.) Prove that a nonempty graph G is randomly Eulerian from x iff G has an Euler circuit and x is contained in every cycle of G.

51. Let F be a forest. Add a vertex x to F and join x to each vertex of odd degree in F. Prove that the graph obtained in this way is randomly Eulerian from x, and every graph randomly Eulerian from x can be obtained in this way.

52. Prove that a graph G is randomly Eulerian from each of two vertices x and y iff G is the union of an even number of x–y paths, any two of which have only x and y in common.

53.$^+$ A *one-way infinite Euler trail* in an infinite multigraph $G = (V, E)$ is an infinite sequence $x_1, e_1, x_2, e_2, \ldots$ such that $x_1, x_2, \ldots \in V$, e_i is the edge x_ix_{i+1}, $e_i \neq e_j$ if $i \neq j$, and $E = \{e_1, e_2, \ldots\}$.
Let G be a connected infinite multigraph with countably many edges and with one vertex of odd degree. (Thus $d(x_1)$ is odd for some vertex x_1; for every other vertex x either $d(x)$ is infinite or it is finite and even.) Show that G has a one-way infinite Euler trail if, and only if, for every finite set $E_0 \subset E$, the graph $G - E_0$ has only one infinite component.

54. Show the necessity of the conditions in Theorem 14.

55. Show that condition (iii) in Theorem 14 can be replaced by the following condition:
(iii') there is a vertex x such that if T is a finite trail starting at x then $G - E(T)$ has at most two infinite components; furthermore, if T is a closed trail (circuit), then $G - E(T)$ has precisely one infinite component.

56. Deduce Theorem 14 from the results in the previous two exercises.

57. Show that for every $n \geq 1$ the graph of the lattice $\mathbb{Z}^n$ has a two-way infinite Euler trail.

58.$^+$ Each of $n \geq 4$ elderly professors know some item of gossip not known to the others. They communicate by telephone and in each conversation they part with all the gossip they know. Show that $2n - 4$ calls are *needed* before each of them knows everything.

59. How would you define the number of sides of a face so that formula (4) continues to hold for graphs with bridges? Rewrite the proof of Theorem 15 accordingly.

60. Let G be a planar graph of order at least 3, with degree sequence $(d_i)_1^n$. Show that

$$\sum_{d_i \le 6}(6-d_i) \ge \sum_{i=1}^{n}(6-d_i) \ge 12.$$

Deduce that if $\delta(G) \ge 5$, then G has at least 12 vertices of degree 5, and if $\delta(G) \ge 4$ then G has at least 6 vertices of degree at most 5.

61.$^+$ Let $(d_i)_1^n$ be the degree sequence of a planar graph of order $n \ge 3$. Prove that for $k \ge 3$ we have

$$\sum_{i=1}^{k} d_i \le 2n + 6k - 16.$$

62.$^-$ Make use of the nonplanarity of K_5 to show that every face of a maximal planar graph is a triangle.

63.$^+$ Prove that every planar graph has a drawing in the plane in which every edge is a straight line segment. [*Hint.* Apply induction on the order of a maximal planar graph by omitting a suitable vertex.]

64. A plane drawing of an infinite graph is defined as that of a finite graph with the additional condition that each point has a neighbourhood containing at most one vertex and meeting only edges incident with that vertex.
Show that Kuratowski's theorem does not hold for infinite graphs; that is, construct an infinite nonplanar graph without TK_5 and $TK_{3,3}$.
Is there an infinite nonplanar graph without a TK_4?

65. Show that there is no bipartite cubic planar graph of order 10, but for every $n \ge 4$, $n \ne 5$, there is a connected bipartite cubic planar graph of order $2n$ (see Fig. I.25).

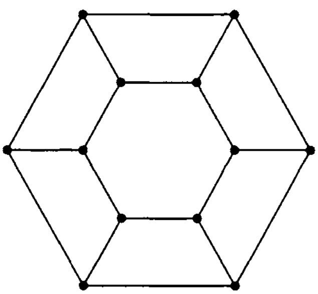


FIGURE I.25. A bipartite cubic planar graph of order 12.

66. There are n convex sets in the plane such that the boundaries of any two of them have at most two common points. Show that the boundary of their union consists of at most $6n - 12$ connected arcs of the boundaries of the sets.

67. Let $(d_i)_1^n$ be the degree sequence of a planar graph G.
(i) By making use of an upper bound for $\sum_1^n d_i$, show that if $\delta(G) \geq 4$ then

$$\sum_1^n d_i^2 < 2(n+3)^2 - 62.$$

(ii) Prove by induction on n that if $n \geq 4$ then

$$\sum_1^n d_i^2 \leq 2(n+3)^2 - 62.$$

Show that equality can hold for every $n \geq 4$.

68.$^+$ Determine the maximum of $\sum_1^n d_i^2$, where $(d_i)_1^n$ is the degree sequence of a planar graph of girth at least 4 (that is without triangles). What is the maximum if the girth is at least $g > 4$?

69. Let G be a graph of order $n \geq 4$ such that every graph obtained from G by deleting a vertex is regular (i.e.; all vertices have the same degree). Show that G is either the complete graph K_n or the empty graph E_n.

70. Show that every graph of maximal degree at most r is an induced subgraph of an r-regular graph: if $\Delta(G) \leq r$, then there is an r-regular graph H and a set $W \subset V(H)$ such that $G = H - W$. Show also that we can always find a pair (H, W) with

$$|W| \leq \max\left\{r - \delta(G), \left\lceil \sum_{x \in G} (r - d(x))/r \right\rceil\right\} + 1.$$

71. Let G be a graph with $\delta(G) \geq 2$. Show that there is a connected graph H with the same degree sequence, that is with $V(H) = V(G)$ and $d_H(x) = d_G(x)$ for all $x \in V(H)$.

72. Show that for every graph $G = (V, E)$ and natural number k, there is a partition $V = \cup_{i=1}^k V_i$ such that if $x \in V_i$ and $i \neq j$, then x is joined to at least as many vertices in V_j as in V_i.

73. Let G be a planar graph, with the edges coloured red and blue. Show that there is a vertex x such that going round the edges incident with x in the clockwise direction, say, we encounter no more than two changes of colour.

74. Suppose we have $n \geq 3$ great circles of the sphere $S^2 \subset \mathbb{R}^3$, not all through the same point, coloured red and blue. Deduce from the result in the previous exercise that there is a point $x \in S^2$ such that there are at least 2 great circles through it and they all have the same colour.

75.$^+$ Let $P_1, P_2, \ldots, P_n$ be points in the plane, not all on a line, coloured red and blue. Prove that there is a line through two of these points such that all points on this line have the same colour.

76. Let $x_1x_2 \cdots x_n$ be a regular n-gon, with $n \geq 2k+1$, $k \geq 1$. Show that if $kn+1$ of the pairs (x_i, x_j) are joined by straight line segments, then $k+1$ of them are pairwise disjoint. Does this hold for kn pairs?

77. Prove that in the game of Hex (played on an $n \times n$ board) precisely one of the players wins.

78. Let $T_1, \ldots, T_k$ be subtrees of a tree T such that for all $1 \leq i < j \leq k$ the trees T_i and T_j have a vertex in common. Show that T has a vertex that is in all the T_i.

79. Given a graph G and an equivalence relation R on $V(G)$, let G/R be the graph whose vertices are the equivalence classes V_i of R, and $V_iV_j \in E(G/R)$ if G contains a $(V_i - V_j)$-edge. Show that for every connected graph H there is a tree T and an equivalence relation R on $V(T)$ such that $H \cong T/R$ and $e(H) = e(T)$.

80. Show that every connected graph with an even number of edges has an orientation in which every vertex x has even outdegree $d^+(x)$.

81. Let G be a connected *infinite* graph and let $\varepsilon : V_F \to \{0, 1\}$, where V_F is the set of vertices of finite degree. Show that G has an orientation such that $d^+(x) \equiv \varepsilon(x) \pmod 2$ for every $x \in V_F$.

82. Show that every multigraph has an orientation in which the out degree and in degree of every vertex differ by at most 1.

83. Show that for every graph G there is a set $W \subset V(G)$ such that every vertex in W has an even number of neighbours in W and every vertex in $V - W$ has an odd number of neighbours in W.

84. Determine all graphs of order n with a *loop* at some of the vertices such that no two vertices have the same degree. [A loop at a vertex x adds 1 to the degree of x.]

85. Let $G = (V, E)$ be a (simple) graph, with $V = \{x_1, \ldots, x_n\}$, and let $t_1, \ldots, t_n$ be distinct real numbers. Show that the map $V \to \mathbb{R}^3$, $x_i \mapsto (t_i, t_i^2, t_i^3)$, gives an embedding of G into $\mathbb{R}^3$ with straight line segments.

86. Let $p_1, q_1, p_2, q_2, \ldots, p_m, q_m$ be $2m$ distinct points in the plane. Show that there are m disjoint polygonal arcs, with the j^{th} arc connecting p_j to q_j.

87. A *k-book* is a topological space homeomorphic to the union of k squares in $\mathbb{R}^3$, with any two sharing the same segment as a common side, called the *spine* of the book. Show that every graph has an embedding into a 3-book. [*Hint.* Put all vertices on a line in a square, parallel to the spine, and join each vertex x with $d(x)$ straight line segments to points on the spine.]

88.$^-$ Let G and H be graphs, with $V(H) = \{y_1, y_2, \ldots, y_k\}$. Justify the remarks before Theorem 18, namely that $G \succ H$ iff G has vertex=disjoint connected subgraphs $G_1, G_2, \ldots, G_k$ such that if $y_i y_j \in E(H)$, then G contains an edge from G_i to G_j.

89. Show that $G \succ K_4$ if and only if $TK_4 \subset G$, that is, G contains a subdivision of K_4.

90. Show that if H is a cubic multigraph and $G \succ H$, then $G \supset TH$.

91. Show that if K_5 is a minor of a graph G, then $G \supset TK_5$ or $G \supset TK_{3,3}$. Check that this implies the equivalence of Theorems 17 and 18.

92. A graph is said to be *outerplanar* if it can be drawn in the plane in such a way that all vertices are on the boundary of the unbounded face (or of any face, of course). Show that a graph is outerplanar iff it contains neither K_4 nor $K_{2,3}$ as a minor.

93. Fill in the small gap in the proof of Lemma 20: show that if cases (i) and (ii) do not apply then there are two adjacent vertices of degree 4.

94. Let $\mathcal{T}$ be the set of spanning trees of a connected graph of order n. Let H be the graph with vertex set $\mathcal{T}$ in which $T_1 \in \mathcal{T}$ is joined to $T_2 \in \mathcal{F}$ if $|E(T_1)\Delta E(T_2)| = 2$, i.e., if T_1 has precisely one edge not in T_2 (and so T_2 has one edge not in T_1). Show that H is connected and has diameter at most $n-2$.

95. Let G be a graph of size $\binom{k}{2} + 1$, and maximal degree at least 2. Show that there is a set $U \subset V(G)$ such that $|U| = k+1$ and $G[U]$ has no isolated vertices.

96. (Exercise 95 ctd.) Let G be a graph with $2k+1$ vertices and $\binom{2k+1}{2} - \binom{k}{2} - 1$ edges. Show that there is a partition $V(G) = U_1 \cup U_2$ of the vertex set such that $\Delta(G[U_i]) \le k-1$ for $i = 1, 2$. Show also that if G has one more edge then such a partition need not exist.

Notes

The first book on graph theory was written by the Hungarian D. Kőnig: *Theorie der endlichen und unendlichen Graphen, Kombinatorische Topologie und Streckenkomplexe*, Akademische Verlagsgesellschaft, Leipzig, 1936, 258 pp.; this book contains all the basic results. (For an English translation with commentaries, see it Theory of Finite and Infinite Graphs, Birkhäuser, Boston, 1990, 426 pp.) Euler's theorem on the bridges of Königsberg had been published 200 years before, in St. Petersburg: L. Euler, Solutio problematis ad geometrian situs pertinentis, *Comm. Acad. Sci. Imper. Petropol.* **8** (1736) 128–140. Theorem 14 is from P. Erdős, T. Grünwald, and E. Weiszfeld, On Euler lines of infinite graphs (in Hungarian), *Mat. Fiz. Lapok* **43** (1936), 129–140. In its full generality, Theorem 13 is due to T. van Aardenne-Ehrenfest and N.G. de Bruijn, Circuits and trees in oriented linear graphs, *Simon Stevin* **28** (1951), 203–217, but the case of degree 4 can be found in C.A.B. Smith and W.T.Tutte, On unicursal paths in a network of degree 4, *Amer. Math. Monthly* **48** (1941), 233–237.

Theorem 17 is in K. Kuratowski, Sur le problème des courbes gauches en topologie, *Fund. Math.* **15** (1930) 271–283; for simpler proofs see A.G. Dirac and S. Schuster, A theorem of Kuratowski, *Indag. Math.* **16** (1954) 343–348, C. Thomassen, Kuratowski's theorem, *J. Graph Theory* **5** (1981) 225–241, and H. Tverberg, A proof of Kuratowski's theorem, in: *Graph Theory in Memory of G.A. Dirac* (eds. L.D. Andersen et al), North Holland, Amsterdam, 1987.

The theorem of A.S. Amitsur and J. Levitzki (Theorem 19) is in Minimal identities for algebras, *Proc. Amer. Math. Soc.* **1** (1950) 449–463; the simpler and more combinatorial proof is based on R. G. Swan, An application of graph theory to algebra, *Proc. Amer. Math. Soc.* **14** (1963) 367–373 and Correction to "An application of graph theory to algebra," *Proc. Amer. Math. Soc.* **21** (1969) 379–380.

Steiner triple systems, mentioned in Exercises 41–43 are named after Jakob Steiner, who, in 1853, asked whether the necessary condition that $n \equiv 1$ or $3 \pmod 6$ is also sufficient for their existence. In fact, the same problem had been posed and answered in the affirmative by the Rev. Thomas Kirkman, On a problem of combinations, *Cambridge and Dublin Math. J.* **2** (1847) 101–204. We shall prove this in Exercises 84–86 of Chapter III.

II

Electrical Networks

This chapter is something of a diversion from the main line of the book, so at the first reading some readers may wish to skip it. The concepts introduced in the first half of Section 3 will be used in Section 2 of Chapter VIII, and in Chapter IX we shall return to electrical networks, when we connect them with random walks.

It does not take long to discover that an electrical network may be viewed as a graph, so the simplest problems about currents in networks are exactly questions about graphs. Does our brief acquaintance with graphs help us tackle the problems? As it will transpire in the first section, the answer is yes; for after a short review of the basic ideas of electricity we make use of spanning trees to obtain solutions. Some of these results can be reformulated in terms of tilings of rectangles and squares, as we shall show in Section 2. The last section introduces elementary algebraic graph theory, which is then applied to electrical networks.

It should be emphasized that in the problems we consider we use hardly more than the terminology of graph theory; virtually the only concept to be used is that of a spanning tree.

II.1 Graphs and Electrical Networks

A simple *electrical network* can be regarded as a graph in which each edge e_i has been assigned a real number r_i, called its *resistance*. If there is a *potential difference* p_i between the endvertices of e_i, say a_i and b_i, then an *electrical current* w_i will flow in the edge e_i from a_i to b_i according to *Ohm's law*:

$$w_i = \frac{p_i}{r_i}.$$

Though to start with we could restrict our attention to electrical networks corresponding to graphs, in the simplifications that follow it will be essential to allow *multiple* edges, that is, to consider *multigraphs* instead of graphs. Furthermore, we *orient* each edge arbitrarily from one endvertex to the other so that we may use p_i to denote the *potential difference in the edge* e_i, meaning the difference between the potentials of the *initial vertex* and the *endvertex.* Similarly, w_i is the *current in the edge* e_i, meaning the current in e_i in the direction of the edge. (Note that we regard a negative current $-w_i$ as a positive current w_i in the other direction.) Thus, throughout the section we consider *directed multigraphs*, that is, directed graphs that may contain *several* edges directed from a_i to b_i. However, in this section there is no danger of confusion if we use a_ib_i to denote an edge *from* a_i *to* b_i; in the next section we shall be more pedantic. Thus,

$$w_{ab} = -w_{ba} \quad \text{and} \quad p_{ab} = -p_{ba}.$$

In many practical problems, electrical currents are made to enter the network at some points and leave it at others, and we are interested in the consequent currents and potential differences in the edges. These are governed by the famous laws of Kirchhoff, another renowned citizen of Königsberg.

Kirchhoff's potential (or *voltage*) *law* states that the potential differences round any cycle $x_1x_2 \cdots x_k$ sum to 0:

$$p_{x_1x_2} + p_{x_2x_3} + \cdots + p_{x_{k-1}x_k} + p_{x_kx_1} = 0.$$

Kirchhoff's current law postulates that the total current outflow from any point is 0:

$$w_{ab} + w_{ac} + \cdots + w_{au} + w_{a\infty} = 0.$$

Here ab, ac, $\ldots$, au are the edges incident with a, and $w_{a\infty}$ denotes the amount of current that leaves the network at a. (In keeping with our convention, $w_{\infty a} = -w_{a\infty}$ is the amount of current entering the network at a.) For vertices not connected to external points we have

$$w_{ab} + w_{ac} + \cdots + w_{au} = 0.$$

Note that if we know the resistances then the potential law can be rewritten as a restriction on the currents in the edges. Thus we may consider that the currents are governed by the Kirchhoff laws only; the physical characteristics of the network (the resistances) affect only the parameters in these laws.

It is also easily seen that the potential law is equivalent to saying that one can assign absolute potentials V_a, V_b, $\ldots$ to the vertices a, b, $\ldots$ so that the potential difference between a and b is $V_a - V_b = p_{ab}$. If the network is connected and the potential differences p_{ab} are given for the edges, then we are free to choose arbitrarily the potential of one of the vertices, say V_a, but then all the other potentials are determined. In this section we shall work with absolute potentials, usually choosing the potential of one of the vertices to be 0, but we must keep in mind that this is the same as the application of the voltage law.

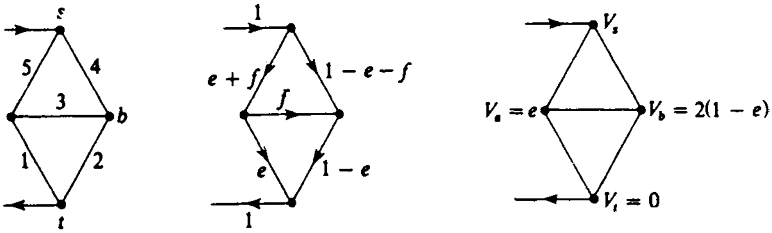


FIGURE II.1. The resistances, the currents, and the potentials.

In the most fundamental problems, current is allowed to enter the network only at a single vertex s, the *source*, and leave it only at another vertex t, the *sink*. (We shall indicate later how the general problem can be reduced to these fundamental problems.) If the size of the current from s to t is w and the potential difference between s and t is p, then by Ohm's law $r = p/w$ is the *total resistance* of the network between s and t. As an example of the use of the Kirchhoff laws we shall evaluate the total resistance between s and t of the simple network shown in Fig. II.1.

This network has 5 resistors, of values $1, 2, 3, 4$, and 5 ohms, as shown in the first picture. If we suppose that a unit current flows into the system at s and leaves it at t, then the consequent edge currents must be as in the second picture, for suitable values of e and f. Finally, the potentials $V_t = 0$, V_a, V_b, V_s assigned to the vertices must satisfy Ohm's law, so $V_a = 1 \cdot e = e$, $V_b = 2(1 - e)$, and $V_s = V_a + 5(e + f) = 6e + 5f$. Ohm's law has to be satisfied in two more edges, ab and bs, giving us

$$V_a = e = V_b + 3f = 2(1 - e) + 3f$$

and

$$V_s = 6e + 5f = V_b + 4(1 - e - f) = 2(1 - e) + 4(1 - e - f).$$

Hence

$$e = 2 - 2e + 3f$$

and

$$6e + 5f = 6 - 6e - 4f,$$

giving $e = 4/7$, $f = -2/21$ and $V_s = 6e + 5f = 62/21$. In particular, the total resistance from s to t is $(V_s - V_t)/1 = 62/21$.

The calculations are often simplified if we note that Kirchhoff's equations are linear and homogeneous in all currents and potential differences. This implies the so-called *principle of superposition*: any combination of solutions is again a solution. As an application of the principle of superposition one can show that any current resulting from multiple sources and sinks can be obtained by superposing flows belonging to one source and sink; that is, solutions of the fundamental problems mentioned above can be used to solve the general problem.

Furthermore, the principle of superposition implies immediately that there is at *most one solution*, no matter how the sources and sinks are distributed. Indeed, the difference of two distinct solutions is a flow in which *no* current enters or leaves the network at any point. If in this flow there is a positive current in some edge from a to b then by the current law a positive current must go from b to c, then from c to d, etc., giving a trail $abcd \cdots$. Since the network is finite, this trail has to return to a point previously visited. Thus we obtain a circuit in whose edges positive currents flow in one direction. But this is impossible, since it implies that the potential of each vertex is strictly greater than that of the next one round the circuit.

Before proving the *existence* of a solution (which is obvious if we believe in the physical interpretation), we shall calculate the total resistance of two networks. Unless the networks are very small, the calculations can get very heavy, and electrical engineers have a number of standard tricks to make them easier.

The very simple networks of Fig. II.2 show two resistors r_1 and r_2 connected first *in series* and then *in parallel*. Let us put a current of size 1 through the networks, from s to t. What are the total resistances? In the first case

$$V_a = r_1 \quad \text{and} \quad V_s = V_a + r_2 = r_1 + r_2,$$

so the total resistance is

$$r = r_1 + r_2.$$

In the second case, when they are connected in parallel, if a current of size e goes through the first resistor and so a current of size $1 - e$ through the second, then

$$V_s = r_1 e = r_2(1 - e), \quad \text{so} \quad e = \frac{r_2}{r_1 + r_2},$$

and the total resistance is given by

$$r = \frac{r_1 r_2}{r_1 + r_2}, \quad \text{or} \quad \frac{1}{r} = \frac{1}{r_1} + \frac{1}{r_2}.$$

This indicates that reciprocals of resistances, or *conductances*, are just as natural as the resistances themselves, and indeed are more convenient in our presentation. (The conductance of an edge of resistance 1 ohm is 1 mho.) What we have shown now is that for series connection the *resistances add* and for parallel connection the *conductances add*.

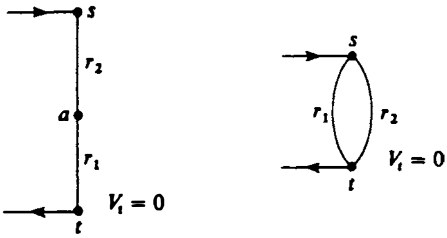


FIGURE II.2. Resistors connected in series and in parallel.

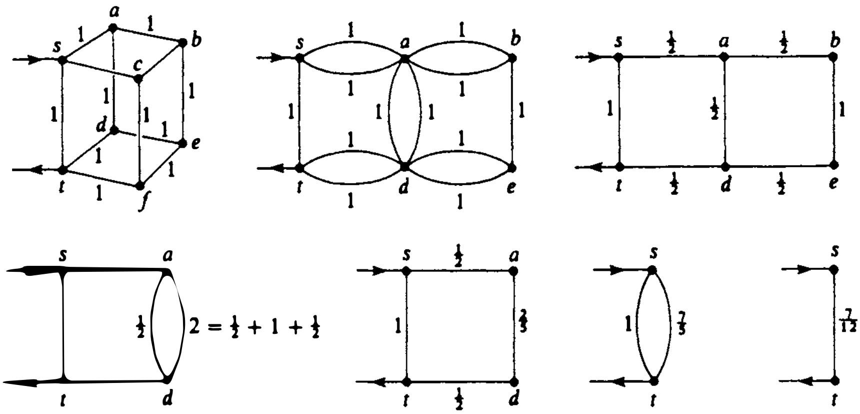


FIGURE II.3. Calculating the total resistance of a cube.

The use of conductances is particularly convenient when considering certain limiting cases of Ohm's law. If the resistance of an edge ab is 0, then we necessarily have $V_a = V_b$, and from an electrical point of view the vertices can be regarded as identical. In the usual slang, a has been "shorted" (short-circuited) to b. Of course, a may be shorted to b if there is some other reason why $V_a = V_b$. At the other extreme, we can introduce edges of 0 conductance without affecting the currents and potentials. Conversely, we make an edge have 0 conductance by "cutting" it. Of course, an edge of 0 resistance is said to have ∞ conductance, and an edge of 0 conductance is said to have ∞ resistance.

Let us see now how the acquaintance with resistors in series and in parallel and the possibility of shorting vertices can help us determine the total resistance. As an example, let us take the network formed by the edges of a cube, in which each edge has 1 ohm resistance. What is the total resistance across an edge st? Using the notation of the first picture in Fig. II.3, we see that by symmetry $V_a = V_c$ and $V_d = V_f$, so c can be shorted to a and f to d, giving us the second picture. From now on we can simplify resistors connected in parallel and in series, until we find that the total resistance is $7/12$. Knowing this, it is easy to recover the entire current flow.

Another important device in practical calculations is the so-called *star–triangle (or star–delta) transformation.* If a vertex v is joined to just three vertices, say a, b, and c, by edges of resistances A, B, and C, then we call v the centre of a *star*, as in the first picture of Fig. II.4. If no current is allowed to enter or leave at v, then we are allowed to replace this star by the *triangle* configuration shown in the second picture of Fig. II.4, because, as the reader should check (see Exercise 11), if the vertices a, b, c are set at potentials V_a, V_b, V_c, then in the two networks we get precisely the same currents $w_{a\infty}, w_{b\infty}, w_{c\infty}$ leaving the network. Needless to say, we may apply the transformation in reverse, replacing A', B', and C' by $A = B'C'/T$, $B = C'A'/T$, and $C = A'B'/T$, where $T = A' + B' + C'$. Incidentally, the formulae become symmetrical if we use resistances in the first transformation

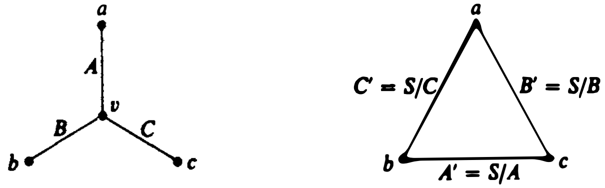


FIGURE II.4. The star–triangle transformation; $S = AB + BC + CA$.

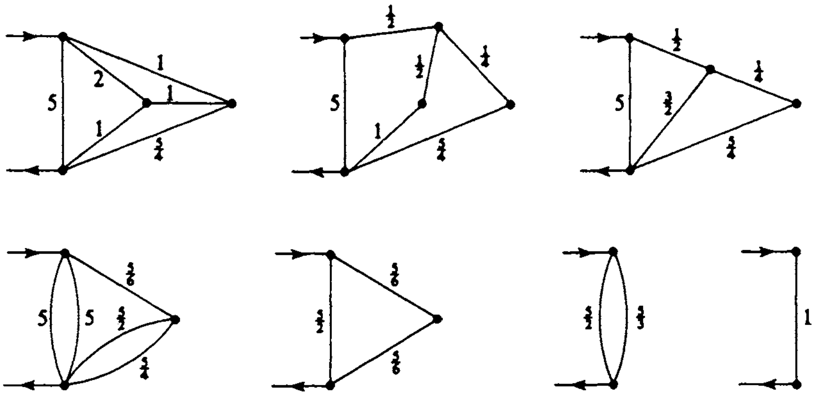


FIGURE II.5. Applications of the star–triangle transformation.

and conductances in the second: $A' = B + C + BC/A$ and $\alpha = \beta' + \gamma' + \beta'\gamma'/\alpha'$, where $\alpha, \beta, \ldots$ are the conductances.

As an application of the star–triangle transformation, let us calculate the total resistance of a tetrahedron across an edge, in which the resistances are as in Fig. II.5. The pictures speak for themselves.

We shall conclude this section on a slightly more theoretical note: we shall prove the existence of a solution. More precisely, we shall present Kirchhoff's theorem stating that, if a current of size 1 is put through a network, then the current in an edge can be expressed in terms of the numbers of certain spanning trees. For simplicity we assume that the graph G of the network is connected, each edge has unit resistance, and a current of size 1 enters at a vertex s and leaves at t.

Theorem 1 *Given an edge ab, denote by $N(s, a, b, t)$ the number of spanning trees of G in which the (unique) path from s to t contains a and b, in this order. Define $N(s, b, a, t)$ analogously and write N for the total number of spanning trees. Finally, let $w_{ab} = \{N(s, a, b, t) - N(s, b, a, t)\}/N$.*

Distribute currents in the edges of G by sending a current of size w_{ab} from a to b for every edge ab. Then there is a total current size 1 from s to t satisfying the Kirchhoff laws.

Proof. To simplify the situation, multiply all currents by N. Also, for every spanning tree T and edge $ab \in E(G)$, let $w^{(T)}$ be the current of size 1 along the unique s–t path in T:

$$w_{ab}^{(T)} = \begin{cases} 1 & \text{if } T \text{ has a path } s \cdots ab \cdots t, \\ -1 & \text{if } T \text{ has a path } s \cdots ba \cdots t, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$N(s, a, b, t) - N(s, b, a, t) = \sum_T w_{ab}^{(T)},$$

where the summation is over all spanning trees T. Therefore, our task is to show that if we send a current of size $\sum_T w_{ab}^{(T)}$ from a to b for every edge ab, then we obtain a total current of size N from s to t satisfying the Kirchhoff laws.

Now, each $w^{(T)}$ is a current of size 1 from s to t satisfying Kirchhoff's current law, and so their sum is a current of size N from s to t satisfying Kirchhoff's current law.

All we have to show then is that the potential law is also satisfied. As all edges have the same resistance, the potential law claims that the total current in a cycle with some orientation is zero. To show this, we proceed as earlier, but first we reformulate slightly the definition of $N(s, a, b, t)$. Call a spanning forest F of G a *thicket* if it has exactly two components, say F_s and F_t, such that s is in F_s and t is in F_t. Then $N(s, a, b, t)$ is the number of thickets $F = F_s \cup F_t$ for which $a \in F_s$ and $b \in F_t$, and $N(s, b, a, t)$ is defined analogously. What is then the contribution of a thicket $F = F_s \cup F_t$ to the total current in a cycle? It is the number of cycle edges from F_s to F_t minus the number of cycle edges from F_t to F_s; so it is zero. □

Let us write out the second part of the proof more formally, to make it even more evident that we use the basic and powerful combinatorial principle of double counting, or reversing the order of summation. For a thicket $F = F_s \cup F_t$ and an edge $ab \in E(G)$, set

$$w_{ab}^{(F)} = \begin{cases} w_{ab}^{(F+ab)} & \text{if } F + ab \text{ is a spanning tree,} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\sum_T w_{ab}^{(T)} = \sum_F w_{ab}^{(F)},$$

where the second summation is over all thickets F. Finally, the total current around a cycle $x_1 x_2 \cdots x_k$ of G, with $x_{k+1} = x_1$, is

$$\sum_{i=1}^{k} \sum_F w_{x_i x_{i+1}}^{(F)} = \sum_F \sum_{i=1}^{k} w_{x_i x_{i+1}}^{(F)} = 0,$$

since $\sum_{i=1}^{k} w_{x_i x_{i+1}}^{(F)} = 0$ for every thicket F.

More importantly, the proof of Theorem 1 can be rewritten to give a solution in the case when the edges have arbitrary conductances. For a spanning tree T define the *weight* $w(T)$ of T as the product of the conductances of its edges. Let N^* be the sum of the weights of all the spanning trees, let $N^*(s, a, b, t)$ be the sum of the weights of all the spanning trees in which b follows a on the (unique) s–t path in the tree, and let $N^*(s, b, a, t) = N^*(t, a, b, s)$.

Theorem 2 *There is a distribution of currents satisfying Ohm's law and Kirchhoff's laws in which a current of size* 1 *enters at* s *and leaves at* t. *The value of the current in an edge* ab *is given by* $\{N^*(s, a, b, t) - N^*(s, b, a, t)\}/N^*$. □

Let us note an immediate consequence of this result.

Corollary 3 *If the conductances of the edges are rational and a current of size* 1 *goes through the network then the current in each edge has rational value.* □

The star–triangle transformation tells us that no matter what the rest of the network is, every 'star' may be replaced by a suitable 'triangle', and vice versa. On an even simpler level, if two networks, N and M, share only two vertices, say a and b, and nothing else, and the total resistance of M from a to b is r, then in $N \cup M$ we may replace M by an edge ab of resistance r. In fact, similar transformations can be carried out for networks with any number of vertices of attachment, not only two or three, as above. To be precise, if a part M of a network is attached to the rest of the network only at a set U of vertices, then we may replace M by edges of certain resistances joining the vertices of U (and introducing no other vertices) without changing the distribution of currents outside M. We leave this as an exercise (Exercise 13^+).

In estimating the resistance of a network, it is frequently convenient to make use of the fact that if the resistance of a wire is increased then the total resistance does not decrease. In particular, if some wires are cut then the total resistance does not decrease; similarly, if some vertices are shorted, i.e., are identified, then the total resistance does not increase. This is obvious if we appeal to physical intuition; however, the problem is that the Kirchhoff laws, together with Ohm's law, determine all currents, potential differences, and so on: having accepted these three laws, we have no right to appeal to any physical intuition. In this chapter we leave this assertion as an exercise (Exercise 14^+), but we shall prove it, several times over, in Chapter IX, when we give a less superficial treatment of electrical networks.

II.2 Squaring the Square

This is a diversion within a diversion; we feel bound to draw attention to a famous problem arising from recreational mathematics that is related to the theory of electrical networks. Is there a *perfect squared square*? In other words, is it possible to subdivide a closed square into finitely many (but at least two) square regions of *distinct* sizes that intersect only at their boundaries?

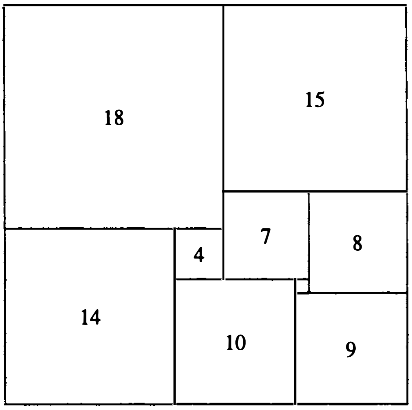


FIGURE II.6. The perfect squaring of the 33 × 32 rectangle, due to Moroń.

The answer to this question is far from obvious: on the one hand, there seems to be no reason why there should *not* be a perfect squared square; on the other hand, it is not easy to find even a perfect squared *rectangle*, a rectangle divided into finitely many (but at least two) squares of distinct sizes.

As it happens, there are perfect squared rectangles: in 1925 Moroń found the perfect squaring of the 33 × 32 rectangle shown in Fig. II.6. This squared rectangle has *order* 9: there are 9 squares in the subdivision; in the figure the number associated with a square is the length of its side.

We shall use Moroń's squared rectangle to illustrate an argument. Let us cut this rectangle out of a sheet of nichrome (or any other material with low conductivity) and let us put rods made of silver (or some other material of high conductivity) at the top and bottom.

What happens if we ensure that the silver rod at the top is at 32 volts while the rod at the bottom is kept at 0? Trivially, a uniform current will flow from top to bottom. In fact, the potential at a point of the rectangle will depend only on the height of the point: the potential at height x will be x volts. Furthermore, there will be no current across the rectangle, only from top to bottom. Thus the current will not change at all if (i) we place silver rods on the horizontal sides of the squares and (ii) cut narrow slits along the vertical sides, as shown in the first picture of Fig. II.7.

Now, since silver is a very good conductor, the points of each silver rod have been shortened, so they can be identified. Thus as an electric conductor the whole rectangle behaves like the *plane* network shown in the second picture of Fig. II.7, in which the conductance of an edge is equal to the conductance of the corresponding square from top to bottom. Clearly, the conductance of a rectangle from top to bottom is proportional to the length of a horizontal side and the resistance is proportional to a vertical side. Consequently, all squares have the same resistance, say unit resistance, so all edges in Fig. II.7 have unit resistance. What is the potential drop in an edge? It is the side length of the corresponding square. What is the resistance of the whole system? The ratio of the vertical side of the original big rectangle to the horizontal side, that is, 32/33.

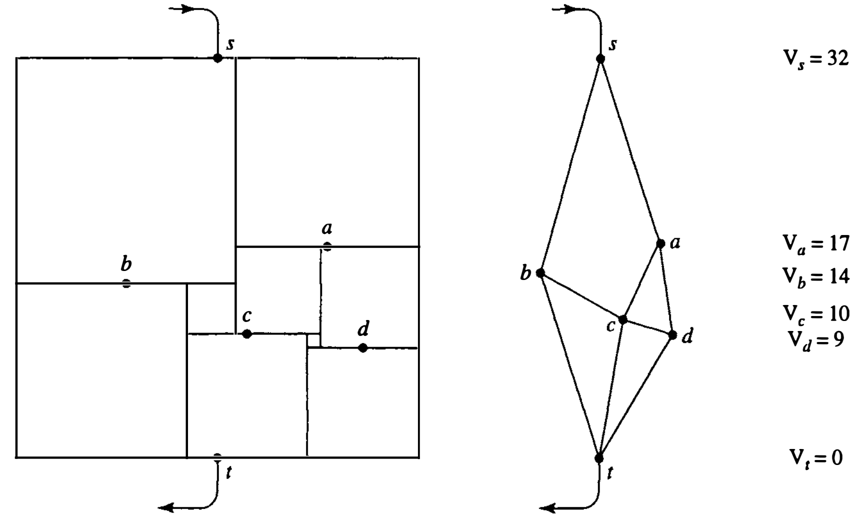


FIGURE II.7. The electrical network associated with our rectangle.

Since the process above is reversible, that is, every squared rectangle can be obtained from some network, we have an effective *tool* to help us in our search for squared squares. Take a connected planar graph G and turn it into an electrical network by giving each edge resistance 1. Calculate the total resistance from a vertex s to a vertex t. If this is also 1, the network may correspond to a suitably squared square. If the potential differences in the edges are all distinct, all squares have different sizes, so we have a perfect squared square.

Of course, at this stage our problem is far from being solved; we do not even know that there *must* exist a squared square. However, we have a chance to search systematically for a solution. What should we look for? A plane graph containing s and t on the outer face, lacking all symmetries, such that the total resistance from s to t is 1.

Many squared squares have been found with the help of computers, but the first examples were found without computers by Sprague in 1939 and by four undergraduates at Cambridge – Brooks, Smith, Stone and Tutte – in 1940. The smallest number of squares that can tile a square is 21; Fig. II.8 shows such a tiling, due to Duijvestijn. In fact, this is the only tiling of order 21. Several other tilings are given among the exercises.

The connection between squaring a rectangle and electrical networks gives us immediately a beautiful result first proved by Dehn in 1903. Corollary 3 tells us that if each edge has resistance 1 and a current of size 1 flows through the system then in each edge the value of the current is rational. This translates to the following result about squared rectangles.

Theorem 4 *If a rectangle can be tiled with squares then the ratio of two neighbouring sides of the rectangle is rational.* □

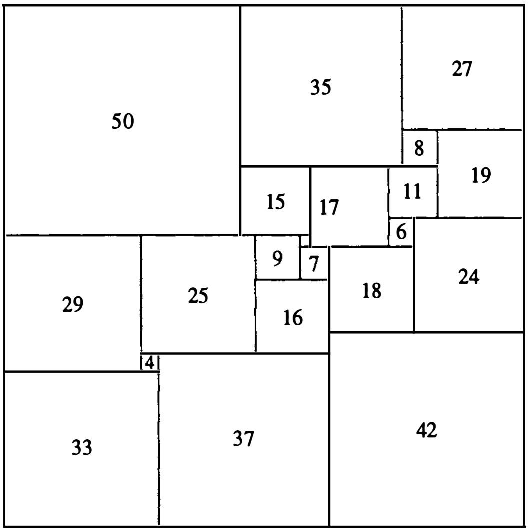


FIGURE II.8. A perfect squared square: a tiling of a square with 21 incongruent squares.

Equivalently, a rectangle can be tiled with squares iff it can be tiled with congruent squares.

It is easily seen that electrical networks can be used to obtain tilings of rectangles of prescribed shapes: an edge e of resistance r corresponds to a rectangle in which the height is r times the base (see Exercise 19).

Numerous questions remain about squared rectangles; here we mention only two. First, which plane networks correspond to perfect squared rectangles? The answer holds no surprises: if s and t are on the outer face of a plane network, with each edge having resistance 1, then this network corresponds to a squared rectangle iff the following condition is satisfied: when a non-zero current is put through the network from s to t, there is a non-zero current in each edge.

Second, which rectangles have perfect squarings? This question is considerably more difficult to answer. The result below, greatly extending Dehn's theorem (Theorem 4), was proved by Sprague in 1940.

Theorem 5 *A rectangle has a perfect squaring if, and only if, the ratio of two neighbouring sides is rational.*

The result can be proved by putting together appropriate perfect rectangles; for the proof we refer the reader to the original paper of Sprague.

In the rest of this section, we consider tilings of *rectangles* by *rectangles*: our aim is to prove some beautiful results that somewhat resemble the results above. Suppose that we have a tiling of a rectangle by 1×8 and 8×1 rectangles. Then, as the total area is a multiple of 8, either one of the sides is a multiple of 4 *and* the other is even, or one of the sides is a multiple of 8. Can both possibilities arise? There are similar questions in higher dimensions. For example, if a box is filled with $1 \times 2 \times 4$ bricks in any position ($1 \times 2 \times 4$, $4 \times 2 \times 1$, $2 \times 1 \times 4$, etc.), then either all sides of the box are even, or one is a multiple of 4, another is even and the third is odd, or else one of the sides is a multiple of 8 and the two other sides are odd. But can all possibilities arise?

The latter problem was posed by de Bruijn in a Hungarian journal in 1959; a decade later he proved considerable extensions of the result, including the theorems below. There is a galaxy of beautiful proofs of the first theorem: here we give four. Call a side of a rectangle *integer* if its length is an integer.

Theorem 6 *Let a rectangle T be tiled with rectangles $T_1, \dots, T_\ell$. If each T_i has an integer side then so does T.*

Remark. In all four proofs we assume, as we may, that $T \subset \mathbb{R}^2$ is in *canonical* position: it has vertices $(0, 0)$, $(a, 0)$, $(0, b)$ and (a, b), where $a, b > 0$. Then the sides of the T_i are also parallel to the axes.

First Proof. Construct a bipartite graph G, with vertex classes L and R, as follows. Let L (for 'left' or 'lattice points') be the set of integer lattice points in the tiled rectangle: $L = \{(x, y) \in \mathbb{Z}^2 : 0 \leq x \leq a,\ 0 \leq y \leq b\}$, and let R (for 'right' or 'rectangles') be the set of tiling rectangles $T_1, \dots, T_\ell$. Our graph G has vertex set $L \cup R$, and $(x, y) \in L$ is joined to $T_i \in R$ if (x, y) is a vertex ('corner') of T_i. Then, since each T_i has an integer side, each T_i has degree 0, 2 or 4, so $e(G)$ is even.

Also, every vertex in L, other than the corners of T, has degree 0, 2 or 4, but the corner $(0, 0) \in L$ has degree 1. Hence G has at least one edge incident with another corner: in particular, at least one other corner belongs to L, and we are done. □

Second Proof. Set $F(x, y) = \sin 2\pi x \sin 2\pi y$. Then

$$\int\int_{T_i} F(x, y)dx\,dy = 0$$

for each i, so

$$\int\int_T F(x, y)dx\,dy = \sum_{i=1}^{\ell} \int\int_i F(x, y)dx\,dy = 0.$$

But

$$\begin{aligned}\int\int_T F(x, y)dx\,dy &= \int_0^b \left(\int_0^a F(x, y)dx\right) dy \\ &= \int_0^a \sin 2\pi x\, dx \int_0^b \sin 2\pi y\, dy \\ &= \left(\frac{1}{2\pi}\right)^2 (1 - \cos 2\pi a)(1 - \cos 2\pi b).\end{aligned}$$

Hence at least one of a and b is an integer. □

Third Proof. Colour the $1/2 \times 1/2$ squares of the square lattice $\frac{1}{2}\mathbb{Z}^2$ in a black and white checkerboard fashion. Then each tile T_i contains an equal amount of black and white, and so T itself has an equal amount of black and white. But it is easily checked that in this case at least one of a and b is an integer. □

Fourth Proof. For $\varepsilon > 0$ and $x \in \mathbb{R}$, set

$$\phi_\varepsilon(x) = \begin{cases} x & \text{if } x \in \mathbb{Z}, \\ x + \varepsilon & \text{otherwise.} \end{cases}$$

Also, for $\mathbf{z} = (x, y) \in \mathbb{R}^2$, define $\phi_\varepsilon(\mathbf{z}) = (\phi_\varepsilon(x), \phi_\varepsilon(y))$, and for a rectangle U with corners $(\mathbf{z}_i)_1^4$ let $\phi_\varepsilon(U)$ be the (possibly degenerate) rectangle with corners $(\phi_\varepsilon(\mathbf{z}_i))_1^4$.

It is an easy exercise to show that if $\varepsilon > 0$ is small enough, say $0 < \varepsilon < \varepsilon_0$, then the rectangles $\phi_\varepsilon(T_i)$ form a tiling of $\phi_\varepsilon(T)$. Writing $|U|$ for the area of a rectangle U, if $0 < \varepsilon < \varepsilon_0$ then

$$|\phi_\varepsilon(T)| = \sum_{i=1}^{\ell} |\phi_\varepsilon(T_i)|.$$

Now, as T_i has an integer side, $|\phi_\varepsilon(T_i)|$ is a linear function of ε for $0 < \varepsilon < \varepsilon_0$. On the other hand, if $a, b \notin \mathbb{Z}$ then $|\phi_\varepsilon(T)| = ab + (a+b)\varepsilon + \varepsilon^2$ is a quadratic function of ε. As this is not the case, our proof is complete. □

It is easy to generalize the result to n-dimensional *boxes*, rectangular parallelepipeds.

Theorem 7 *Let a box B in $\mathbb{R}^n$ be tiled with boxes $B_1, \dots, B_\ell$. If each B_i has at least k integer sides, then B itself has at least k integer sides.*

Proof. The fourth proof above carries over, *mutatis mutandis*. Defining ϕ_ε as before, with $\phi_\varepsilon(\mathbf{z}) = (\phi_\varepsilon(z_1), \dots, \phi_\varepsilon(z_n))$, we find that, if $\varepsilon > 0$ is small enough, each $|\phi_\varepsilon(B_i)|$ is a polynomial of degree at most $n - k$. Also, if B has precisely h integer sides then $|\phi_\varepsilon(B)|$ has degree $n - h$. □

Theorem 8 *Let $a_1, \dots, a_n$ be natural numbers with $a_1|a_2, \dots, a_{n-1}|a_n$, and let B be an $A_1 \times \cdots \times A_n$ box filled with $a_1 \times \cdots \times a_n$ bricks standing in any position. Then B can also be filled with these bricks positioned the same way. Equivalently, there is a permutation π of $\{1, \dots, n\}$ such that a_i divides $A_{\pi(i)}$.*

Proof. By Theorem 6, we know that a_n divides an A_i: let $\pi(n)$ be such that a_n divides $A_{\pi(n)}$. Next, we know by Theorem 7 that a_{n-1} divides at least two A_i: let $\pi(n-1) \neq \pi(n)$ be such that a_{n-1} divides $A_{\pi(n-1)}$. Continuing in this way, we get a permutation as desired. □

For some more proofs and extensions of Theorem 6, see Exercises 28–35.

II.3 Vector Spaces and Matrices Associated with Graphs

The *vertex space* $C_0(G)$ of a graph G is the complex vector space of all functions from $V(G)$ into $\mathbb{C}$. Similarly, the *edge space* $C_1(G)$ is the complex vector space of

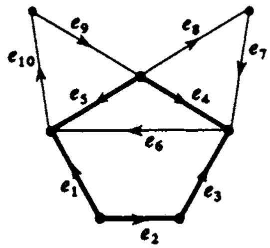


FIGURE II.9. If the thick cycle L is oriented anti-clockwise, its vector in $C_1(G)$ is $\mathbf{z}_L = (-1, 1, 1, -1, 0, \ldots, 0)$.

all linear functions from $E(G)$ into $\mathbb{C}$. In these definitions it is sometimes convenient to replace the complex field by F_2, the field of order 2, or by other fields. We shall take $V(G) = \{v_1, v_2, \ldots, v_n\}$ and $E(G) = \{e_1, e_2, \ldots, e_m\}$, so that $\dim C_0(G) = n$ and $\dim C_1(G) = m$. The elements of $C_0(G)$ are usually written in the form $\mathbf{x} = \sum_{i=1}^n x_i v_i$ or $\mathbf{x} = (x_i)_1^n$. The sum $\sum_{i=1}^n x_i v_i$ is a *formal sum* of the vertices, but if we think of v_i as the function $V(G) \to \mathbb{C}$ that is 0 everywhere, except at the vertex v_i, where it is 1, then $v_i, \ldots, v_n$ is a basis of $C_0(G)$ and the sum above simply expresses an element in terms of the basis elements. Similarly, an element of $C_1(G)$ may be written as $\mathbf{y} = \sum_{i=1}^m y_i e_i$ or $\mathbf{y} = (y_i)_1^m$. We call $(v_1, \ldots, v_m)$ the *standard basis* of the vertex space $C_0(G)$ and $(e_1, \ldots, e_m)$ the *standard basis* of the edge space. We shall endow these spaces with the inner product in which the standard bases are orthonormal: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i \overline{y}_i$.

In this section we shall be concerned mostly with the edge space $C_1(G)$; to start with we define two subspaces which will turn out to be orthogonal complements of each other. Let L be a cycle in G with a given cyclic orientation $L = u_1 u_2 \cdots u_l$. If $e_i = u_j u_{j+1}$ and e_i is oriented from u_j to u_{j+1} then we say that e_i *is oriented as* L. This oriented cycle L can be identified with an element $\mathbf{z}_L$ of $C_1(G)$:

$$\mathbf{z}_L(e_i) = \begin{cases} 1 & \text{if } e_i \in E(L) \text{ and } e_i \text{ is oriented as } L, \\ -1 & \text{if } e_i \in E(L) \text{ and } e_i \text{ is not oriented as } L, \\ 0 & \text{if } e_i \notin E(L). \end{cases}$$

A simple example is shown in Fig. II.9. Denote by $Z(G)$ the *subspace* of $C_1(G)$ spanned by the vectors $\mathbf{z}_L$ as L runs over the set of cycles; $Z(G)$ is the *cycle space* of G.

Now let P be a partition $V = V_1 \cup V_2$ of the vertex set of G. Consider the set $E(V_1, V_2)$ of edges from V_1 to V_2; such a set of edges is called a *cut*. There is a vector $\mathbf{u}_P$ in $C_1(G)$ called a *cut vector*, or *cocycle vector* naturally associated with this partition P:

$$\mathbf{u}_P(e_i) = \begin{cases} 1 & \text{if } e_i \text{ goes from } V_1 \text{ to } V_2, \\ -1 & \text{if } e_i \text{ goes from } V_2 \text{ to } V_1, \\ 0 & \text{if } e_i \notin E(V_1, V_2). \end{cases}$$

We write $U(G)$ for the *subspace* of the edge space $C_1(G)$ spanned by all the cut vectors $\mathbf{u}_P$, and we call it the *cut* (or *cocycle*) space of G.

Theorem 9 *The inner product space $C_1(G)$ is the orthogonal direct sum of the cycle space $Z(G)$ and the cut space $U(G)$. If G has n vertices, m edges and k components then*

$$\dim Z(G) = m - n + k \quad \text{and} \quad \dim U(G) = n - k.$$

Proof. Let us see first that $Z(G)$ and $U(G)$ are orthogonal. Let L be a cycle and P a partition $V = V_1 \cup V_2$. What is the product $\langle \mathbf{z}_L, \mathbf{u}_P \rangle$? It is simply the number of edges of L going from V_1 to V_2 in the orientation of L, minus the number of edges of L from V_2 to V_1. Thus $\langle \mathbf{z}_L, \mathbf{u}_P \rangle = 0$ for every cycle L and partition P, so $Z(G)$ and $U(G)$ are indeed orthogonal.

Since the dimension of $C_1(G)$ is the number of edges, m, both assertions will be proved if we show that $\dim Z(G) \geq m - n + k$ and $\dim U(G) \geq n - k$. We shall first prove this under the assumption that G is connected; the general case will follow easily.

Thus let us assume that G is connected, that is, $k = 1$. Let T be a spanning tree of G. We shall make use of T to exhibit $m - n + 1$ independent vectors in $Z(G)$ and $n - 1$ independent vectors in $U(G)$. We may choose the indices of the edges in such a way that $e_1, e_2, \ldots, e_{n-1}$ are the tree edges and $e_n, e_{n+1}, \ldots, e_m$ are the remaining edges, the *chords* of T.

We know that for every chord e_i there is a (unique) oriented cycle C_i such that $\mathbf{z}_{C_i}(e_i) = 1$ and $\mathbf{z}_{C_i}(e_j) = 0$ for every other chord e_j, that is, whenever $j \geq n$ and $j \neq i$. (For short: $\mathbf{z}_{C_i}(e_j) = \delta_{ij}$ if $j \geq n$, where δ_{ij} is the Kronecker delta.) We call C_i the *fundamental cycle* belonging to e_i (with respect to T); also, $\mathbf{z}_{C_i}$ is a *fundamental cycle vector* (see Fig. II.10). Similarly, by deleting an edge e_i of T the remainder of the spanning tree falls into two components. Let V_1^i be the vertex set of the component containing the initial vertex of e_i and let V_2^i be the vertex set of the component containing the terminal vertex of e_i. If P_i is the partition $V = V_1^i \cup V_2^i$ then clearly $\mathbf{u}_{P_i}(e_j) = \delta_{ij}$ for $1 \leq j \leq n - 1$. The cut $E(V_1^i, V_2^i)$ is the *fundamental cut*, or *fundamental cocycle*, belonging to e_i (with respect to T), and $\mathbf{u}_{P_i}$ is the *fundamental cut vector*, or *fundamental cocycle vector*.

It is easily seen that $\{\mathbf{z}_{C_i} : n \leq i \leq m\}$ is an independent set of cycle vectors. Indeed, if $\mathbf{z} = \sum_{i=n}^{m} \lambda_i \mathbf{z}_{C_i} = 0$ then for every $j \geq n$ we have $0 = \mathbf{z}(e_j) = \sum_{i=n}^{m} \lambda_i \delta_{ij} = \lambda_j$, and so every coefficient λ_j is 0. Similarly, the fundamental cut vectors $\mathbf{u}_{P_i}$, $1 \leq i \leq n - 1$, are also independent. Hence $\dim Z(G) \geq m - n + 1$ and $\dim U(G) \geq n - 1$, as required.

Finally, the general case $k \geq 1$ follows immediately from the case $k = 1$. For if G has components $G_1, G_2, \ldots, G_k$ then $C_1(G)$ is the orthogonal direct sum of the subspaces $C_1(G_i)$, $i = 1, 2, \ldots, k$; furthermore, $Z(G_i) = Z(G) \cap C_1(G_i)$ and $U(G_i) = U(G) \cap C_1(G_i)$. $\square$

The proof above shows that $n(G) = \dim Z(G)$, called the *nullity* of G, and $r(G) = \dim U(G)$, the *rank* of G, are independent of the field over which the edge space is defined. The nullity is also called the *cyclomatic number* or *corank*

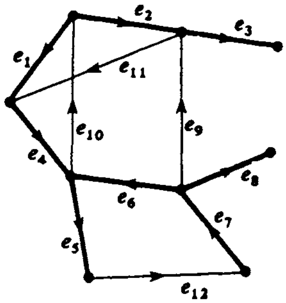


FIGURE II.10. The fundamental cycle vector belonging to e_9 is $\mathbf{z}_{C_9} = e_9 - e_2 + e_1 + e_4 - e_6$; the fundamental cut vector belonging to e_4 is $\mathbf{u}_{P_4} = e_4 - e_{10} - e_9$.

of G. The use of a spanning tree in the proof is not compulsory; in some cases, for instance in the case of a planar graph, there are other natural cycle and cut bases (cf. Exercise 37).

There are several matrices naturally associated with a graph and its vector spaces discussed above. The adjacency matrix $A = A(G) = (a_{ij})$ of a graph G is the $n \times n$ matrix given by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i v_j \in E(G), \\ 0 & \text{otherwise.} \end{cases}$$

In order to define the incidence matrix of a graph, we again consider an *orientation* of the edges, as in the definition of the cycle and cut spaces. The *incidence matrix* $B = B(G) = (b_{ij})$ of G is the $n \times m$ matrix defined by

$$b_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is the initial vertex of the edge } e_j, \\ -1 & \text{if } v_i \text{ is the terminal vertex of the edge } e_j, \\ 0 & \text{otherwise.} \end{cases}$$

There is a simple connection between the two matrices A and B. As usual, we write M^t for the transpose of a matrix M.

Theorem 10 *Let $D = (D_{ij})$ be the $n \times n$ diagonal matrix with $D_{ii} = d(v_i)$, the degree of v_i in G. Then*

$$BB^t = D - A.$$

Proof. What is $(BB^t)_{ij}$? It is $\sum_{l=1}^m b_{il}b_{jl}$, which is $d(v_i)$ if $i = j$, -1 if $v_i v_j$ is an edge (if $e_l = v_i v_j$ is directed from v_i to v_j, then $b_{il}b_{jl} = 1(-1) = -1$ and all other products are 0), and 0 if $v_i v_j$ is not an edge and $i \neq j$. □

The matrix $L = D - A$, the *combinatorial Laplacian* or *Kirchhoff matrix* of a graph, is of great importance in spectral graph theory: we shall return to it at the end of this section and in Chapter IX.

We may and will identify the matrices A and B with the linear maps $A : C_0(G) \to C_0(G)$ and $B : C_1(G) \to C_0(G)$ that they define in the standard bases:

$(A\mathbf{x})_i = \sum_{j=1}^{n} a_{ij}x_j$ and $(B\mathbf{y})_i = \sum_{j=1}^{m} b_{ij}y_j$. If we wanted to be pedantic, we would write the vectors in the vertex and edge spaces as *column vectors*, or we would put $A\mathbf{x}^t$ and $B\mathbf{y}^t$, where t stands for transposition; we shall not do this since there is no danger of confusion. If C is a cycle then clearly $B\mathbf{z}_C = 0 \in C_0(G)$; in fact, it is easily shown (cf. Exercise 38) that the cycle space is exactly the *kernel* of B. Thus the rank of B is $r(B) = m - (m - n + k) = n - k$, the rank of G, and its nullity is the corank, or cyclomatic number, of G. Furthermore, the transpose of B maps $C_0(G)$ into $C_1(G)$, and the *image* of B^t is exactly the cut space (cf. Exercise 39).

In Chapter VIII we shall discuss in some detail the eigenvalues and eigenvectors of the adjacency matrix; in this section we shall use the matrices to solve the electrical network problem discussed in the first section. In fact, it was Kirchhoff who first realized the applicability of matrix algebra to graph theory, exactly in connection with the electrical network problem.

How can we formulate the Kirchhoff laws in terms of matrices and vectors in the edge space? Let us assume that G' is the graph of our electrical network, $V(G') = \{v_1, v_2, \ldots, v_{n-1}\}$, $E(G') = \{e_1, e_2, \ldots, e_{m'}\}$, the network is connected and we have a voltage generator ensuring that the potential difference between v_i and v_j is $g_i - g_j$ volts for $1 \leq i \leq j \leq k$. In order to express Kirchhoff's laws in a neat form, we add a vertex v_n to G', and join it to $v_1, v_2, \ldots, v_k$; the new graph is G. Let $m = m' + k$ and $e_{m'+i} = v_n v_i$, $i = 1, 2, \ldots, k$, so that $V(G) = \{v_1, v_2, \ldots, v_n\}$ and $E(G) = \{e_1, e_2, \ldots, e_m\}$.

Give the edges of G' an arbitrary orientation and let w_i be the amount of current flowing in the edge e_i; thus $w_i = -1$ means a current of 1 ampere in the opposite direction. Direct each new edge $e_{m'+i}$ from v_n to v_i and let $w_{m'+i}$ be the *total current* entering the network at v_i. Once again, $w_{m'+i} = -1$ means that a current of 1 ampere leaves the network at v_i. The vector $\mathbf{w} = (w_1, w_2, \ldots, w_m) \in C_1(G)$ is the *current vector*. In this notation Kirchhoff's current law takes the form

$$B\mathbf{w} = \mathbf{0}. \tag{1}$$

It is just as easy to formulate Kirchhoff's potential law in matrix form. Let p_i be the potential difference in the edge e_i and let $\mathbf{p} = (p_1, p_2, \ldots, p_m) \in C_1(G)$ be the *potential vector*. The potential law states that $\langle \mathbf{z}, \mathbf{p} \rangle = 0$ for every cycle $\mathbf{z} \in C_1(G)$. Instead of postulating this about every cycle, we collect all the necessary information into a single matrix. As before, we choose a spanning tree T in G and label the edges so that $e_1, e_2, \ldots, e_{n-1}$ are the tree edges and $e_n, e_{n+1}, \ldots, e_m$ are the chords. Let C be the $m \times (m - n + 1)$ matrix whose i th column is the fundamental cycle vector $\mathbf{z}_{C_{n-1+i}}$ belonging to the edge e_{n-1+i}, $i = 1, 2, \ldots, m - n + 1$. Since the fundamental cycle vectors form a basis of the cycle space, the potential law takes the form

$$C^t\mathbf{p} = \mathbf{0}, \tag{2}$$

where C^t denotes the transpose of C.

Now, in order to find the current through the edges of G' we need one more equation, namely the equation relating the potential to the current, the resistance and the voltage generator. For $i \leq m'$, let r_i be the resistance of the edge e_i, and postulate that each new edge e_j, $j \geq m'+1$, has resistance $r_j = 0$. We may assume that $r_i > 0$ for every $i \leq m'$, since otherwise the edge e_i could have been cut. Let $R = (R_{ij})$ be the $m \times m$ diagonal matrix with $R_{ii} = r_i$. Finally, let $\mathbf{g} = (0, \ldots, 0, g_1, g_2, \ldots, g_k) \in C_1(G)$ be the vector of the voltage generator. Then clearly,

$$\mathbf{p} = R\mathbf{w} + \mathbf{g}. \tag{3}$$

This equation contains all the information we have about the electric current in addition to the Kirchhoff laws.

In order to solve (1), (2) and (3) for $\mathbf{w}$ and $\mathbf{p}$, we shall split $C_1(G)$ as $E_T + E_N$, where E_T is the subspace spanned by the tree edges and E_N is spanned by the chords, the edges not belonging to T. Let $\mathbf{w} = (\mathbf{w}_T, \mathbf{w}_N)$ and $\mathbf{p} = (\mathbf{p}_T, \mathbf{p}_N)$ be the corresponding splittings; furthermore, writing $\widetilde{B}$ for the matrix obtained from B by omitting the last row, we have

$$C = \begin{pmatrix} C_T \\ C_N \end{pmatrix} \quad \text{and} \quad \widetilde{B} = (B_T B_N).$$

As the columns of C are the fundamental cycles, C_N is the $(m-n+1)\times(m-n+1)$ identity matrix I_{m-n+1}. Since the kernel of B contains all cycle vectors, $BC = 0$ and so $\widetilde{B}C = 0$, giving $B_T C_T = -B_N$. Now, B_T is invertible, as the reader should check (cf. Exercise 40), so

$$C_T = -B_T^{-1} B_N.$$

After this preparation we can easily solve our equations.

Theorem 11 *The electric current* $\mathbf{w}$ *satisfying* $\mathbf{p} = R\mathbf{w} + \mathbf{g}$ *is given by* $\mathbf{w} = -C(C^t RC)^{-1} C^t \mathbf{g}$.

Proof. Equation (1) implies that $B_T \mathbf{w}_T + B_N \mathbf{w}_N = \mathbf{0}$, so $\mathbf{w}_T = -B_T^{-1} B_N \mathbf{w}_N = C_T \mathbf{w}_N$. Hence $\mathbf{w} = C\mathbf{w}_N$. Combining (2) and (3) we find that $C^t R\mathbf{w} + C^t\mathbf{g} = \mathbf{0}$ and so $(C^t RC)\mathbf{w}_N = -C^t\mathbf{g}$. As $C^t RC$ is easily shown to be invertible, the result follows. □

Clearly, Theorem 11 is valid in a somewhat more general situation, not only when G and $\mathbf{g}$ are defined as above. In fact, the following conditions are sufficient (and more or less necessary) for the existence of a unique current: $g_i r_i = 0$ for every i and the edges e_j with $r_j = 0$ form a connected subgraph.

Furthermore, the results hold for multigraphs: all the concepts (incidence matrix, cycle and cut spaces, fundamental cycles and cuts) can be defined as before and the proofs of the results remain unchanged.

By considering multigraphs one can set up Theorem 7 in a slightly simpler form, without adding a new vertex to the graph G' of the network. Thus if the current enters G' at a vertex a and leaves it at a vertex b, then we join a to b by a *new edge* e of 0 resistance (even if a and b had been joined before) and postulate (by choosing $\mathbf{g} = (0, 0, \ldots, 0, 1)$, where e is the last edge) that the potential difference in e is 1. Using this set-up one can check that the ratio of the current in

e_i to the total current (that is, the current in e) is indeed given by Theorem 1 of Section 1, though this checking is rather tedious and involved. On the other hand, as we shall show now, it is very easy to express the total number of spanning trees in a graph in terms of the combinatorial Laplacian.

In fact, let us consider the case of electrical networks with differing resistances and weighted spanning trees, as in Theorem 2. Let then G be a graph with $V(G) = \{v_1, \ldots, v_n\}$ and *conductance matrix* $C = (c_{ij})$: if $i = j$ or $v_i v_j$ is not an edge then c_{ij} is ∞, otherwise it is the conductance of the edge $v_i v_j$.

As in Theorem 2, given a spanning tree T, write $w(T)$ for the product of the conductances of the edges of T, and let $N^*(G) = \sum_T w(T)$, with the summation over all spanning trees.

The *combinatorial Laplacian*, or *Kirchhoff matrix*, of our electrical network is $L = D - C$, where D is the diagonal matrix whose i th diagonal entry is $\sum_{j=1}^n c_{ij} = \sum_{j=1}^n c_{ji}$. As in L all rows and columns sum to 0, all the first cofactors of L are equal: denote by $K^*(G)$ this common value. Here then is the matrix–tree theorem for electrical networks.

Theorem 12 *With the notation above, $N^*(G) = K^*(G)$.*

Proof. We may assume that G is connected, since otherwise $N^*(G) = K^*(G) = 0$. Also, the result is trivial for $n = 1$ since then $N^*(G) = K^*(G) = 1$.

Let us apply induction on the number of edges of G. As the result holds for no edges, we turn to the proof of the induction step. Suppose then that $n > 1$, G is connected, and the assertion holds for networks with fewer edges. Assuming, as we may, that v_1 and v_2 are adjacent, let $G - v_1 v_2$ be obtained from G by cutting (deleting) the edge $v_1 v_2$, and let $G/v_1 v_2$ be obtained from G by fusing (contracting) the edge $v_1 v_2$. Thus in $G/v_1 v_2$ the vertices v_1 and v_2 are replaced by a new vertex, v_{12}, say, which is joined to a vertex v_i, $i > 2$, by an edge of conductance $c_{1i} + c_{2i}$, provided $c_{1i} + c_{2i} > 0$.

The crunch of the proof is that N^* and K^* satisfy the same cut-and-fuse relation:

$$N^*(G) = N^*(G - v_1 v_2) + c_{12} N^*(G/v_1 v_2), \tag{4}$$

and

$$K^*(G) = K^*(G - v_1 v_2) + c_{12} K^*(G/v_1 v_2). \tag{5}$$

Indeed, $N^*(G - v_1 v_2)$ ‘counts’ the spanning trees not containing $v_1 v_2$, and $c_{12} N^*(G/v_1 v_2)$ ‘counts’ the remaining spanning trees. To see (5), simply consider the cofactors belonging to v_1 and v_{12}.

This is all: by the induction hypothesis, the right-hand sides of (4) and (5) are equal. $\square$

A special case of Theorem 12 concerns multigraphs (or even graphs): all we have to do is to write c_{ij} for the number of edges joining v_i to v_j.

Corollary 13 *The number of spanning trees in a multigraph is precisely the common value of the first cofactors of the combinatorial Laplacian.*

A similar result holds for *directed* multigraphs; however, this time we have to count spanning trees oriented towards a vertex, as in Section I.3.

Theorem 14 *Let G be a directed multigraph with vertex set $V(G) = \{v_1, \ldots, v_n\}$. For $1 \leq i \leq n$, denote by $t_i(G)$ the number of spanning trees oriented towards v_i. Also, let $L = (\ell_{ij})$ be the combinatorial Laplacian of G: for $i \neq j$, $-\ell_{ij}$ is the number of edges from v_i to v_j, and $\ell_{ii} = -\sum_{j \neq i} \ell_{ij}$. Then $t_i(G)$ is precisely the first cofactor of L belonging to ℓ_{ii}.*

The proof is entirely along the lines of the proof of Theorem 12: when considering $t_1(G)$, say, all we have to take care is to contract all edges from v_i *to* v_1 for some $i > 1$. Note that this result contains Corollary 13: given a multigraph, replace each edge by two edges, oriented in either direction, and apply Theorem 14.

II.4 Exercises

In exercises 1–7 every graph is taken as a simple electrical network, with every edge having resistance 1.

1. Calculate the resistance of the network shown in Fig. II.1 measured between the vertices 2 and 3.

2. For each different pair of vertices of a cube calculate the resistance between them.

3. What is the resistance between two adjacent vertices of (a) an octahedron, (b) a dodecahedron and (c) an icosahedron?

4. Suppose each edge of a connected network is in the same number of spanning trees. Prove that the total resistance between two adjacent vertices is $(n-1)/e$, where n is the order and e is the size of the network. Verify your answers to Exercise 3.

5. By applying suitable star–triangle transformations, calculate the resistance of a dodecahedron between the midpoints of two adjacent edges.

6. Show that the resistance across an edge of K_n is $2/n$, and so is the resistance between two vertices of $K_{n,m}$ that belong to the second class (having m vertices).

7. Calculate the resistance between two nonadjacent vertices of the complete three-partite graph $K_{n,n,n}$.

8. Give a detailed proof of Theorem 2.

9. Construct the tilings associated with the networks in Fig. II.11.

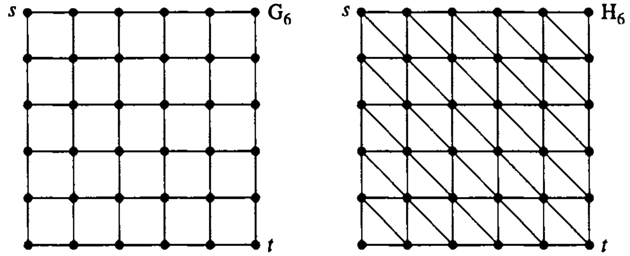


FIGURE II.11. Networks with edges of differing resistances.

10.$^-$ Consider an electrical network on a complete graph. Indicate a simple way of measuring the resistances of the edges by setting the vertices at certain potentials and measuring the currents leaving or entering the network at the vertices.

11. Let M_1 and M_2 be electrical networks, each containing a set U of vertices, the vertices of *attachment*. We say that (M_1, U) is *equivalent to* (M_2, U) if whenever N is a network sharing with each M_i the set U and nothing else, and we set some vertices of N at certain potentials, then in $N \cup M_1$ and $N \cup M_2$ we obtain precisely the same distribution of currents in the edges of N. For $a, b \in U, a \neq b$, let $w_{ab}(M_i, U)$ be the amount of current leaving M_i at b if the vertex a is set at potential 1 and all other vertices of U are set at 0. Show that (M_1, U) is equivalent to (M_2, U) if $w_{ab}(M_1, U) = w_{ab}(M_2, U)$ for all $a, b \in U, a \neq b$.
Use this to verify the star–triangle transformation.

12. (Exercise 11 contd.) Show that a network M, with attachment set U, is equivalent to a network with vertex set U (and attachment set U) if, and only if, $w_{ab}(M, U) = w_{ba}(M, U)$ for all $a, b \in U, a \neq b$.

13.$^+$ Show that every network M with attachment set U is equivalent to a network with vertex set U. [*Hint*. By the result in the previous exercise, it suffices to show that $w_{ab}(M, U) = w_{ba}(M, U)$, where $a, b \in U, a \neq b$. Short all vertices of U, other than a and b, to a vertex c. Let V_x be the potential of a vertex x when we set a at 1, and b and c at 0, and let it be V_x' when we set b at 1, and a and c at 0. For a vertex x, set $P_x = (V_x, V_x') \in \mathbb{R}^2$ and, for each edge xy, let the point P_x pull P_y with a force $c_{xy}(P_x - P_y)$. Note that if $P_a = (1,0)$, $P_b = (0,1)$ and $P_c = (0,0)$ are fixed, then this system is in equilibrium, so the torque at P_c is 0.]

14.$^+$ Show that if the resistance of a wire is increased (in particular, if it is cut) then the total resistance of a network does not decrease, and if a wire is shorted (or just some vertices are shorted) then the total resistance does not increase.

15. Given a multigraph G and an edge e, write $G - e$ for G without the edge e, and G/e for the multigraph obtained by *contracting* the edge e, i.e., for the

graph obtained from $G - e$ by identifying the endvertices of e. Also, for a connected multigraph G and an edge e, write $\mathbb{P}_G(e \in T)$ for the probability that a random spanning tree contains e. Thus

$$\mathbb{P}_G(e \in T) = N(G/e)/N(G),$$

where $N(H)$ denotes the number of spanning trees of a graph H.
Show that the result in the previous exercise is equivalent to the assertion that if e and f are distinct edges of G then

$$\mathbb{P}_{G/f}(e \in T) \le \mathbb{P}_G(e \in T) \le \mathbb{P}_{G-f}(e \in T).$$

16. The n-dimensional (hyper-)cube has vertex set $\{0,1\}^n$, with two sequences $a = (a_i)_1^n, b = (b_i)_1^n \in \{0,1\}^n$ joined by an edge if they differ in exactly one term (so $a_i \ne b_i$ for precisely one suffix i). Show that the resistance across an edge is $(2^n - 1)/(n2^{n-1}) = \frac{2}{n} - \frac{1}{n2^{n-1}} \sim \frac{2}{n}$, and calculate the resistance between two opposite vertices.

17.$^+$ (Exercise 16 contd.) Show that the resistance between any two vertices of the n-dimensional cube is at least $(2^n - 1)/(n2^{n-1}) \sim \frac{2}{n}$ and at most $(n+1)/\binom{n}{2} \sim \frac{2}{n}$.

18. Let G_n be the n by n grid, with s and t in the opposite corners, and let H_n be its diagonal variant, as shown in Fig. II.12. Estimate the total resistance between s and t in the two networks if every edge has resistance 1.

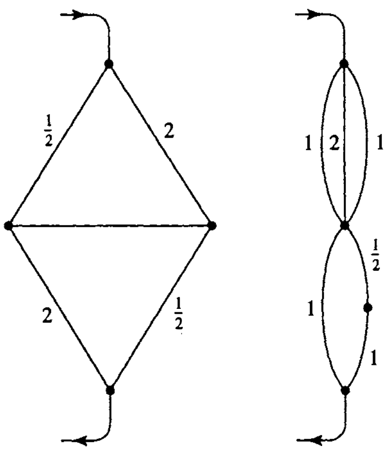


FIGURE II.12. The networks G_6 and H_6.

19. For $k \ge 1$, let L_k, M_k and N_k be the networks indicated in Fig. II.13. Thus M_k has $2k+1$ edges, with resistances $1, 2, \ldots, k, k+1$, and $1, \frac{1}{2}, \ldots, \frac{1}{k}$. For each network, calculate the resistance from s to t, and find the associated tiling.

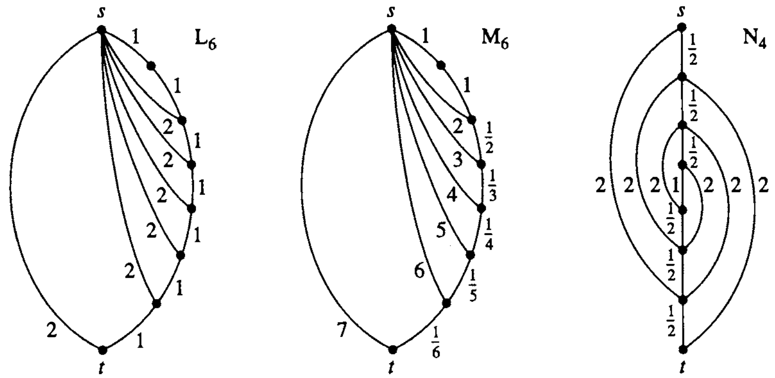


FIGURE II.13. The networks L_6, M_6 and N_4; the numbers indicate the resistances.

20.+ Let s and t be vertices of the boundary of the outer face of a plane network, in which each edge has unit resistance. Suppose that when a non-zero current is put through the network from s to t then there is a non-zero current in each edge. Show that this network corresponds to a squared rectangle.

21.+ Show that there is no perfect squared rectangle of order less than 9 (that is, made up of at most 8 squares).

22.++ Show that there are two essentially different squared rectangles of order 9; the squaring of the 33×32 rectangle in Fig. II.6 and the squaring of the 69×61 rectangle in Fig. II.14.

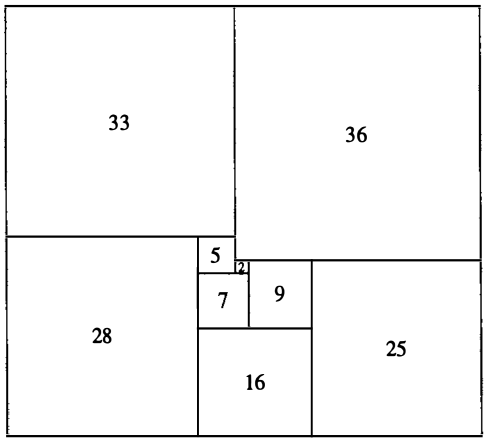


FIGURE II.14. A squaring of the 69×61 rectangle.

23. Find the perfect squared square indicated in Fig. II.15. (This was found by A.J.W. Duijvestijn.)

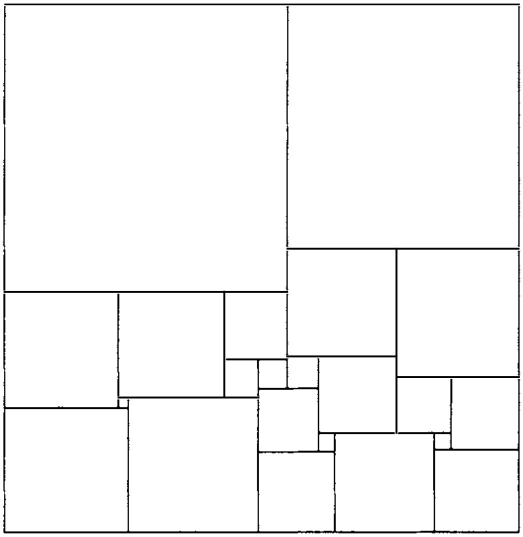


FIGURE II.15. A perfect squaring of the 110×110 square: the largest squares have side lengths 60, 50, 28, 27, 26, 24, 23, 22, 21 and 18.

24. Find the simple perfect squared square given by the network in Fig. II.16. (This example was found by T.H. Willcox.)

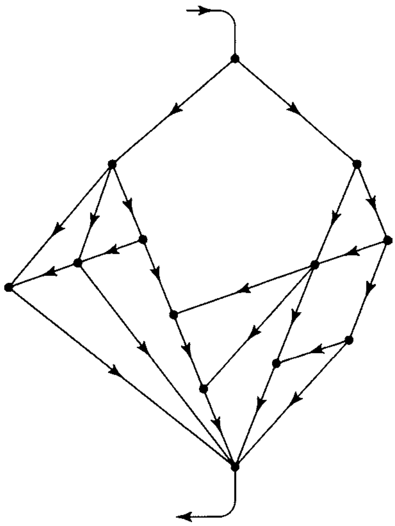


FIGURE II.16. A network giving a perfect squared square: the main square has side length 110, and the constituent squares have side lengths 60, 50, 28, 27, 26, 24, 23, 22, 21, 19, 18, 17, 16, 14, 12, 9, 8, 6, 4, 3, 2 and 1.

25.$^+$ Show that an equilateral triangle cannot be dissected into finitely many incongruent equilateral triangles.

26. Prove that if a rectangular parallelepiped can be decomposed into cubes then the ratios of its sides are rational.

27. Show that a cube cannot be dissected into finitely many incongruent cubes.

28. Dot the i's in the following proof of Theorem 6. For a rectangle $U = [x_1, x_2] \times [y_1, y_2]$, set $\psi(U) = (x_2 - x_1) \otimes (y_2 - y_1) \in \mathbb{Z}(\mathbb{R}/\mathbb{Z}) \otimes \mathbb{Z}(\mathbb{R}/\mathbb{Z})$. Then $\psi(T) = \sum \psi(T_i) = 0$, and so T has an integer side.

29. Fill in the details in the following proof of Theorem 6. Let M be the free $\mathbb{Z}$-module with basis $\mathbb{R}^2/\mathbb{Z}^2$. For a rectangle $U = [x_1, x_2] \times [y_1, y_2] \subset \mathbb{R}^2$, set $\mu(U) = \sum_{i,j=1}^{2} (-1)^{i+j}(x_i, y_j) \in M$. Then $\mu(T) = \sum_{i=1}^{\ell} \mu(T_i) = 0$, so T has an integer side.

30. Prove Theorem 6 in the following way.
(i) Let $p \geq 2$ be a prime. Check that if each T_i has only integer sides and one of them is divisible by p, then one of the sides of T is divisible by p.
(ii) For a prime $p \geq 2$ and $x = (x_1, x_2) \in \mathbb{R}^2$, let $\phi_p(x) = (\lceil px_1 \rceil, \lceil px_2 \rceil) \in \mathbb{Z}^2$. Assume that $T = [0, a] \times [0, b]$. Show that if p is large enough then $\phi_p(T)$ is tiled with $\phi_p(T_1), \ldots, \phi_p(T_\ell)$, where $\phi_p(U)$ is the rectangle whose vertices are the images of the vertices of U under ϕ_p. Apply (i) to this tiling, and deduce Theorem 6.

31. Prove the following extension of Theorem 6. Let $T_1, \ldots, T_\ell$ be rectangles tiling the rectangle $T = [0, a] \times [0, b] \subset \mathbb{R}^2$. Suppose that each T_i has 0, 2 or 4 vertices (corners) in $\mathbb{Z}^2$. Then T has an integer side. [*Hint.* First proof of Theorem 6.]

32. Adapt the second proof of Theorem 6 to prove Theorem 7.

33. Let $T_1, \ldots, T_\ell$ be rectangles contained in a rectangle T such that every point of T that is not on the boundary of some T_i is contained in the same $m \geq 1$ number of rectangles T_i. Show that if each T_i has an integer side then so does T.

34. We know from Corollary 8 that if an $a \times b \times c$ box B in $\mathbb{R}^3$ is filled with $1 \times 2 \times 4$ bricks, then it can also be filled with these bricks all standing in the same way. Prove this as follows. First, note that $8|abc$ and each of ab, bc and ca is even. Hence, we are done unless each of a, b and c is even. Assume then that we are in this case. Replace the box B by an appropriate set of lattice points: $B' = \{(x, y, z) \in \mathbb{Z}^3 : 1 \leq x \leq a, 1 \leq y \leq b, 1 \leq z \leq c\}$. Check that the sum of the coordinates of the points of B' is $\frac{1}{2}abc(a + b + c + 3)$ and that the sum of the coordinates of the points in a box is of the form $8s + 16$. Deduce that at least one of a, b and c is divisible by 4. [This was de Bruijn's original problem he published in a Hungarian journal in 1959; the solution above is his own: it was published in 1960.]

35. Prove the result in the previous exercise in the following way. As before, we may assume that each of a, b and c is even. Divide the box B into $2 \times 2 \times 2$ cubes, and consider a black and white checkerboard colouring of these cubes. Check that each $1 \times 2 \times 4$ brick has exactly as much black as white, and so there are as many black cubes as white ones. Deduce the result from this. [This solution to de Bruijn's problem, given by G. Katona and D. Szász in 1960, is the origin of the 'checkerboard' proof of Theorem 6.]

36. Let $\mathcal{T}$ be the set of tilings of a simply connected domain with 2×1 and 1×2 dominoes. Let H be the graph with vertex set $\mathcal{T}$ in which a tiling T_1 is joined to a tiling T_2 if T_1 and T_2 agree in all but two dominoes. Show that H is connected. Show also that the assertion need not hold if the domain is not simply connected.

37. Show that in a plane graph the boundaries of the bounded faces form a cycle basis.

38. Show that the cycle space is the kernel of the map $C_1(G) \to C_0(G)$ defined by the incidence matrix B.

39. Let B^t be the transpose of the incidence matrix B of a graph G. Show that the cut space is the image of the map $C_0(G) \to C_1(G)$ defined by B^t.

40. Let F be a set of $n - 1$ edges of a graph of order n with incidence matrix B. Let $\widetilde{B}_F$ be an $(n - 1) \times (n - 1)$ submatrix of B whose columns correspond to the edges of F. Prove that $\widetilde{B}_F$ is invertible iff F is the edge set of a tree.

41. Deduce from Corollary 13 that there are n^{n-2} trees on n distinguishable vertices.

42. Which squared rectangle corresponds to the network in Fig. II.17. Rotate the rectangle through 90° and draw the network for this rectangle.

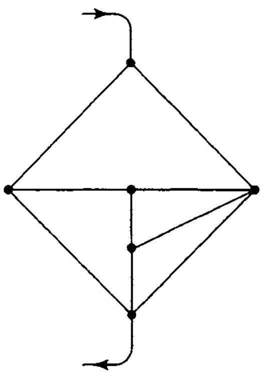


FIGURE II.17. A plane network.

43. How many essentially different squared rectangles correspond to the network of the cube in Exercise 2?

44. Show that a graph is planar if, and only if, its cycle space has a basis of cycles such that every edge belongs to at most two of these cycles.

45. Given a tiling of a rectangle by rectangles, write S for the number of *segments*: the number of maximal segments that are unions of some sides of the rectangles, T for the number of tiles, and C for the number of *crosses*: the number of points in four tiles (see Fig. II.18). Prove that $S - T + C = 3$. [*Hint.* Let G be the plane graph of the tiling, with n vertices, m edges and $f = T + 1$ faces. Write n_i for the number of vertices of degree i so that $n_2 = 4$ and $n_4 = C$. Check that $2m = 8 + 3n_3 + 4C$ and $S = n_3/2 + 4$. Apply Euler's formula.]

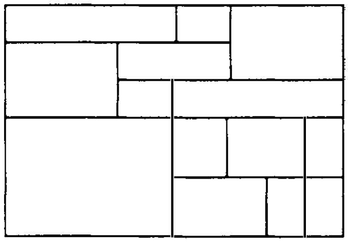


FIGURE II.18. A tiling of a rectangle by rectangles, with $S = 15$, $T = 14$ and $C = 2$.

46. Note that in every triangulation of a convex n-gon there are at least $2n - 3$ segments that occur twice among the sides of the triangles of the triangulation and the sides of the original n-gon. Show that the same holds for every *tiling* of a convex n-gon with triangles, as in Fig. II.19. [*Hint.* Suppose that our tiling is made up of T *triangles*, and there are t segments that occur *twice*. The polygon and the triangles have, altogether, $n + 3T$ sides, so $s = n + 3T - 2t$ sides occur once ('singly'). Suppose also that there are b *boundary* vertices, i.e., vertices of the triangles that are also on a side of a triangle or of the n-gon. Check that $s \leq 3b$ so $n + 3T = 2t + s \leq 2t + 3b$. Counting angles, check that $T \geq b + n - 2$, and deduce the assertion.]

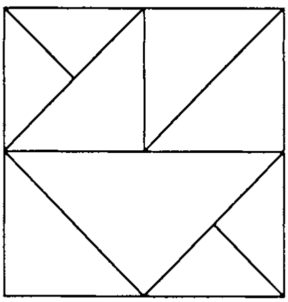


FIGURE II.19. A tiling of a square with triangles; the parameters are $n = 4$, $T = 9$, $t = 5$, $s = 21$ and $b = 7$.

47. (Exercise 46 contd.) Show that in a tiling of a convex n-gon with triangles there are precisely $2n - 3$ segments that occur twice among the sides of the triangles and the n-gon if, and if only if,
    (i) every vertex (of a triangle) in the interior of the n-gon is in the interior of a side of a triangle,
    (ii) if a segment is a union of sides then it is itself a side.

II.5 Notes

The origin of the fundamental results on the distribution of currents, Theorems 1 and 2, is G. Kirchhoff, Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der Linearen Vertheilung galvanischer Ströme geführt wird, *Ann. Phys. Chem.* **72** (1847) 497–508.

Theorem 4 is one of the simplest results from M. Dehn, Über die Zerlegung von Rechtecken in Rechtecke, *Math. Ann.* **57** (1903) 314–322; its extension, Theorem 5, is from R. Sprague, Über die Zerlegung von Rechtecken in lauter verschiedene Quadrate, *J. für die Reine und Angewandte Mathematik* **182** (1940) 60–64.

The first perfect squared squares were published independently by R. Sprague, Beispiel einer Zerlegung des Quadrats in lauter verschiedene Quadrate, *Math. Zeitschrift* **45** (1939) 607–608, and by R.L. Brooks, C.A.B. Smith, A.H. Stone and W. T. Tutte, The dissection of rectangles into squares, *Duke. Math. J.* **7** (1940) 312–340. The square shown in Fig. II.8 was published in A.J.W. Duijvestijn, Simple perfect square of lowest order, *J. Combinatorial Theory Ser. B* **25** (1978) 240–243.

Two survey articles the reader may wish to look at are W.T. Tutte, The quest of the perfect square, *Amer. Math. Monthly* **72** (1965) 29–35 and N.D. Kazarinoff and R. Weitzenkamp, Squaring rectangles and squares, *Amer. Math. Monthly* **80** (1973) 877–888. A recent compendium of squaring results is a privately published volume by J.D. Skinner II, *Squared Squares – Who is Who, and What is What*, Lincoln, Nebraska, 1993, 167 pp.

The origins of the results of de Bruijn at the end of Section 2 are two problems he published in the Hungarian *Matematikai Lapok* in 1959 and 1961; the material presented is from N.G. de Bruijn, Filling boxes with bricks, *Amer. Math. Monthly* **76** (1969) 37–40. For a rich variety of proofs and generalizations of these results, see S. Wagon, Fourteen proofs of a result about tiling a rectangle, *Amer. Math. Monthly* **94** (1987) 601–617.

III

Flows, Connectivity and Matching

Given a collection of boys and girls, when can all the girls find husbands that they know? For a subgroup H of a finite group G, are there group elements $g_1, \dots, g_n$ such that $\{g_1H, \dots, g_nH\}$ is the collection of left cosets and $\{Hg_1, \dots, Hg_n\}$ is the collection of right cosets of H? Given sets $A_1, \dots, A_m$, are there distinct elements $a_1 \in A_1, \dots, a_m \in A_m$?

These seemingly disparate questions are, in fact, closely related: they all concern sets of independent edges, called *matchings*, in bipartite graphs, and are answered by the same basic theorem in various guises, attributed to Hall, König and Egerváry. This theorem, which we shall call *Hall's marriage theorem*, is a prime example of several results we shall present in this chapter giving necessary and sufficient conditions for the existence of certain objects; in each case the beauty of the theorem is that a condition whose necessity is obvious is shown to be also sufficient. In the natural formulation of our results we shall have two functions, say f and g, clearly satisfying $f \leq g$, and we shall show that $\max f = \min g$. The results of this chapter are closely interrelated, and so the order they are proved in is a matter of taste; to emphasize this, some results will be given several proofs.

In the previous chapter we discussed flows in electrical networks: in Section 1 of this chapter we shall study rather different aspects of flows in directed graphs. Our main aim is to present the simple but very powerful max-flow min-cut theorem of Ford and Fulkerson, proved in 1962. This result not only implies the central results of the next two sections, but it also has a number of other important consequences concerning undirected graphs.

Connectivity of graphs is our theme in the second section: the main result is Menger's theorem, first proved in 1927. Hall's marriage theorem and its variants are presented in Section 3.

In the first instance we shall deduce the theorems of both Menger and Hall from the max-flow min-cut theorem. However, as these results are closely related, and are of fundamental importance, we shall also give independent proofs of each.

Hall's theorem tells us, in particular, when a bipartite graph has got a *1-factor*, a subgraph whose vertex set is that of the original graph and in which every vertex has degree 1. The question of the existence of a 1-factor in an arbitrary graph is considerably harder. It is answered by the theorem of Tutte we shall present in Section 4.

The last section is about so-called *stable matchings* in bipartite graphs. These are matchings which are compatible with 'preferences' at all the vertices: in a well defined sense, such a matching is a local maximum for every pair of vertices, one from each class. The fundamental result is a theorem of Gale and Shapley, proved in 1962: this result is not only of great interest in its own right, but it also has numerous applications. Some of these applications will be given here; another important recent application, to list colourings, will be given in Chapter V.

III.1 Flows in Directed Graphs

Let $\vec{G}$ be a (finite) directed graph with vertex set V and edge set $\vec{E}$. We shall study (static) flows in $\vec{G}$ from a vertex s (the *source*) to a vertex t (the *sink*). A *flow* f is a non-negative function defined on the edges; the value $f(\vec{xy})$ is the amount of *flow* or *current* in the edge $\vec{xy}$. For notational simplicity we shall write $f(x, y)$ instead of $f(\vec{xy})$ and a similar convention will be used for other functions. Also, we take $f(x, y)$ to be 0 whenever $\vec{xy} \notin \vec{E}$. The only condition a flow from s to t has to satisfy is Kirchhoff's current law: *the total current flowing into each intermediate vertex* (that is, vertex different from s and t) *is equal to the total current leaving the vertex*. Thus if for $x \in V$ we put

$$\Gamma^+(x) = \{y \in V : \vec{xy} \in \vec{E}\},$$
$$\Gamma^-(x) = \{y \in V : \vec{yx} \in \vec{E}\},$$

then a *flow from s to t* satisfies the following condition:

$$\sum_{y\in\Gamma^+(x)} f(x, y) = \sum_{z\in\Gamma^-(x)} f(z, x)$$

for each $x \in V - \{s, t\}$. Since

$$\begin{aligned} 0 &= \sum_{x\in V-\{s,t\}} \left\{ \sum_{y\in\Gamma^+(x)} f(x, y) - \sum_{z\in\Gamma^-(x)} f(z, x) \right\} \\ &= \sum_{u\in\{s,t\}} \left\{ \sum_{z\in\Gamma^-(u)} f(z, u) - \sum_{y\in\Gamma^+(u)} f(u, y) \right\}, \end{aligned}$$

we find that

$$\sum_{y \in \Gamma^+(s)} f(s, y) - \sum_{y \in \Gamma^-(s)} f(y, s) = \sum_{y \in \Gamma^-(t)} f(y, t) - \sum_{y \in \Gamma^+(t)} f(t, y).$$

In other words, the net current leaving s equals the net current flowing into t. The common value, denoted by $v(f)$, is called the *value of* f or the *amount of flow* from s to t.

We wish to determine the maximal flow value from s to t provided the flow satisfies certain constraints. First we shall deal with the case when the so called *capacity* of an edge restricts the current through the edge. It will turn out that several other seemingly more complicated restrictions can be reduced to this case.

Let us fix our directed graph $\vec{G} = (V, \vec{E})$ and two vertices in it, say s and t. With each edge $\overrightarrow{xy}$ of $\vec{G}$ we associate a non-negative number $c(x, y)$, called the *capacity* of the edge. We shall assume that the current flowing through the edge $\overrightarrow{xy}$ cannot be more than the capacity $c(x, y)$.

Given two subsets X, Y of V, we write $\vec{E}(X, Y)$ for the set of directed $X - Y$ edges:

$$\vec{E}(X, Y) = \{\overrightarrow{xy} \in \vec{E} : x \in X,\ y \in Y\}.$$

Whenever $g : \vec{E} \to \mathbb{R}$ is a function, we put

$$g(X, Y) = \sum g(x, y),$$

where the summation is over $\vec{E}(X, Y)$. If S is a subset of V containing s but not t then $\vec{E}(S, \overline{S})$ is called a *cut* separating s from t. Here $\overline{S} = V - S$ is the complement of S. If we delete the edges of a cut then no positive-valued flow from s to t can be defined on the remainder. Conversely, it is easily seen that if F is a set of edges after whose deletion there is no flow from s to t (that is, $v(f) = 0$ for every flow from s to t) then F contains a cut (Exercise 1). The *capacity* of a cut $\vec{E}(S, \overline{S})$ is $c(S, \overline{S})$ (see Fig. III.1). It is easily seen (Exercise 2) that the capacity of a cut is at least as large as the value of any flow, so the minimum of all cut capacities is at least as large as the maximum of all flow values. The celebrated max-flow min-cut theorem of Ford and Fulkerson states that this trivial inequality is, in fact, an equality. Before stating this theorem and getting down to the proof, let us justify the above use of the words 'minimum' and 'maximum'. Since there are only finitely many cuts, there is a cut whose capacity is minimal. The existence of a flow with maximal value is only slightly less trivial. Indeed, rather crudely,

$$v(f) \leq \sum_{\overrightarrow{xy} \in \vec{E}} c(x, y)$$

for every flow f, so $v = \sup v(f) < \infty$. Let $f_1, f_2, \ldots$ be a sequence of flows with $\lim_n v(f_n) = v$. Then, by passing to a subsequence, we may assume that for each $\overrightarrow{xy} \in \vec{E}$ the sequence $(f_n(x, y))$ is convergent, say to $f(x, y)$. The function f is a flow with value v, that is, a flow with maximal value. In a similar way one

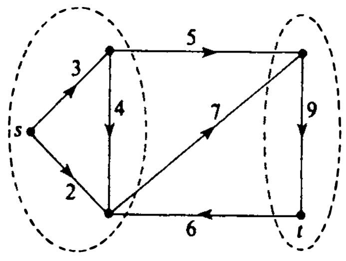


FIGURE III.1. A cut with capacity 12. (The numbers next to the edges indicate their capacity.)

can show that even if some of the edges have infinite capacity, there is a flow with maximal value which can be either finite or infinite (Exercise 3).

Theorem 1 (Max-Flow Min-Cut Theorem.) *The maximal flow value from s to t is equal to the minimum of the capacities of cuts separating s from t.*

Proof. We have remarked already that there is a flow f with maximal value, say v, and the capacity of every cut is at least v. Thus, in order to prove the theorem we have to show that there is a cut with capacity v. We shall, in fact, do considerably more than this: we shall give a very simple procedure for constructing such a cut from a flow f with maximal value.

Define a subset $S \subset V$ recursively as follows. Let $s \in S$. If $x \in S$, and

$$c(x, y) > f(x, y)$$

or

$$f(y, x) > 0,$$

then let $y \in S$.

We claim that $\vec{E}(S, \overline{S})$ is a cut separating s from t with capacity $v = v(f)$. Let us see first why t cannot belong to S. If t belongs to S, we can find vertices $x_0 = s, x_1, \dots, x_\ell = t$ such that

$$\varepsilon_i = \max\{c(x_i, x_{i+1}) - f(x_i, x_{i+1}), f(x_{i+1}, x_i)\} > 0$$

for every i, $0 \le i \le l - 1$. Put $\varepsilon = \min_i \varepsilon_i$. Then f can be augmented to a flow f^* in the following way: if $\varepsilon_i > f(x_{i+1}, x_i)$ then increase the flow in $\overrightarrow{x_i x_{i+1}}$ by ε; otherwise, decrease the flow in $\overrightarrow{x_{i+1} x_i}$ by ε. Clearly, f^* is a flow and its value is $v(f^*) = v(f) + \varepsilon$, contradicting the maximality of f. This shows that $t \notin S$ so $\vec{E}(S, \overline{S})$ is a cut separating s from t.

Now, $v(f)$ is equal to the value of the flow from S to $\overline{S}$ defined in the obvious way:

$$\sum_{x \in S, y \in \overline{S}} f(x, y) - \sum_{x \in \overline{S}, y \in S} f(x, y).$$

By the definition of S the first sum is exactly

$$\sum_{x \in S, y \in \overline{S}} c(x, y) = c(S, \overline{S}),$$

and each summand in the second sum is zero. Hence $c(S, \overline{S}) = v(f)$, as required. $\square$

The max-flow min-cut theorem is the cornerstone of the theory to be presented in this chapter. Note that the theorem remains valid (with exactly the same proof) if some of the edges have infinite capacity but the maximal flow value is finite.

The above proof of the theorem also provides a surprisingly efficient *algorithm* for finding a flow with maximal value if the capacity function is *integral*, that is, if $c(x, y)$ is an integer for every edge $\vec{xy}$. We start with the identically zero flow: $f_0(x, y) = 0$ for every $\vec{xy} \in \vec{E}$. We shall construct an increasing sequence of flows $f_0, f_1, f_2, \ldots$ that has to terminate in a maximal flow. Suppose we have constructed f_i. As in the proof above, we find the set S belonging to f_i. Now, if $t \notin S$ then f_i is a maximal flow (and $E(S, \overline{S})$ is a minimal cut) so we terminate the sequence. If, on the other hand, $t \in S$, then f_i can be augmented to a flow f_{i+1} by increasing the flow along a path from s to t, precisely as in the proof. Since each $v(f_i)$ is an integer, we have $v(f_{i+1}) \geq v(f_i) + 1$, and the sequence must end in at most $\sum_{x,y} c(x, y)$ steps.

Moreover, if c is integral then the algorithm constructs a maximal flow which is also integral, that is, a flow whose value is an integer in every edge. Indeed, f_0 is integral, and if f_i is integral then so is f_{i+1}, since it is obtained from f_i by increasing the flow in a path by a value that is the minimum of a set of positive integers. This result is often called the integrality theorem.

Theorem 2 *If the capacity function is integral then there is a maximal flow that is also integral.* $\square$

We shall rely on this simple result when we use the max-flow min-cut theorem to find various paths in graphs. It is important to note that the results do not claim uniqueness: the algorithm finds one of the maximal flows (usually there are many), and Theorem 2 claims that one of the maximal flows is integral.

The existence of the algorithm proves some other intuitively obvious results as well. For instance, there is a maximal *acyclic* flow, that is, one that does not contain a flow around a cycle (see Exercise 4); in other words, for no cycle $x_1 x_2 \cdots x_k$ do we have

$$f(x_1, x_2) > 0, \ f(x_2, x_3) > 0, \ \ldots, \ f(x_{k-1}, x_k) > 0, \ f(x_k, x_1) > 0.$$

Just as in the case of electrical networks, if instead of one source and one sink we take several of each, the problem becomes only a little more complicated. In fact, the only difference is that we have to be careful when we define a cut. If $s_1, \ldots, s_k$ are the sources and $t_1, \ldots, t_l$ are the sinks then $\vec{E}(S, \overline{S})$ is a *cut* if $s_i \in S$ and $t_j \in \overline{S}$ for every i, j, $1 \leq i \leq k$, $1 \leq j \leq l$.

In order to be able to apply the max-flow min-cut theorem, let us add a new source s and a new sink t to $\vec{G}$, together with all the edges $\overrightarrow{ss_i}$ and $\overrightarrow{t_jt}$, each having *infinite capacity*. Let $\vec{H}$ be the graph obtained in this way. Consider those flows from $s_1, \ldots, s_k$ to $t_1, \ldots, t_l$ in $\vec{G}$ in which the total current entering (leaving) a source (sink) is not greater than the total current leaving (entering) it. These flows can easily be extended to a flow from s to t in $\vec{H}$, and this extension establishes a 1-to-1 correspondence between the two sets of flows. Furthermore, a cut separating s from t in $\vec{H}$ that has finite capacity cannot contain an edge of the form $\overrightarrow{ss_i}$ or $\overrightarrow{t_jt}$ so it corresponds to a cut of the same capacity in $\vec{G}$, separating $s_1, \ldots, s_k$ from $t_1, \ldots, t_l$. Thus Theorem 1 has the following extension.

Theorem 3 *The maximum of the flow value from a set of sources to a set of sinks is equal to the minimum of the capacities of cuts separating the sources from the sinks.* □

Let us assume now that we have capacity restrictions on the *vertices*, except for the source and the sink. Thus we are given a function $c : V - \{s, t\} \to \mathbb{R}^+$ and every flow f from s to t has to satisfy the following inequality for every $x \in V - \{s, t\}$:

$$\sum_{y \in \Gamma^+(x)} f(x, y) = \sum_{z \in \Gamma^+(x)} f(z, x) \leq c(x).$$

How should we define a cut in this situation? A *cut* is a subset S of $V - \{s, t\}$ such that no positive-valued flow from s to t can be defined on $G - S$. In order to distinguish the two kinds of cuts, we sometimes call this a vertex-cut and the other one an edge-cut. However, it is almost always clear which cut is in question. Can we carry over the max-flow min-cut theorem to this case? Yes, very easily, if we notice that a flow can be interpreted to flow in a vertex as well, namely from the part where all the currents enter it to the part where all the currents leave it. More precisely, we can turn each vertex of $\vec{G}$ into an edge (without changing the nature of the directed graph) in such a way that any current entering (and leaving) the vertex will be forced through the edge. To do this, replace each vertex $x \in V - \{s, t\}$ by two vertices, say x_- and x_+; send each incoming edge to x_- and send each outgoing edge out of x_+. Finally, for each x, add an edge from x_- to x_+ with capacity $c(x_-, x_+) = c(x)$ (see Fig. III.2).

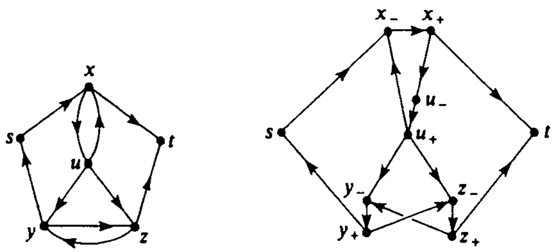


FIGURE III.2. Replacing a graph $\vec{G}$ with restrictions on the capacity of the vertices by a graph $\vec{H}$ with restrictions on the capacity of the edges.

There is a simple 1-to-1 correspondence between the flows from s to t in $\vec{G}$ and the flows in the new graph $\vec{H}$ satisfying the capacity restrictions on (some of) the edges. Since in $\vec{H}$ only the edges $\overrightarrow{x_-x_+}$ have finite capacities, an edge-cut of finite capacity in $\vec{H}$ consists entirely of edges of the form $\overrightarrow{x_-x_+}$, so it corresponds to a vertex-cut in $\vec{G}$ of the same capacity. Thus we have the following form of Theorem 1.

Theorem 4 *Let $\vec{G}$ be a directed graph with capacity bounds on the vertices other than the source s and the sink t. Then the minimum of the capacity of a vertex-cut is equal to the maximum of the flow value from s to t.* □

Theorems 1, 3 and 4 can easily be combined into a single theorem. We leave this to the reader (Exercise 6).

III.2 Connectivity and Menger's Theorem

Recall that a graph is *connected* if any two of its vertices can be joined by a path, and otherwise it is *disconnected*. A maximal connected subgraph of a graph G is a *component* of G.

If G is connected and, for some set W of vertices or edges, $G - W$ is disconnected, then we say that W *separates* G. If in $G - W$ two vertices s and t belong to different components then W *separates* s *from* t. For $k \geq 2$, we say that a graph G is *k-connected* if either G is a complete graph K_{k+1} or else it has at least $k+2$ vertices and no set of $k-1$ vertices separates it. Similarly, for $k \geq 2$, a graph G is *k-edge-connected* if it has at least two vertices and no set of at most $k-1$ edges separates it. A connected graph is also said to be *1-connected* and *1-edge-connected*. The maximal value of k for which a connected graph G is k-connected is the *connectivity* of G, denoted by $\kappa(G)$. If G is disconnected, we put $\kappa(G) = 0$. The *edge-connectivity* $\lambda(G)$ is defined analogously.

Clearly, a graph is 2-connected iff it is connected, has at least 3 vertices and contains no cutvertex. Similarly, a graph is 2-edge-connected iff it is connected, has at least 2 vertices and contains no bridge. It is often easy to determine the connectivity of a given graph. Thus if $1 \leq l \leq n$ then $\kappa(P_\ell) = \lambda(P_\ell) = 1$, $\kappa(C_n) = \lambda(C_n) = 2$, $\kappa(K_n) = \lambda(K_n) = n-1$ and $\kappa(K_{\ell,n}) = \lambda(K_{\ell,n}) = \ell$. In order to correct the false impression that the vertex-connectivity is equal to the edge-connectivity, note that if G is obtained from the disjoint union of two complete graphs K_ℓ by adding a new vertex x and joining x to every old vertex, then $\kappa(G) = 1$, since x is a cutvertex, but $\lambda(G) = \ell$ (see also Exercise 11). This last example shows that $\lambda(G - x)$ may be 0 even when $\lambda(G)$ is large. However, it is clear from the definitions that for every vertex x and edge xy we have

$$\kappa(G) - 1 \leq \kappa(G - x) \quad \text{and} \quad \lambda(G) - 1 \leq \lambda(G - xy) \leq \lambda(G).$$

If G is nontrivial (that is, has at least two vertices), then the parameters $\delta(G)$, $\lambda(G)$ and $\kappa(G)$ satisfy the following inequality:

$$\kappa(G) \leq \lambda(G) \leq \delta(G).$$

Indeed, if we delete all the edges incident with a vertex, the graph becomes disconnected, so the second inequality holds. To see the other inequality, note first that if G is complete then $\kappa(G) = \lambda(G) = |G| - 1$, and if $\lambda(G) \leq 1$ then $\lambda(G) = \kappa(G)$. Suppose now that G is not complete, $\lambda(G) = k \geq 2$ and $\{x_1y_1, x_2y_2, \ldots, x_ky_k\}$ is a set of edges disconnecting G. If $G - \{x_1, x_2, \ldots, x_k\}$ is disconnected then $\kappa(G) \leq k$. Otherwise, each vertex x_i has degree at most k (and so exactly k), as shown in Fig. III.3. Deleting the neighbours of x_1, we disconnect G. Hence $\kappa = \lambda(G)$.

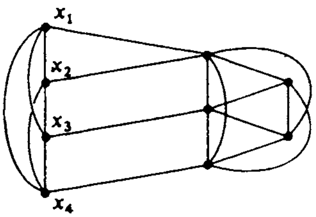


FIGURE III.3. A 4-edge-connected graph G such that $G - \{x_1, x_2, x_3, x_4\}$ is connected.

Another property immediate from the definition of vertex-connectivity is that for $k \geq 1$, if G_1 and G_2 are k-connected subgraphs of a graph G having at least k common vertices, then $G_1 \cup G_2$ is also k-connected. Indeed, if $W \subset V(G_1) \cup V(G_2)$ has at most $k-1$ vertices, then there is a vertex x in $(V(G_1) \cap V(G_2)) \setminus W$. Therefore, the connected subgraphs $G_1 - W$ and $G_2 - W$ of G have at least one vertex, namely x, in common, so $G_1 \cup G_2 - W = (G_1 - W) \cup (G_2 - W)$ is connected.

Having seen in Chapter I how useful it is to partition a graph into its components, that is, into its maximal connected subgraphs, let us attempt a similar decomposition using all maximal 2-connected subgraphs. A subgraph B of a graph G is a *block of* G if either it is a bridge (together with the vertices incident with the bridge) or else it is a maximal 2-connected subgraph of G. The remarks above show that any two blocks have at most one vertex in common, and if x, y are distinct vertices of a block B then $G - E(B)$ contains no x–y path. Therefore, every vertex belonging to at least two blocks is a cutvertex of G, and, conversely, every cutvertex belongs to at least two blocks. Recalling that a cycle is 2-connected and an edge is a bridge iff no cycle contains it, we find that G decomposes into its blocks $B_1, B_2, \ldots, B_p$ in the following sense:

$$E(G) = \bigcup_1^p E(B_i), \quad \text{and} \quad E(B_i) \cap E(B_j) = \emptyset \quad \text{if } i \neq j.$$

Suppose now that G is a nontrivial connected graph. Let $bc(G)$ be the graph whose vertices are the blocks and cutvertices of G and whose edges join cutvertices to blocks: each cutvertex is joined to the blocks containing it. Then $bc(G)$, called the *block–cutvertex* graph of G, is a tree. Each endvertex of $bc(G)$, is a block of G, called an *endblock* of G. If G is 2-connected or is a K_2 (an "edge") then it

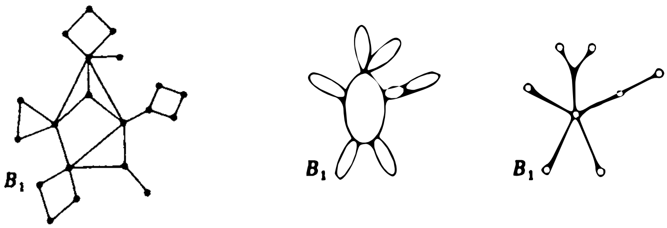


FIGURE III.4. The construction of the block–cutvertex tree $bc(G)$. The subgraph B_1 is an endblock.

contains only one block, namely itself: otherwise there are at least two endblocks, and a block is an endblock iff it contains exactly one cutvertex (Fig. III.4).

The basic result in the theory of connectivity was proved by Menger in 1927. It is the analogue of the max-flow min-cut theorem for (undirected) graphs. Recall that two s–t paths are *independent* if they have only the vertices s and t in common.

Theorem 5 (i) *Let s and t be distinct nonadjacent vertices of a graph G. Then the minimal number of vertices separating s from t is equal to the maximal number of independent s–t paths.*

(ii) *Let s and t be distinct vertices of G. Then the minimal number of edges separating s from t is equal to the maximal number of edge-disjoint s–t paths.*

Proof. (i) Replace each edge xy of G by two *directed* edges, $\overrightarrow{xy}$ and $\overrightarrow{yx}$, and give each vertex other than s and t capacity 1. Then by Theorem 4 the maximal flow value from s to t is equal to the minimum of the capacity of a cut separating s from t. By the integrality theorem (Theorem 2) there is a maximal flow with current 1 or 0 in each edge. Therefore, the maximal flow value from s to t is equal to the maximal number of independent s–t paths. The minimum of the cut capacity is clearly the minimal number of vertices separating s from t.

(ii) Proceed as in (i), except instead of restricting the capacity of the vertices, give each directed edge capacity 1. □

The two parts of the above theorem are called the *vertex form* and the *edge form* of Menger's theorem. One can easily deduce the edge form from the vertex form (Exercise 15), but the other implication is not so easy. Since, as we have mentioned already, the max-flow min-cut theorem can also be deduced from Menger's theorem, we shall give another proof of the vertex form of Menger's theorem from first principles.

Second Proof of the Vertex Form of Menger's Theorem. Denote by k the minimal number of vertices separating s and t. Then clearly there are at most k independent s–t paths and for $k \leq 1$ *there are* k independent s–t paths.

Suppose the theorem fails. Take the minimal $k \geq 2$ for which there is a counterexample to the theorem and let G be a counterexample (for this minimal k) with the minimal number of edges. Then there are at most $k-1$ independent s–t paths and no vertex x is joined to both s and t, otherwise, $G - x$ would be a counterexample for $k-1$.

Let W be a set of k vertices separating s from t. Suppose neither s nor t is adjacent to every vertex in W. Let G_s be obtained from G by replacing the component of $G - W$ containing s by a single vertex s' and joining s' to each vertex in W. In G_s we still need k vertices to separate s' from t, and since the component we collapsed had at least two vertices, G_s has fewer edges than G. Now, as G is a counterexample of minimal size, in G_s there are k independent s'–t paths. The segments of these k paths from t to W are such that any two of them have nothing but the vertex t in common. In particular, for every $w \in W$ one of these paths is a t–w path. If we carry out the analogous procedure for t instead of s then we get k paths from s to W. These two sets of paths can be put together to give k independent s–t paths, contradicting our assumption. Hence for any set W of k vertices separating s from t *either s or t is adjacent to all vertices of W*.

Let $sx_1x_2 \cdots x_lt$ be a shortest s–t path. Then $l \geq 2$ and, by the minimality of G, in the graph $G - x_1x_2$ we can find a set W_0 of $k-1$ vertices separating s from t. Then both $W_1 = \{x_1\} \cup W_0$ and $W_2 = \{x_2\} \cup W_0$ are k-sets separating s from t. Since t is not joined to x_1, the vertex s is joined to every vertex in W_1. Similarly, s is not joined to x_2, and so t is joined to every vertex in W_2. This implies the contradiction that s and t have at least one common neighbour: every vertex in W_0 is a common neighbour of s and t, and $|W_0| = k - 1 \geq 1$. □

Corollary 6 *For $k \geq 2$, a graph is k-connected iff it has at least two vertices and any two vertices can be joined by k independent paths. Also, for $k \geq 2$, a graph is k-edge-connected iff it has at least two vertices and any two vertices can be joined by k edge disjoint paths.* □

Another characterization of k-connectivity is given in Exercise 12.

Corresponding to the max-flow min-cut theorem for multiple sources and sinks, one has the following version of Menger's theorem. If S and T are arbitrary subsets of vertices of G, then the maximal number of vertex-disjoint (including endvertices!) S–T paths is $\min\{|W| : W \subset V(G), G - W$ has no S–T path$\}$. To see this, add two new vertices to G, say s and t, join s to every vertex in S and t to every vertex in T, and apply Menger's theorem to the vertices s and t in the new graph.

III.3 Matching

Given a finite group G and a subgroup H of index m, can you find m elements of G, say $g_1, g_2, \ldots, g_m$ such that $\{g_1H, g_2H, \ldots, g_mH\}$ is the set of all left cosets of H and $\{Hg_1, Hg_2, \ldots, Hg_m\}$ is the set of all right cosets? A reformulation of this problem turns out to be a special case of the following problem, which arises frequently in diverse branches of mathematics. Given a family $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ of subsets of a set X, can we find m distinct elements of X, one from each A_i? A set $\{x_1, x_2, \ldots, x_m\}$ with these properties (i.e., $x_i \in A_i$, $x_i \neq x_j$ if $i \neq j$) is called a *set of distinct representatives* of the family $\mathcal{A}$. The set system $\mathcal{A}$ is naturally identifiable with a bipartite graph with vertex classes

$V_1 = \mathcal{A}$ and $V_2 = X$ in which $A_i \in \mathcal{A}$ is joined to every $x \in X$ contained in A_i. A system of distinct representatives is then a *set of m independent edges* (thus each vertex in V_1 is incident with one of these edges). We also say that there is a *complete matching from* V_1 *to* V_2.

It is customary to formulate this problem in terms of marriage arrangements. Given m girls and n boys, under what conditions can we marry off all the girls, provided that we do not want to carry matchmaking so far as to marry a girl to a boy she does not even know?

It is clear that both the max-flow min-cut theorem and Menger's theorem imply a necessary and sufficient condition for the existence of a complete matching. In fact, because of the special features of a bipartite graph, there is a particularly simple and pleasant necessary and sufficient condition.

If there are k girls who know at most $k-1$ boys altogether, then we cannot find suitable marriages for these girls. Equivalently, if there is a complete matching from V_1 to V_2, then for every $S \subset V_1$ there are at least $|S|$ vertices of V_2 adjacent to a vertex in S; that is,

$$|\Gamma(S)| \geq |S|.$$

The result that this necessary condition is also sufficient is usually called *Hall's theorem*. This fundamental theorem was proved by Hall in 1935, but an equivalent form of it had been proved by König and Egerváry in 1931, but both versions follow immediately from Menger's theorem from 1927. We shall give three proofs. The first is based on Menger's theorem or the max-flow min-cut theorem, the other two prove the result from first principles.

Theorem 7 *A bipartite graph G with vertex sets* V_1 *and* V_2 *contains a complete matching from* V_1 *to* V_2 *iff*

$$|\Gamma(S)| \geq |S| \textit{ for every } S \subset V_1.$$

We have already seen that the condition is necessary so we have to prove only the sufficiency.

First Proof. Both Menger's theorem (applied to the *sets* V_1 and V_2 as at the end of Section 2) and the max-flow min-cut theorem (applied to the *directed graph* obtained from G by sending each edge from V_1 to V_2, and giving each vertex capacity 1) imply the following. If G does not contain a complete matching from V_1 to V_2 then there are $T_1 \subset V_1$ and $T_2 \subset V_2$ such that $|T_1| + |T_2| < |V_1|$ and there is no edge from $V_1 - T_1$ to $V_2 - T_2$. Then $\Gamma(V_1 - T_1) \subset T_2$ so

$$|\Gamma(V_1 - T_1)| \leq |T_2| < |V_1| - |T_1| = |V_1 - T_1|.$$

This shows the sufficiency of the condition. □

Second Proof. In this proof, due to Halmos and Vaughn, we shall use the matchmaking terminology. We shall apply induction on $m = |V_1|$, the number of girls. For $m = 1$ the condition is clearly sufficient, so we assume that $m \geq 2$ and the condition is sufficient for smaller values of m.

Suppose first that any k girls $(1 \leq k < m)$ know at least $k + 1$ boys. Then we arrange one marriage arbitrarily. The remaining (sets of) girls and boys still satisfy the condition, so the other $m - 1$ girls can be married off by induction.

Suppose now that for some k, $1 \leq k < m$, there are k girls who know exactly k boys altogether. These girls can clearly be married off by induction. What about the other girls? We can marry them off (again by induction) if they also satisfy the condition, provided that we do not count the boys who are already married. But the condition is satisfied, since if some ℓ girls to be married know fewer than ℓ remaining boys, then these girls together with the first k girls would know fewer than $k + \ell$ boys. □

Third Proof. This proof is due to Rado. Let G be a minimal graph satisfying the condition. It suffices to show that G consists of $|V_1|$ *independent edges.*

If this is not so, then G contains two edges of the form a_1x, a_2x, where $a_1, a_2 \in V_1$, $a_1 \neq a_2$, and $x \in V_2$. Since the deletion of either of these edges invalidates the condition, there are sets $A_1, A_2 \subset V_1$ such that for $i = 1, 2$ we have $|\Gamma(A_i)| = |A_i|$, and a_i is the only vertex of A_i adjacent to x. Then

$$\begin{aligned}|\Gamma(A_1) \cap \Gamma(A_2)| &\geq |\Gamma(A_1 - \{a_1\}) \cap \Gamma(A_2 - \{a_2\})| + 1\\ &\geq |\Gamma(A_1 \cap A_2)| + 1 \geq |A_1 \cap A_2| + 1.\end{aligned}$$

But this implies the following contradiction:

$$\begin{aligned}|\Gamma(A_1 \cup A_2)| &= |\Gamma(A_1) \cup \Gamma(A_2)|\\ &= |\Gamma(A_1)| + |\Gamma(A_2)| - |\Gamma(A_1) \cap \Gamma(A_2)|\\ &\leq |A_1| + |A_2| - |A_1 \cap A_2| - 1\\ &= |A_1 \cup A_2| - 1.\end{aligned}$$ □

A regular bipartite graph satisfies the conditions of Hall's theorem, so it has a complete matching. In turn this implies that we can indeed find group elements $g_1, g_2, \ldots, g_m$, as required at the beginning of the section.

Let us reformulate the marriage theorem in terms of sets of distinct representatives.

Theorem 8 *A family $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ of sets has a set of distinct representatives iff*

$$\left|\bigcup_{i \in F} A_i\right| \geq |F| \text{ for every } F \subset \{1, 2, \ldots, m\}.$$ □

In the next four results we present two natural extensions of the marriage theorem. The first two of the these concern deficient forms of the theorem. Suppose that the marriage condition is not satisfied. How near can we come to marrying off all the girls? When can we marry off all but d of the girls? Clearly, only if any k of them know at least $k - d$ boys. This obvious necessary condition is again sufficient.

Corollary 9 *Suppose that a bipartite graph $G = G_2(m, n)$, with vertex sets V_1, V_2, satisfies the following condition:*

$$|\Gamma(S)| \geq |S| - d$$

for every $S_1 \subset V_1$. Then G contains $m - d$ independent edges.

Proof. Add d vertices to V_2 and join them to each vertex in V_1. The new graph $G*$ satisfies the conditions for a complete matching. At least $m - d$ of the edges in a complete matching of $G*$ belong to G. □

Let us give another deficient form of the marriage theorem. If an edge e is incident with a vertex x, then we say that *e covers x*, and *x covers e*. Furthermore, a vertex is said to *cover* itself (and no other vertex).

Corollary 10 *Let $G = G_2(m, n)$ be a bipartite graph. Write h for the maximal number of independent edges, i for the maximal number of independent vertices, and j for the minimal number of edges and vertices covering all the vertices. Then*

$$i = j = m + n - h.$$

Proof. Let $E' \cup V'$ be a set of j edges and vertices covering all vertices, with $E' \subset E$ and $V' \subset V$. If $e, f \in E'$ share a vertex, then in the cover $E' \cup V'$ we may replace f by its other endvertex. Hence we may assume that E' consists of independent edges. This shows that $j = m + n - h$.

Also, $m + n - i \geq h$, since if I is a set of i independent vertices (in any graph), then every edge is incident with at least one vertex not in I.

Finally, let $S \subset V_1$ be such that $|\Gamma(S)| = |S| - (m - h)$, as guaranteed by Corollary 9. Then, with $T = V_2 - \Gamma(S)$, the set $S \cup T$ is a set of $|S| + n - |\Gamma(S)| = m + n - h$ independent vertices, proving that $i \geq m + n - h$. □

The next extension concerns matchmaking *for boys* in a polygynous country, where the i th boy intends to marry d_i girls.

Corollary 11 *Let G be a bipartite graph with vertex classes $V_1 = \{x_1, \ldots, x_m\}$ and $V_2 = \{y_1, \ldots, y_n\}$. Then G contains a subgraph H such that $d_H(x_i) = d_i$ and $0 \leq d_H(y_j) \leq 1$ iff*

$$|\Gamma(S)| \geq \sum_{x_i \in S} d_i$$

for every $S \subset V_1$.

Proof. Replace each vertex x_i by d_i vertices joined to every vertex in $\Gamma(x_i)$. Then G has such a subgraph H iff the new graph has a matching from the new first vertex class to V_2. The result follows from Theorem 7. □

Of course, Corollary 11 also has a defect form which the reader is encouraged to state and deduce from this.

The alert reader is probably aware of the fact that these corollaries are still special cases of the max-flow min-cut theorem. In fact, the bipartite graph version of the max-flow min-cut theorem is considerably more general than the corollaries above.

Theorem 12 *Let $G = G_2(m, n)$ be a bipartite graph with vertex classes $V_1 = \{x_1, \ldots, x_m\}$ and $V_2 = \{y_1, \ldots, y_n\}$. For $S \subset V_1$ and $1 \le j \le n$ denote by S_j the number of edges from y_j to S. Let $d_1, \ldots, d_m$ and $e_1, \ldots, e_n$ be natural numbers and let $d \ge 0$. Then there exists a subgraph H of G with*

$$e(H) \ge \sum_{i=1}^{m} d_i - d,$$

$$d_H(x_i) \le d_i, \qquad 1 \le i \le m$$

and

$$d_H(y_j) \le e_j, \qquad 1 \le j \le n,$$

iff for every $S \subset V_1$ we have

$$\sum_{x_i \in S} d_i \le \sum_{j=1}^{n} \min\{S_j, e_j\} + d.$$

Proof. Turn G into a directed graph $\vec{G}$ by sending each edge from V_1 to V_2. Give each edge capacity 1, a vertex x_i capacity d_i, and a vertex y_j capacity e_j. Then there is a subgraph H with the required properties iff in $\vec{G}$ there is a flow from V_1 to V_2 with value at least $\sum_1^m d_i - d$, and by the max-flow min-cut theorem, this happens iff every cut has capacity at least $\sum_1^m d_i - d$. Now, minimal cuts are of the form $T \cup U \cup E(V_1 - T, V_2 - U)$, where $T \subset V_1$ and $U \subset V_2$. Given a set T, the capacity of such a cut will be minimal if a vertex y_j belongs to U iff its capacity is smaller than the number of edges from $S = V_1 - T$ to y_j. With this choice of U the capacity of the cut is exactly

$$\sum_{x_i \in T} d_i + \sum_{1}^{n} \min\{S_j, e_j\}.$$

The condition that this is at least $\sum_1^m d_i - d$ is clearly the condition in the theorem. □

The reader is invited to check that the second proof of Theorem 7 can be rewritten word for word to give a proof of the exact form of this result (that is, with $d = 0$) and the defect form (the case $d \ge 0$) can be deduced from it as Corollary 10 was deduced from Theorem 7 (Exercise 33).

To conclude this section we prove another extension of the marriage theorem. This is *Dilworth's theorem* concerning partially ordered sets. A *partial order* $<$ on a set is a transitive and irreflexive relation defined on some ordered pairs of elements. Thus if $x < y$ and $y < z$ then $x < z$, but $x < y$ and $y < x$ cannot both hold. A set with a partial order on it is a *partially ordered set*. The relation $x \le y$ expresses the fact that either $x = y$ or else $x < y$. A subset C of a partially ordered set P is a *chain* (or *tower*) if for $x, y \in C$ either $x \le y$ or $y < x$. A set

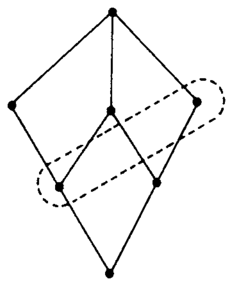


FIGURE III.5. A partially ordered set and a maximal antichain. (An edge indicates that its upper endvertex is greater than its lower endvertex.)

$A \subset P$ is an *antichain* if $x < y$ implies that $\{x, y\} \not\subset A$. See Fig. III.5 for an example.

What is the smallest number of chains into which we can decompose a partially ordered set? Since no two elements of an antichain can belong to the same chain, we need at least as many chains as the maximal size of an antichain. Once again, the trivial necessary condition is, in fact, sufficient.

Theorem 13 *If every antichain in a (finite) partially ordered set P has at most m elements, then P is the union of m chains.*

Proof. Let us apply induction on $|P|$. If $P = \emptyset$, there is nothing to prove, so we suppose that $|P| > 0$ and the theorem holds for sets with fewer elements.

Let C be a maximal chain in P. (Thus if $x \notin C$, then $C \cup \{x\}$ is no longer a chain.) If no antichain of $P - C$ has m elements, then we are home by induction. Therefore, we may assume that $P - C$ contains an antichain $A = \{a_1, a_2, \dots, a_m\}$.

Define the *lower shadow* of A as

$$S^- = \{x \in P : x \leq a_i \text{ for some } i\},$$

and define the *upper shadow* S^+ of A analogously. Then P is the union of the two shadows, since otherwise A could be extended to an antichain with $m+1$ elements. Furthermore, neither shadow is the whole of P, since the maximal element of C does not belong to S^- and the minimal element of C does not belong to S^+. By the induction hypothesis both shadows can be decomposed into m chains, say

$$S^- = \bigcup_{i=1}^{m} C_i^- \quad \text{and} \quad S^+ = \bigcup_{i=1}^{m} C_i^+.$$

Since different a_i belong to different chains, we may assume that $a_i \in C_i^-$ and $a_i \in C_i^+$.

The proof will be completed if we show that a_i is the maximal element of C_i^- and the minimal element of C_i^+, since in that case the chains C_i^- and C_i^+ can be strung together to give a single chain C_i, and then $P = \bigcup_1^m C_i$.

Suppose then that, say, a_i is not the maximal element of C_i^- : $a_i < x$ for some $x \in C_i^-$. Since x is in the lower shadow of A, there is an $a_j \in A$ with $x \leq a_j$. However, this implies the contradiction $a_i < a_j$. □

In fact, Dilworth's theorem holds for all partially ordered sets: we leave this to the reader (Exercise 53).

III.4 Tutte's 1-Factor Theorem

A *factor* of a graph is a spanning subgraph: a subgraph whose vertex set is that of the whole graph. If every vertex of a factor has degree r, then we call it an *r-factor*. How can we characterize graphs with a 1-factor? If G has a 1-factor H and we delete a set S of vertices of G, then in a component C of $G - S$ an even number of vertices are on edges of H contained in C, and the other vertices of C are on edges of H joining a vertex of C to a vertex of S. In particular, for every *odd* component C of $G - S$ (that is, a component with an odd number of vertices) there is an edge of H joining a vertex of C to a vertex of S. Now, the edges of H are independent, so this implies that the graph $G - S$ has *at most* $|S|$ *odd components*, one for each vertex in S (see Fig. III.6).

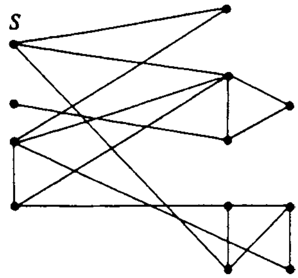


FIGURE III.6. A graph G with a 1-factor: $|S| = 4$ and $G - S$ has 2 odd components.

The necessity of the condition we have just found is rather trivial, but it is not clear at all that the condition is also sufficient. This surprising and deep result was first proved by Tutte in 1947. It will be convenient to denote by $q(H)$ the number of odd components of a graph H, that is, the number of components of odd order.

Theorem 14 *A graph G has a 1-factor iff*

$$q(G - S) \leq |S| \tag{1}$$

for every $S \subset V(G)$.

Proof. We know that the condition is necessary. We shall prove the sufficiency by induction on the order of G. For $|G| = 0$ there is nothing to prove. Now let G be a graph of order at least one satisfying (1) and suppose that the theorem holds for graphs of smaller order.

Suppose that $S_0 \subset V(G)$ is a non-empty set for which equality holds in (1). Denote by $C_1, C_2, \ldots, C_m$, $m = |S_0| \geq 1$, the odd components of $G - S_0$ and let $D_1, D_2, \ldots, D_k$ be the even components of $G - S_0$. If the theorem is true and G does contain a 1-factor F, then for each C_i there is at least one edge of F that joins a vertex of C_i to a vertex in S_0. Since $m = |S_0|$, for each C_i there is exactly one such edge, say $c_i s_i$, $c_i \in C_i$, $s_i \in S_0$. Each $C_i - c_i$ contains a 1-factor (a subgraph of F), and each D_j contains a 1-factor (a subgraph of F). Finally, the edges $s_1c_1, s_2c_2, \ldots, s_mc_m$ form a complete matching from S_0 into the set $\{C_1, C_2, \ldots, C_m\}$.

The proof is based on the fact that one can find an S_0 that has all the properties described above. How shall we find such a set S_0? Let S_0 be a maximal non-empty subset of $V(G)$ for which equality holds in (1). Of course, *a priori* it is not even clear that there is such a set S_0. With $S = \emptyset$ the condition (1) implies that G has even order. If s is any vertex of G, then $G - \{s\}$ has odd order, so it has a least one odd component. Since (1) holds, $G - \{s\}$ has exactly one odd component. Hence for every $S = \{s\}$ we have equality in 1. This establishes the existence of S_0.

As before, let $C_1, C_2, \ldots, C_m$, $m = |S_0|$ be the odd components of $G - S_0$ and $D_1, D_2, \ldots, D_k$ the even components.

(i) *Each D_j has a 1-factor.* Indeed, if $S \subset V(D_j)$ then

$$q(G - S_0) + q(D_j - S) = q(G - S_0 \cup S) \leq |S_0 \cup S| = |S_0| + |S|,$$

so

$$q(D_j - S) \leq |S|.$$

Hence by the induction hypothesis D_j has a 1-factor.

(ii) *If $c \in C_i$, then $C_i - c$ has a 1-factor.* Assume that this is false. Then by the induction hypothesis there is a subset S of $V(C_i) - \{c\}$ such that

$$q(C_i - \{c\} \cup S) > |S|.$$

Since

$$q(C_i - \{c\} \cup S) + |S \cup \{c\}| \equiv |C_i| \equiv 1 \pmod 2,$$

this implies that

$$q(C_i - \{c\} \cup S) \geq |S| + 2.$$

Consequently,

$$\begin{aligned} |S_0 \cup \{c\} \cup S| = |S_0| + 1 + |S| &\geq q(G - S_0 \cup \{c\} \cup S) \\ &= q(G - S_0) - 1 + q(C_i - \{c\} \cup S) \\ &\geq |S_0| + 1 + |S|, \end{aligned}$$

so in (1) we have equality for the set $S_0 \cup \{c\} \cup S$ as well. This contradicts the maximality of S_0.

(iii) *G contains m independent edges of the form s_ic_i, $s_i \in S_0$ and $c_i \in C_i$, $i = 1, 2, \ldots, m$.* To show this, let us consider the bipartite graph $H = G_2(m, m)$

with vertex classes $V_1 = \{C_1, C_2, \dots, C_m\}$ and $V_2 = S_0$, in which C_i is joined to a vertex $s \in S_0$ if and only if G contains an edge from s to C_i. The assertion above is true iff H has a 1-factor, that is, a matching from V_1 to V_2. Fortunately, we have the weapon to check this: Hall's theorem. Given $A \subset V_1$, put $B = \Gamma_H(A) \subset V_2$ (see Fig. III.7). Then (1) implies that

$$|A| \le q(G - B) \le |B|.$$

Hence the graph H satisfies Hall's condition, so it has a 1-factor.

We are almost done. To complete the proof we just put together the information from (i), (ii), and (iii). We start with the m independent edges $s_i c_i$, $s_i \in S_0, c_i \in C_i$. Adding to this set of edges a 1-factor of each $C_i - c_i$, $1 \le i \le m$, and a 1-factor of each D_j, $1 \le j \le k$, we arrive at a 1-factor of G. □

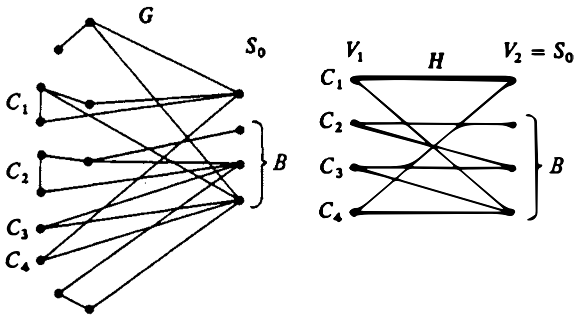


FIGURE III.7. The construction of H from G. The set $A = \{C_2, C_3\}$ determines $B \subset S_0$ by the rule $B = \Gamma_H(A)$.

It is once again very easy to obtain a defect form of the above result.

Corollary 15 *A graph G contains a set of independent edges covering all but at most d of the vertices iff*

$$q(G - S) \le |S| + d$$

for every $S \subset V(G)$.

Proof. Since the number of vertices not covered by a set of independent edges is congruent to $|G|$ modulo 2, we may assume that

$$d \equiv |G| \pmod 2.$$

Put $H = G + K_d$; that is, let H be obtained from G by adding to it a set W of d vertices, and joining every new vertex to every other vertex, old and new. Then G contains a set of independent edges covering all but d of the vertices iff H has a 1-factor. When does (1) hold for H? If $\emptyset \ne S' \subset V(H)$ and $W - S' \ne \emptyset$, then $H - S'$ is connected, so $q(H - S') \le 1$, and then (1) does hold; if $W \subset S'$ then, setting $S = S' - W$, we have $q(H - S') = q(G - \{S' \setminus W\}) = q(G - S)$, so (1) is equivalent to

$$q(G - S) \leq |S'| = |S| + d. \qquad \square$$

Tutte's theorem has numerous beautiful consequences: for example, it implies that every 2-edge-connected cubic graph has a 1-factor (Exercise 32).

III.5 Stable Matchings

Let us return to the problem of finding matchings in bipartite graphs. This time we shall study so-called *stable* matchings, that is, matchings satisfying certain conditions. These matchings were first studied by Gale and Shapley in 1961, and our main aim is to prove their fundamental result. Before we turn to the complete graphs studied by Gale and Shapley, we consider general bipartite graphs.

As in the case of Hall's theorem, it is customary to formulate the conditions and results in terms of marriage arrangements between n boys and m girls. Suppose then that we have an n by m bipartite graph $G = G_2(n, m)$ with bipartition (V_1, V_2), where $V_1 = \{a, b, \dots\}$ is the set of boys and $V_2 = \{A, B, \dots\}$ is the set of girls. For the moment we do not assume that $n = m$, i.e., that we have the same number of girls and boys. As before, an edge aA means that the boy a *knows* the girl A. Suppose that each boy has an order of preferences on the set of girls he knows, and each girl has an order of preferences on the set of boys she knows. We assume that these orders are linear orders but place no other restriction on them. Given the preferences, a *stable matching* in G is a set M of independent edges of G such that if $aB \in E(G) - M$, then either $aA \in M$ for some girl A preferred to B by a, or $bB \in M$ for some boy b preferred to a by B. Thus if a is not married to B, then either a is married to a girl he prefers to B, or else B is married to a boy she prefers to a. Otherwise, a and B could (and eventually would) get married, perhaps divorcing their present spouses, to the benefit of *both*. This makes the (somewhat unrealistic) assumption that it is always better to be married (to an acquaintance) than to stay single.

Note that a stable matching is not assumed to be complete. However, it is clear that every stable matching is a *maximal* matching in G; that is, it cannot be enlarged to a strictly larger matching. Indeed, suppose $M \cup \{aB\}$ is a matching in G for some edge $aB \in E(G) - M$. Then under the marriage arrangement M, the boy a is a bachelor, so he is certainly not married to a girl he prefers to B, and the girl B is a spinster, so she is not married to a boy she prefers to a. This contradicts the fact that M is stable.

Although every stable matching is maximal, it need not be a *maximum* matching; that is, it *need not have maximal cardinality*. A trivial example is shown in Fig. III.8. However, as we shall see later, all stable matchings have the same cardinality.

The stability condition for a matching is fairly complex, so *a priori* it is not clear that there is always a stable matching. In fact, we shall show that not only is there always a stable matching, but there is also a stable matching that is *optimal* for each boy. The existence of an optimal stable matching follows free of charge

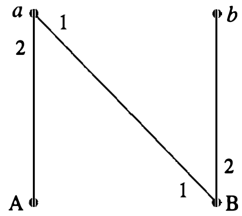


FIGURE III.8. If a prefers B to A, and B prefers a to b, then $M = \{aB\}$ is the only stable matching.

from the algorithm used to construct a stable matching, so to start with we shall not bother with optimality. It is rather quaint that this *fundamental algorithm* is simply the codification of the rules of old-fashioned etiquette: every boy proposes to his highest preference and every girl refuses all but her best proposer. This goes on until no changes occur; then every boy marries the girl to whom he last proposed, and every girl marries her only proposer she has not yet refused.

Note that the algorithm is such that once a girl gets a proposal, at the end of the process she does end up with a husband, for she will refuse a suitor only for somebody she finds more desirable. Also, every boy gets married unless in the algorithm he is refused by every girl. Finally, as the algorithm progresses, every girl gets better and better suitors, and every boy has to be resigned to marrying less and less desirable girls. With this we have come close to proving the stable matching theorem of Gale and Shapley.

Theorem 16 *For every assignment of preferences in a bipartite graph, there is a stable matching.*

Proof. Let us describe a variant of the fundamental algorithm we have just mentioned, in which all boys and all girls act simultaneously, in *rounds*. In every odd round (1 st, 3 rd, ...), each boy proposes to his highest preference among those girls whom he knows and who have not yet refused him, and in every even round (2 nd, 4 th, ...), each of the m girls refuses all but her highest suitor. The process ends when no girl refuses a suitor: then every girl marries her (only) suitor, if she has one.

The algorithm terminates after at most $2nm$ rounds, since at most $m(n-1)$ proposals are refused.

We claim that this fundamental algorithm produces a stable matching. It clearly produces a (partial) matching M, since at every stage each boy proposes to at most one girl, and each girl rejects all but at most one boy. To see that M is stable, let $aB \in E(G) - M$. Then *either* a never proposed to B, *or* a was refused by B during the algorithm. In the former case a marries a girl he prefers to B, as he never goes as low as B, and in the latter case B refused a for a boy she prefers to a, so eventually ends up with a husband she prefers to her suitor a. □

The fundamental algorithm we have just described can be run at various speeds: we do not have to have uniform action, in rounds. Every boy and every girl may act individually: each boy keeps proposing to the best girl (in *his* estimation) who has not yet refused him, and each girl maximizes her satisfaction by being willing to accept only the very best boy (in her estimation) who has ever proposed to her. When the dust settles, we have a stable matching *independent* of the speed at which we have run the algorithm (see Exercise 42).

What can we say about the collection of stable matchings? Somewhat surprisingly, all stable matchings are incident with the same set of vertices; that is, every vertex is either matched in every stable matching or remains unmatched in every stable matching. This will follow easily from the lemma below.

Call a cycle *preference-oriented* if it can be written in the form $aAbB\cdots zZ$ such that A prefers b to a, b prefers B to $A, \ldots,$ and Z prefers a to z.

Lemma 17 *Let M and M' be two stable matchings in a bipartite graph with certain preferences, and let C be a component of the subgraph H formed by the edges of $M \cup M'$. If C has at least three vertices, then it is a preference-oriented cycle. In particular, if aA, $bB \in M$ and $aB \in M'$, then a prefers A to B iff B prefers a to b.*

Proof. In this proof it is best not to distinguish between boys and girls: we shall write $x_1, x_2, \ldots$ for either of them. We know that C is either a path of length at least two or a cycle of length at least four.

Suppose that C contains a path $x_1x_2x_3x_4$, with x_2 preferring x_3 to x_1. Assuming, as we may, that $x_2x_3 \notin M$, we see that x_3 prefers x_4 to x_2, since M is stable.

This simple observation implies that if C is a *cycle*, then it is preference-oriented. Indeed, if $x_1x_2x_3\cdots x_k$ is a cycle and x_2 prefers x_3 to x_1, then looking at the path $x_1x_2x_3x_4$ we see that x_3 prefers x_4 to x_2. Next, looking at the path $x_2x_3x_4x_5$ we see that x_4 prefers x_5 to x_3. Continuing in this way, we find that x_k prefers x_1 to x_{k-1} and x_1 prefers x_2 to x_k.

Also, if C is a *path* $x_1x_2\cdots x_\ell$ with $l \geq 3$ and $x_1x_2 \notin M$, say, then x_2 prefers x_3 to x_1, since M is stable and $x_1x_2 \notin M$. Similarly, $x_{\ell-1}$ prefers $x_{\ell-2}$ to x_ℓ. However, this is impossible, since, arguing as above, x_2 prefers x_3 to x_1, x_3 prefers x_4 to x_2, x_4 prefers x_5 to x_3, and so on, $x_{\ell-1}$ prefers x_ℓ to $x_{\ell-2}$.

The second assertion is immediate from the fact that the component of H containing the path $AaBb$ is a preference-oriented cycle. □

Theorem 18 *For every assignment of preferences in a bipartite graph with bipartition (V_1, V_2), there are subsets $U_1 \subset V_1$ and $U_2 \subset V_2$ such that every stable matching is a complete matching from U_1 to U_2. In particular, all stable matchings have the same cardinality.*

Proof. Suppose that the assertion fails. Then we may assume that some edge aA of M is such that a is not incident with any edge of M'. As M' is a maximal matching, $bA \in M'$ for some $b \in V_1$, $b \neq a$. But then the component of a in the subgraph formed by the edges of $M \cup M'$ contains a, A, and b, and is not a cycle, contradicting Lemma 17. □

It is tempting to expect that every stable matching of the subgraph spanned by $U_1 \cup U_2$ is a stable matching in the entire graph, but this is not the case (see Exercise 52).

Let us state an immediate consequence of Theorem 18 and Lemma 17.

Corollary 19 *Let M and M' be stable matchings in a bipartite graph, with some assignment of preferences. Suppose $aB \in M$ and $aB \notin M'$. Then in M' both a and B have mates; also, one of a and B is better off in M' than in M, and the other is worse off.* □

The matching constructed by the fundamental algorithm in the proof of Theorem 16 is not only a stable matching, but it is also 'best' for the boys: every boy ends up with his highest preference among all stable matchings.

To make this definition more formal, let G be a bipartite graph with bipartition (V_1, V_2), with a certain assignment of preferences. A stable matching M is said to be V_1*-optimal* (or *optimal for the boys*) if for every stable matching M' and every vertex $a \in V_1$, if $aB \in M'$, then $aA \in M$ for some girl A, and either $A = B$ or else a prefers A to B. In other words, M is a V_1-optimal stable matching if in M every boy is at least as well off as in any other stable matching, once again with the assumption that it is better to be married than stay single. It is not clear that there is a V_1-optimal stable matching, but it is obvious that if there is one, then it is *unique*.

Theorem 20 *For every assignment of preferences in an n by n complete bipartite graph with bipartition (V_1, V_2), there is a V_1-optimal stable matching.*

Proof. Let us denote by $S(a)$ the set of girls a boy a could marry in some stable matching: this is the set of *possible* girls for a. We claim that in the fundamental algorithm no girl in $S(a)$ refuses a, so every boy marries his favourite possible girl, and thus the stable matching is optimal for the boys.

Suppose that this is not the case. Let us stop the algorithm when it happens for the very first time that a boy, say a, is refused by one of his possible girls, say A. By definition, this happens because A prefers another of her suitors at the time, say b. At that time b prefers A to all others that have not yet refused him. Hence, *a fortiori*, b prefers A to all others that are possible for him. As A is possible for a, there is a stable matching M in which a marries A and b marries a girl B. But this is impossible, since b prefers A to B, and A prefers b to a. This contradiction completes the proof. □

By definition, the V_1-optimal stable matching is 'best' for every boy (element of V_1). How 'good' is it for the girls? Recalling Corollary 19, we see that, somewhat surprisingly, it is the *worst* for every girl, independently of the assignment of preferences. To be a little more precise, call a stable matching M V_2*-pessimal* if in M no girl is better off than in any other stable matching. Once again, *a priori* it is not clear that there is a V_2-pessimal stable matching, but the definition implies that if there is a V_2-pessimal stable matching, then it is unique. Corollary 19 implies that the V_1-optimal stable matching is precisely the V_2-pessimal stable matching.

There are a good many extensions and variants of the results above; here we shall consider only stable *complete* matchings in (not necessarily complete) bipartite graphs with *equal* colour classes and stable matchings in a *polygynous society*.

Let us set the scene again, in a slightly different way. This rather heavy-handed definition of a stable matching will be frequently useful in applications. Suppose that we have a set V_1 of n boys and a set V_2 of n girls. Every boy and every girl has a possibly *incomplete* set of preferences. Thus for every girl A there is a list $L(A) = \{a_1, a_2, \ldots, a_k\}$, signifying that A is willing to marry *only* the boys $a_1, a_2, \ldots, a_k$, and this is exactly her order of preferences. Similarly, every boy a has a possibly incomplete list $L(a)$. We call (V_1, V_2, L) an *incomplete system of preferences*. This setup clearly corresponds to the bipartite graph $(V_1 \cup V_2, E)$, where $E = \{aA : a \in L(A) \text{ and } A \in L(a)\}$, with the preferences given by the lists. In this formulation a matching M from V_1 to V_2 is *stable* if any two matched vertices appear on each other's lists, and if $a \in L(B)$, $B \in L(a)$ but $aB \notin M$ then either $aA \in M$ for some $A \in L(a)$ that a prefers to B, or $bB \in M$ for some $b \in L(B)$ that B prefers to a.

How can we decide whether an incomplete system has a stable complete matching? Resembling our trick in the proof of Corollary 9, we can enlarge the incomplete system to a complete system in such a way that the stable matchings in the original incomplete system correspond to easily identifiable stable matchings in the enlarged complete system. To be precise, let us add a fictitious boy w and a fictitious girl W to the system: w is the *widower* and W is the *widow*. Set $V_1' = V_1 \cup \{w\}$ and $V_2' = V_2 \cup \{W\}$. Let us define a complete set of preferences for (V_1', V_2') as follows: each person slots in the widow (widower) after her (his) genuine preferences, and follows it with an arbitrary enumeration of the boys (girls) she (he) is unwilling to marry at all. Finally, the widow puts the widower at the end of her list, and the widower puts the widow at the end of his preferences. We say that this now complete system (V_1', V_2', L') is *associated* to the original system.

Theorem 21 *An incomplete system (V_1, V_2, L) with $|V_1| = |V_2|$ has a stable complete matching iff the associated complete system (V_1', V_2', L') has a stable matching in which the widow marries the widower.*

Proof. Let M be a complete matching from V_1 to V_2, and let M' be the complete matching from V_1' to V_2' obtained from M by adding to it the edge wW. To prove the theorem, we shall show that M is a stable matching in the incomplete system (V_1, V_2, L) iff M' is a stable matching in the associated complete system (V_1', V_2', L').

Suppose M is a stable matching in (V_1, V_2, L). Then M' is stable, since if $aA \in M$ then A prefers a to w and a prefers A to W.

Also, if M' is a stable matching in (V_1', V_2', L'), then M is a stable matching in (V_1, V_2, L). Indeed, if $aA \in M$, then $A \in L(a)$, since otherwise a prefers W to A, and W prefers a to w. Similarly, $a \in L(A)$, since otherwise A prefers w to a and w prefers A to W. Hence every edge aA of M satisfies $A \in L(a)$ and $a \in L(A)$, so M is a stable matching in (V_1, V_2, L). □

Since the V_1'-optimal stable matching in (V_1', V_2', L') is precisely the V_2'-pessimal stable matching, if some stable matching in (V_1', V_2', L') contains wW, then *every* stable matching contains wW. This gives us the following slightly stronger form of Theorem 21.

Theorem 21′. *Let (V_1, V_2, L) be an incomplete system, with associated complete system (V_1', V_2', L'), and let M' be a stable matching for (V_1', V_2', L'). Then there is a stable complete matching for (V_1, V_2, L) iff M' contains wW.* □

For the last variant of the stable matching theorem, it is convenient to use politically correct terms. In the *college admissions problem* we are given n applicants, $a_1, \ldots, a_n$, wishing to enter m colleges, $A_1, \ldots, A_m$, with college A_i willing to admit n_i undergraduates, such that $\sum_{i=1}^m n_i = n$. Each applicant orders the colleges according to his preferences and each college orders the applicants according to his or her preferences. Once again, an assignment of the applicants to the colleges, with n_i students assigned to college A_i, is said to be a *stable admissions scheme* if whenever a student a_i is admitted by a college $A_{i'}$ and a student a_j by a college $A_{j'}$, then either a_i prefers $A_{i'}$ to $A_{j'}$, or $A_{j'}$ prefers a_j to a_i. A stable admissions scheme is said to be *optimal* (for the applicants) if every applicant gets as good a college as possible under any stable admissions scheme.

Theorem 22 *No matter what the orders of preferences are, there is always an optimal stable admissions scheme.*

Proof. For the sake of argument, call the students *boys*, and replace each college A_i by n_i *girls*, say $A_i^{(1)}, A_i^{(2)}, \ldots, A_i^{(n_i)}$, with each $A_i^{(j)}$ having the same order of preferences among the boys $a_1, \ldots, a_n$ as A_i. Also, each boy orders the girls by taking the girls corresponding to the highest-rated college first (in any order) followed by the girls corresponding to the second college (in any order), and so on. In the bipartite graph we have just defined, take a stable matching that is optimal for $\{a_1, \ldots, a_n\}$: the admissions scheme this induces is clearly optimal (for the applicants). □

There are many ways of relaxing the conditions in the college admissions problem. The condition $\sum_{i=1}^k n_i = n$ need not by kept, an applicant may not wish to go to a certain college at all, and a college may not be willing to accept a student under any circumstances. For the sake of convenience, we can discard the last possibility by declaring that a student will not apply to a college if the college is unwilling to accept him under any circumstances. In this more general setup the analogue of the fundamental algorithm goes as follows. All students apply to their highest-rated colleges. A college of size n_i puts on its waiting list the n_i applicants it rates highest, or all of them if it gets no more than n_i applications, and rejects the others. The rejected students apply to their second choices, and again a college with quota n_i rejects all but its n_i highest applicants, and so on. The process stops when in a round no student gets rejected by a college, and then every college admits all the students on its waiting list. It is easy to show (see Exercise 51) that the admissions scheme obtained is such that (i) every college admits at most n_i of the students who applied to it, (ii) if a college does not admit its full quota then no student left on the shelf has applied to that college, and (iii) the assignment is *stable* in the sense that if a student a is admitted by a college A, and a student b by a college B, then *either* a is unwilling to go to B, *or* a prefers A to B *or* B prefers b to a. Furthermore, every student goes to as good a college as under any other admissions scheme satisfying these conditions.

III.6 Exercises

1. Suppose F is a set of edges after whose deletion there is no flow from s to t with positive value. Prove that F contains a cut separating s from t.

2. By summing an appropriate set of equations show that the capacity of a cut is at least as large as the value of a flow.

3. Let $\vec{G} = (V, \vec{E})$ be a directed graph and let c be an extended-real-valued capacity function on $\vec{E}$. (Thus $c(x, y)$ is a non-negative real or $+\infty$.) Let s and t be two vertices. Prove that either there is a flow from s to t with infinite value or else there is a flow with maximal finite value.

4. By successively reducing the number of circular flows in $\vec{G}$, prove that there is a maximal flow without circular flows in which no current enters the source and no current leaves the sink.

5. Use the method of Exercise 4 to show that if the capacity function is integral, then there is a maximal flow that is, also integral.

6. Formulate and prove the max-flow min-cut theorem of Ford and Fulkerson for multiple sources and sinks with bounds on the capacities of the edges and vertices.

7. (Circulation theorem.) A *circulation* in a directed graph $\vec{G}$ is a flow without a source and a sink. Given a lower capacity $l(x, y)$ and an upper capacity $c(x, y)$ for each edge $\vec{xy}$ with $0 \le l(x, y) \le c(x, y)$, we call a circulation g *feasible* if

$$l(x, y) \le g(x, y) \le c(x, y)$$

for every edge $\vec{xy}$. Prove that there exists a feasible circulation iff

$$l(S, \overline{S}) \le c(\overline{S}, S) \text{ for every } S \subset V.$$

[Note that the necessity of the condition is trivial, since in a feasible circulation the function l forces at least $l(S, \overline{S})$ current from S to $\overline{S}$ and the function c allows at most $c(\overline{S}, S)$ current from $\overline{S}$ back to S. To prove the sufficiency, adjoin a sink s and a source t to $\vec{G}$, send an edge from s to each vertex of $\vec{G}$ and send an edge from each vertex of $\vec{G}$ to t. Define a capacity function c^*

on the edges of the new graph $\vec{G^*}$ by putting $c^*(x, y) = c(x, y) - l(x, y)$, $c^*(s, x) = l(V, x)$, and $c^*(x, t) = l(x, V)$. Then the relation

$$f(x, y) = g(x, y) - l(x, y)$$

sets up a 1-to-1 correspondence between the feasible circulations g in $\vec{G}$ and flows f in $\vec{G^*}$ from s to t with value $l(V, V)$. Rewrite the condition given in the max-flow min-cut theorem in the form required in this result.]

8.[+] Let H be a bipartite multigraph without loops, with vertex classes V_1 and V_2. (Thus H may contain multiple edges; that is, two vertices belonging to different classes may be joined by several edges, which are said to be *parallel*.) As usual, given a vertex x, denote by $\Gamma(x)$ the set of edges incident with x and by $d(x) = |\Gamma(x)|$ the degree of x. Prove that, given any natural number k, the set E of edges can be partitioned into sets $E_1, E_2, \ldots, E_k$ such that for every vertex x and every set E_i we have

$$\left\lfloor \frac{d(x)}{k} \right\rfloor \leq |\Gamma(x) \cap E_i| \leq \left\lceil \frac{d(x)}{k} \right\rceil,$$

where, as in the rest of the book, $\lceil z \rceil$ is the least integer not less than z and $\lfloor z \rfloor = -\lceil -z \rceil$.

Thus if we think of the partition $\bigcup_1^k E_i$ as a colouring of the edges with k colours, then the colouring is *equitable* in the sense that in each vertex the distribution of colours is as equal as possible. [*Hint*. Construct a directed graph $\vec{H} = (V_1 \cup V_2, \vec{E})$ from H by sending an edge from x to y iff $x \in V_1$, $y \in V_2$ and H contains at least one xy edge. Let $\vec{G}$ be obtained from $\vec{H}$ by adding a vertex u and all the edges $\vec{ux}$, $\vec{yu}$, for $x \in V_1$ and $y \in V_2$, as shown in Fig. III.9. Define an appropriate upper and lower capacity for each edge of $\vec{G}$ and prove that there is a feasible integral circulation. Use this circulation to define one of the colour classes.]

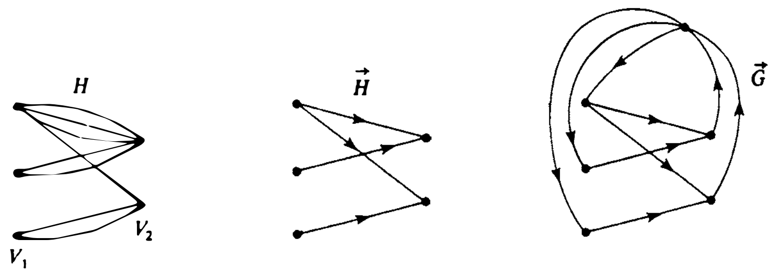


FIGURE III.9. The graphs H, $\vec{H}$, and $\vec{G}$.

9. (Exercise 8 contd.) Show that we may require that, in addition to the property above, the colour classes be as equal as possible, say $|E_1| \leq |E_2| \leq \cdots \leq |E_k| \leq |E_1| + 1$, and that in each set of parallel edges the distribution of colours is as equal as possible.

10. Let $d_1 \le d_2 \le \cdots \le d_n$ be the degree sequence of a graph G. Suppose that
$$d_j \ge j + k - 1 \text{ for } j = 1, 2, \ldots, n - 1 - d_{n-k+1}.$$
Prove that G is k-connected.

11. Let k and l be integers with $1 \le k \le l$. Construct graphs G_1, G_2, and G_3 such that
(i) $\kappa(G_1) = k$ and $\lambda(G_1) = l$,
(ii) $\kappa(G_2) = k$ and $\kappa(G_2 - x) = l$ for some vertex x,
(iii) $\lambda(G_3 - x) = k$ and $\lambda(G_3 - xy) = l$ for some edge xy.

12. Let G be a regular bipartite graph of degree at least 2. Show that $\kappa(G) \ne 1$.

13. Given $U \subset V(G)$ and a vertex $x \in V(G) - U$, an $x - U$ *fan* is a set of $|U|$ paths from x to U any two of which have exactly the vertex x in common. Prove that a graph G is k-connected iff $|G| \ge k + 1$ and for any k-set $U \subset V(G)$ and vertex x not in U, there is an $x - U$ fan in G. [*Hint.* Given a pair (x, U), add a vertex u to G and join it to each vertex in U. Check that the new graph is k-connected if G is. Apply Menger's theorem for x and u.]

14. Prove that if G is k-connected ($k \ge 2$), then every set of k vertices is contained in a cycle. Is the converse true?

15. The *line graph* $L(G)$ of a graph $G = (V, E)$ has vertex set E and two vertices $e, f \in E$ are adjacent iff they have exactly one vertex of G in common. By applying the vertex form of Menger's theorem to the line graph $L(G)$, prove that the vertex form of Menger's theorem implies the edge form.

16. Show that if $\lambda(G) = k \ge 2$, then the deletion of k edges from G results in a graph with at most 2 components. Is there a similar result for vertex-connectivity?

17. Let G be a connected graph with minimum degree $\delta(G) = k \ge 1$. Prove that G contains a path $x_1 x_2 \cdots x_k$ such that $G - \{x_1, x_2, \ldots, x_k\}$ is also connected. [*Hint.* Let $x_1 x_2 \cdots x_\ell$ be a longest path. Note that $\ell \ge k + 1$. Suppose that $G - \{x_1, x_2, \ldots, x_k\}$ is disconnected, and let $y_0 y_1 \cdots y_m$ be a longest path in a component C not containing x_{k+1}. Then $d_C(y_0) \le m$, but y_0 cannot be joined to $k - m$ of the vertices $x_1, \ldots, x_k$.]

18. Let $G = G_2(m, n)$ be a bipartite graph with vertex classes V_1 and V_2 containing a complete matching from V_1 to V_2.
(i) Prove that there is a vertex $x \in V_1$ such that for every edge xy there is a matching from V_1 to V_2 that contains xy.
(ii) Deduce that if $d(x) = d$ for every $x \in V_1$, then G contains at least $d!$ complete matchings if $d \le m$ and at least $d(d-1)\cdots(d-m+1)$ complete matchings if $d > m$.

19. Let $A = (a_{ij})_1^n$ be an $n \times n$ *doubly stochastic matrix*; that is, let $a_{ij} \ge 0$ and $\sum_{i=1}^n a_{ij} = \sum_{j=1}^n a_{ij} = 1$ for all i, j. Show that A is in the convex

hull of the $n \times n$ permutation matrices, i.e., there are $\lambda_i \geq 0$, $\sum_1^m \lambda_i = 1$, and permutation matrices $P_1, P_2, \ldots, P_m$ such that $A = \sum_1^m \lambda_i P_i$. [Let $a_{ij}^* = \lceil a_{ij} \rceil$, $A^* = (a_{ij}^*)_1^n$, and let $G = G_2(n, n)$ be the bipartite graph naturally associated with A^*. Show that G has a complete matching and deduce that there are a permutation matrix P and a real λ, $0 < \lambda \leq 1$, such that $A - \lambda P = B = (b_{ij})_1^n$ satisfies $b_{ij} \geq 0$, $\sum_{i=1}^n b_{ij} = \sum_{j=1}^n b_{ij} = 1 - \lambda$ for all i, j, *and* B has at least one more 0 entry than A.]

20. Prove the following form of the Schröder–Bernstein theorem. Let G be a bipartite graph with vertex classes X and Y having arbitrary cardinalities. Let $A \subset X$ and $B \subset Y$. Suppose there are complete matchings from A into Y and from B into X. Prove that G contains a set of independent edges covering all the vertices of $A \cup B$. [*Hint.* Consider the components of the union of the matchings.]

21. Let G be a bipartite graph with vertex sets V_1, V_2. Let A be the set of vertices of maximal degree. Show that there is a complete matching from $A \cap V_1$ into V_2.

22. Deduce from the previous exercise that every bipartite graph contains a set of independent edges such that each vertex of maximal degree (that is, degree $\Delta(G)$) is incident with one of the edges. Deduce that a non-empty regular bipartite graph has a 1-factor.

23. We say that G is an $(r, r-k)$*-regular* graph if $r - k \leq \delta(G) \leq \Delta(G) \leq r$. Prove that for $1 \leq k \leq s \leq r$ every $(r, r-k)$-regular graph contains an $(s, s-k)$-regular factor. [*Hint.* Assume $s = r - 1$. Take a minimal $(r, r-k)$-regular factor. Note that in this factor no two vertices of degree r are adjacent. Remove a set of independent edges covering the vertices of degree r.]

24. Let G be a graph with $\kappa(G) = k \geq 1$ and let $V_1 \cup W \cup V_2$ be a partition of $V(G)$ with $|W| = k$, $V_i \neq \emptyset$, $i = 1, 2$, and G containing no $V_1 - V_2$ edges. Show that, for each i, G contains either a matching from W into V_i or a matching from V_i into W.

25. Let G be a connected graph of order at least four such that every edge belongs to a 1-factor of G. Show that G is 2-connected. Show also that if $|G| \geq 2k$ and every set of $k-1$ independent edges in contained in a 1-factor, then G is k-connected.

26. Show that if a graph G has a 1-factor, $|G| \geq 2k + 2$, and every set of k independent edges is contained in a 1-factor, then every set of $k-1$ independent edges is contained in a 1-factor.

27. Let G be a connected graph of order at least 4, and let $F = \{f_1, \ldots, f_m\}$ be a 1-factor of G. Show that F contains two edges, $f_i = a_i b_i$ and $f_j = a_j b_j$, say, such that $G - \{a_i, b_i\}$ and $G - \{a_j, b_j\}$ are both connected.

28. An $r \times s$ *Latin rectangle* based on $1, 2, \ldots, n$ is an $r \times s$ matrix $A = (a_{ij})$ such that each entry is one of the integers $1, 2, \ldots, n$ and each integer occurs

in each row and column at most once. Prove that every $r \times n$ Latin rectangle A can be extended to an $n \times n$ *Latin square*. [*Hint*. Assume that $r < n$ and extend A to an $(r+1) \times n$ Latin rectangle. Let A_j be the set of possible values of $a_{r+1,j}$; that is, let $A_j = \{k : 1 \leq k \leq n, k \neq a_{ij}\}$. Check that the system $\{A_j : 1 \leq j \leq n\}$ has a set of distinct representatives.]
Prove that there are at least $n!(n-1)! \cdots (n-r+1)!$ distinct $r \times n$ Latin rectangles based on $1, 2, \ldots, n$. [*Hint*. Apply Exercise 18(ii).]

29. Let A be an $r \times s$ Latin rectangle and denote by $A(i)$ the number of times the symbol i occurs in A. Show that A can be extended to an $n \times n$ Latin square iff $A(i) \geq r + s - n$ for every $i = 1, 2, \ldots, n$.

30. The fundamental theorem of Tychonov's from general topology states that the product of a family of compact topological spaces is compact in the product topology. In combinatorics this result is frequently needed in the following simple form: the product of a family of finite sets is compact. Prove this in the following formulation.
Let Γ be an index set and $\Gamma^{(<\omega)}$ the collection of finite subsets of Γ. For each $\gamma \in \Gamma$, let S_γ be a finite set and, for $\Delta \subset \Gamma$, define $S_\Delta = \bigcup_{\gamma \in \Delta} S_\gamma$. For every $I \in \Gamma^{(<\omega)}$, let $\mathcal{F}_I$ be a non-empty set of functions $f : I \to S_I$, with $f(\gamma) \in S_\gamma$ for every $\gamma \in I$. Suppose that, for all $I' \subset I \in \Gamma^{(<\omega)}$ and $f \in \mathcal{F}_I$, the restriction $f|I'$ of f to I' belongs to $\mathcal{F}_{I'}$. Show that there is a function $F : \Gamma \to S_\Gamma$ such that $F|I \in \mathcal{F}_I$ for every $I \in \Gamma^{(<\omega)}$. [*Hint*. For $I \in \Gamma^{(<\omega)}$, define $\mathcal{G}_I = \{g \in \mathcal{F}_I$: for every $\mathcal{J} \in \Gamma^{(<\omega)}$, $\mathcal{J} \supset I$, there is an $f \in \mathcal{F}_{\mathcal{J}}$ with $f|I = g\}$. Note that $\mathcal{G}_I \neq \emptyset$ for every $I \in \Gamma^{(<\omega)}$. Find a function $F : \Delta \to S_\Delta$, with $\Delta \subset \Gamma$ maximal, such that $F|I \in \mathcal{G}_I$ for every $I \in \Delta^{(<\omega)}$.]

31.$^+$ Deduce from Tychonov's theorem in the previous exercise the following extension of Hall's theorem.
Let G be an infinite bipartite graph with vertex classes X and Y, such that each vertex in X is incident with finitely many edges. Then there is a complete matching from X into Y iff $|\Gamma(A)| \geq |A|$ for every finite subset A of X.
Show that the finiteness condition cannot be omitted.

32. Prove that a 2-edge-connected cubic graph has a 1-factor. [This result is called *Petersen's theorem.* In order to prove it, check that the condition of Tutte's theorem is satisfied. If $\emptyset \neq S \subset V(G)$ and C is an odd component of $G - S$, then there are at least two $S - C$ edges, since G is 2-edge connected. Furthermore, since G is cubic, there are at least three $S - C$ edges. Deduce that $q(G - S) \leq |S|$.]
Show also that a cubic graph need not have a 1-factor.

33. Imitate the second proof of Theorem 7 to give a direct proof of the case $d = 0$ of Theorem 12 and then deduce from it the general case $d \geq 0$.

34. Let G be a graph of order n with at most $r \geq 2$ independent vertices. Prove that if $\vec{G}$ is any orientation of G that does not contain a directed cycle (*acyclic*

orientation), then $\vec{G}$ contains a directed path of length at least $\lceil n/r \rceil - 1$. [*Hint.* Apply Dilworth's theorem, Theorem 13.]

35. Let $I_1, I_2, \ldots, I_{rs+1}$ be intervals in $\mathbb{R}$, with $r, s \geq 1$. Show that either some $r+1$ of these intervals have a non-empty intersection or some $s+1$ of them are pairwise disjoint.

36. Let $R_1, R_2, \ldots, R_m$ be rectangular parallelepipeds in canonical position in $\mathbb{R}^n$, so that $R_i = \prod_{i=1}^n [a_i, b_i]$. Show that if $m \geq rs^n + 1$ then either some $r+1$ of these parallelepipeds have a non-empty intersection, or some $s+1$ of them are pairwise disjoint.

37. Deduce from Exercise 34 the following result. Given a set of $rk+1$ distinct natural numbers, either there exists a set of $r+1$ numbers, none of which divides any of the other r numbers, or else there exists a sequence $a_0 < a_1 < \cdots < a_k$ such that if $0 \leq i < j \leq k$, then a_i divides a_j.

38. Describe all maximal graphs of order $n = 2l$ that do not contain a 1-factor. [*Hint.* Read it out of Tutte's theorem (Theorem 14).]

39. Make use of Exercise 38 and the convexity of the binomial coefficient $\binom{x}{2}$, $x \geq 2$, to prove that if $n \geq 2k+1$ then the maximal size of a graph of order n with at most k independent edges is

$$\max\left\{\binom{2k+1}{2}, \binom{k}{2} + k(n-k)\right\}.$$

Show also that the extremal graphs (that is, the graphs for which equality holds) are one or both of the graphs $K_{2k+1} \cup E_{n-2k-1}$ and $K_k + E_{n-k}$ (see Fig. III.10).

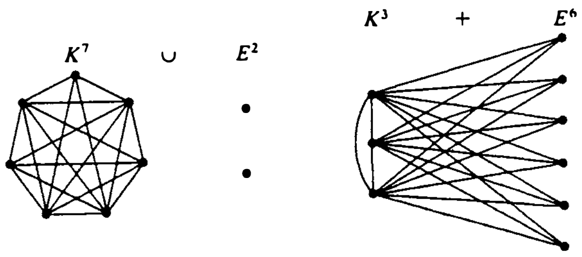


FIGURE III.10. For $k = 3$, $n = 9$ there are two extremal graphs: $K_7 \cup E_2$ and $K_3 + E_6$.

40. Call a sequence $d_1, d_2, \ldots, d_n$ of integers *graphic* if there is a graph G with vertex set $V(G) = \{x_1, x_2, \ldots, x_n\}$ such that $d(x_i) = d_i$, $1 \leq i \leq n$. (The graph G is said to *realize* $(d_i)_1^n$.) Show that $d_1 \geq d_2 \geq \cdots \geq d_n$ is graphic iff so is the sequence

$$d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, d_{d_1+3}, \ldots, d_n.$$

41. Use the algorithm given in Exercise 40 to decide which of the following sequences are graphic: 5, 4, 3, 2, 2, 2; 5, 4, 4, 2, 2, 1; 4, 4, 3, 3, 2, 2, 2; and 5, 5, 5, 4, 2, 1, 1, 1. Draw the graphs realizing the appropriate sequences constructed by the algorithm.

42. The general form of the fundamental algorithm for stable matching goes as follows. In each step of the algorithm *either* a boy with no outstanding proposal proposes to the next girl on his list *or* a girl with at least two outstanding proposals refuses all but her best suitor. These steps can be taken in any order. Show that the algorithm always constructs the same matching, namely the unique stable matching optimal for the boys.

43. Define a stable matching in a bipartite *multigraph* by defining, for each vertex, an order of preference on the set of edges incident with the vertex. Show that every bipartite multigraph has a stable matching.

44. Show that, for every $n \geq 1$, there is an assignment of preferences in an n by n bipartite graph such that there is precisely one stable matching (and so an optimal stable matching is also pessimal).

45. Show that in the stable matching optimal for the boys at most one boy ends up with his worst choice.

46. Suppose that in a set of n boys and n girls all boys have the same order of preferences. How many proposals are made in the fundamental stable matching algorithm?

47. What is the maximal number of proposals made in the fundamental stable matching algorithm when applied to n boys and n girls?

48. Let $m(n)$ be the maximal number of stable matchings in a set of n boys and n girls. Show that $m(n_1 + n_2) \geq m(n_1)m(n_2)$, and deduce that $m(n) \geq 2^{n/2}$ for $n \geq 2$. (What is $m(3)$?)

49. Show that if a stable matching contains an edge aA, with A being the worst for a and a being the worst for A, then every stable matching contains aA.

50. Let us say that a stable matching M is *less than* a stable matching M' (in notation, $M \prec M'$), if for all $aA \in M$ and $aA' \in M'$, either $A = A'$ or a prefers A' to A. Show that the set $\mathcal{M}$ of all stable matchings, endowed with the partial order $\prec$, is a distributive lattice. (All one needs is that if a, A, and A' are as above and A'' is the 'better' of A and A' for a, then the edges aA'' form a suitable matching, and so do the edges aA''', where A''' is the 'worse' of A and A' for a.) Deduce from this that if there are stable matchings, then there is a stable matching optimal for the boys.

51. Show that the admissions scheme produced by the analogue of the fundamental algorithm, described at the end of Section 5, has the claimed properties, and is optimal for the students.

52. Construct a bipartite graph G with preferences such that some stable matching of the subgraph spanned by $U_1 \cup U_2$ is not a stable matching in G, where U_1 and U_2 are as in Theorem 18.

53. Let $P = (X, <)$ be a partially ordered set containing no antichain on $m + 1$ elements. Show that P is the union of m chains. [*Hint.* Give the space $[m]^X$ of all functions $f : X \to [m] = \{1, 2, \dots, m\}$ the product topology. By Tychonov's theorem, this space is compact. For two incomparable elements, x and y, set

$$V_{xy} = \left\{ f \in [m]^X : f(x) \neq f(y) \right\}.$$

Note that each V_{xy} is closed and any finitely many of them have a non-empty intersection.]

54. Deduce from Dilworth's theorem the following result of Erdős and Szekeres. Every sequence $(x_i)_1^n$ of real numbers with more than $k\ell$ terms contains either an increasing subsequence with $k + 1$ terms or a decreasing subsequence with $\ell + 1$ terms. (A subsequence $(x_{i_j})_0^m$ is increasing if $x_{i_0} \leq x_{i_1} \leq \cdots \leq x_{i_m}$; it is decreasing if $x_{i_0} \geq x_{i_1} \geq \cdots \geq x_{i_m}$.) Show that a sequence of length $k\ell$ need not contain either.

55. Show that an incomplete regular graph on n vertices does not contain a complete graph on more than $n/2$ vertices.

56. Show that every connected regular bipartite graph with more than 2 vertices is 2-connected.

57. Let G be a bipartite graph with bipartition (U, W), $|U| = |W| = n$, and minimal degree at least $n/2$. Show that G has a complete matching.

58. Let G be an r-regular bipartite graph, and let E_0 be a set of $r - 1$ edges. Show that $G - E_0$ has a complete matching.

59. Let $\mathcal{M}$ be the set of complete matchings of an n by n bipartite graph. Let H be the graph with vertex set $\mathcal{M}$ in which $M_1 \in \mathcal{M}$ is joined to $M_2 \in \mathcal{M}$ if M_1 and M_2 agree in all but two edges. Is H necessarily connected?

60.$^-$ Let G be a connected graph that is not complete such that for any two distinct nonadjacent vertices there are k independent paths joining them. Show that $\kappa(G) \geq k$.

61. Let G be an r-regular graph of order $2r-2$ and vertex-connectivity $\kappa(G) = k$. Show that $k^2 - k + 12 \geq 4r$. Show also that equality holds if and only if $k \equiv 0$ or $1 \pmod 4$ and $V(G) = V_1 \cup V_2 \cup V_3$, $|V_1| = r + 1 - k$, $|V_2| = k$, $|V_3| = r - 3$, with $G[V_1]$ and $G[V_3]$ complete, each $x_1 \in V_1$ joined to all vertices in V_2, each $x_2 \in V_2$ joined to all vertices in V_1 and to $k - 1$ vertices in V_3, and each $x_3 \in V_3$ joined to four vertices in V_2. Note, in particular, that $k = \kappa(G) \geq 4$, and there is a unique r-regular graph of order $2r - 2$ and connectivity 4, namely a certain graph of order 10.

62. An n by n matrix $S = (s_{ij})$ is said to be *doubly substochastic* if $\sum_{i=1}^n s_{ij} \leq 1$ for every j and $\sum_{j=1}^n s_{ij} \leq 1$ for every i. Show that every doubly substochastic matrix is element-wise dominated by some doubly stochastic matrix; that is, if $S = (s_{ij})$ is doubly substochastic, then $s_{ij} \leq d_{ij}$ for some doubly stochastic matrix $D = (d_{ij})$.

63. (Exercise 19 contd.) Let $\lambda_n \geq 0$ be the maximal real number such that for every $n \times n$ doubly stochastic matrix $A = (a_{ij})_1^n$ there is a permutation matrix P for which all entries of $A - \lambda_n P$ are non-negative. Show that $\lambda_n = 1/\lfloor (n+1)^2/4 \rfloor$.

64. Let $n \geq 1$. Show that every $n \times n$ doubly stochastic matrix is the convex linear combination of $(n-1)^2 + 1$ permutation matrices but it need not be expressible as a convex linear combination of $(n-1)^2$ permutation matrices.

65. Let S_n^+ be the set of sequences $x = (x_1, \dots, x_n)$ such that $x_i \geq 0$ for all i and $\sum x_i = 1$. The *decreasing rearrangement* of a sequence $x \in S_n^+$ is the sequence $x^* = (x_{[i]})$ where $x_{[i]}$ is the i th largest term of x. Let us write $x \prec y$ to denote the fact that $\sum_1^k x_{[i]} \leq \sum_1^k y[i]$ for every k. Show that if $x, y \in S_n^+$, then $x \prec y$ if and only if $x = Dy$ for some doubly stochastic matrix D.

66. Show that for $y \in S_n^+$ the set $\{x \in S_n^+ : x \prec y\}$ is the convex hull of the points obtained by permuting the elements of y.

67. Let G be a 3-edge-connected cubic graph without a cutvertex. Show that G is also 3-connected. (Thus if $\lambda(G) = 3$ and $\kappa(G) \geq 2$ then $\kappa(G) = 3$.) Show also that there are cubic graphs G of arbitrarily large order with $\lambda(G) = 3$ and $\kappa(G) = 1$.

68. Let G be a bipartite graph with bipartition (V_1, V_2), and let $A_i \subset V_i, i = 1, 2$. Let I_i be a set of independent edges covering A_i, $i = 1, 2$. (Thus I_1 is a complete matching from A_1 into V_2, and I_2 is a complete matching from A_2 into V_1.) Show that $I_1 \cup I_2$ contains a set of independent edges covering $A_1 \cup A_2$.

69. Let G be a connected bipartite graph with $2k+3$ vertices in each class and each vertex having degree k or $k+1$. Show that G has a complete matching unless it is a certain graph, to be determined. [*Hint.* The exceptional graph consists of two not quite full copies of $K_{k+2,k+1}$, joined by an edge.]

70. Let $G = (V_1, V_2; E)$ be a bipartite graph such that $|\Gamma(A)| \geq |A| + d$ for every $A \subset V_1$, $A \neq \emptyset$. Show that G has a subgraph H such that $d_H(x) = d+1$ for every $x \in V_1$ and $|\Gamma_H(A)| \geq |A| + d$ for every $A \subset V_1$, $A \neq \emptyset$.

71. Let $1 \leq d_1 \leq d_2 \leq \cdots \leq d_n$. Show that $(d_i)_1^n$ is the degree sequence of some graph if and only if $\sum_1^n d_i$ is even and

$$\sum_{i=n-k+1}^{n} d_i \leq k(k-1) + \sum_{i=1}^{n-k} \min\{d_i, k\}.$$

72. Let G and H be graphs with vertex set V such that $d_G(x) = d_H(x) + 1$ for every $x \in V$. Show that there is a graph $\widetilde{G} = (V, \widetilde{E})$ such that $d_{\widetilde{G}}(x) = d_G(x)$ for every $x \in V$ and $\widetilde{G}$ contains a 1-factor (i.e., a set of $|V|/2$ independent edges).

73.$^-$ Let G be a graph with vertex set $V(G) = \{x_1, x_2, \ldots, x_n\}$ such that every vertex x_i is joined to at most $k-1$ of the vertices $x_1, x_2, \ldots, x_{i-1}$. Show that G is k-partite (i.e., $V(G) = \cup_{i=1}^k V_i$, with no edge joining vertices in the same class V_i).

74.$^-$ Show that for all $1 \le k \le \ell$ there is a graph G with vertex-connectivity $\kappa(G) = k$ and edge-connectivity $\lambda(G) = \ell$.

75.$^-$ Deduce from Hall's theorem the following theorem of König and Egerváry. Let A be an $m \times n$ matrix of 0s and 1s, and call a column or row of A a *line*. Then the minimal number of lines containing all the 1s of A is precisely the maximal number of 1s with no two in the same line.

76. Let x and y be vertices of G at distance d. Suppose that after the deletion of any $k-1$ of the vertices the distance between x and y is still d. Show that G contains k independent $x-y$ paths, each of length d.

77. Let x and y be adjacent vertices of degree at least k in a graph G. Show that if G/xy is k-connected then, so is G. [The graph G/xy is obtained from G by *contracting* the edge xy, i.e., by identifying x and y, and joining the new vertex to all other vertices of G that are joined in G to at least one of x and y.]

78. Let G be a graph of minimal degree 3 without two edge-disjoint cycles. Show that G is either K_4 or $K_{3,3}$ (i.e., it is either a complete graph on 4 vertices or a complete bipartite graph with 3 vertices in each class).

79. Determine all multigraphs (graphs with loops and multiple edges) of minimal degree 3 without two edge-disjoint cycles. [In a multigraph, a loop at x adds 2 to the degree of x; a loop forms a cycle of length 1, and two edges joining the same two vertices form a cycle of length 2.]

80. Deduce from the result of the previous exercise that every graph of order n and size $n+4$ contains two edge-disjoint cycles.

81. Show that a graph with n vertices and m edges has an independent set of at least $2n/3 - m/3$ vertices. For what graphs is equality attained? (What are the extremal graphs?)
The *transversal number* $\tau(G)$ of a graph G is the minimal number of vertices meeting every edge. Show that the transversal number of a graph with n vertices and m edges is at most $(n+m)/3$. What are the extremal graphs?

82. For $n \ge 2$ even, let F_n be the number of 1-factors of K_n. Show that $F_n = (n-1)!! = (n-1)(n-3)(n-5)\cdots = n!/\{(n/2)!2^{n/2}\}$.

83. Let $n \geq 6$ be even, and let $e_1, \ldots, e_{n-1}$ be edges of K_n. Show that $K_n - \{e_1, \ldots, e_{n-1}\}$ has a 1-factor unless each e_i is incident with the same vertex.

84. Let $\mathcal{A}$ be a Steiner triple system on $[n] = \{1, \ldots, n\}$, as defined in Exercise I.41, and let $U = \{x_1, \ldots, x_n\} \cup \{y_1, \ldots, y_n\} \cup \{z\}$ be a set of $2n+1$ elements. Show that

$$\mathcal{B} = \{x_i y_i z : 1 \leq i \leq n\} \cup \{x_i x_j x_k : ijk \in \mathcal{A}\} \cup \{x_i y_j y_k \in \mathcal{A}\}$$

is a Steiner triple system on U.

85. (Exercise 85 contd.) Show that $\mathcal{A}$ contains a set $\mathcal{A}_0$ of n triples such that each $i \in [n]$ is in precisely three triples that belong to $\mathcal{A}_0$. Let

$$U = X \cup Y \cup Z = \{x_1, \ldots, x_n\} \cup \{y_1, \ldots, y_n\} \cup \{z_0, \ldots, z_6\}$$

be a set of $2n+7$ elements, and let H be the bipartite graph with bipartition $X \cup Y$ and edge set $\{x_i y_j : ijk \in \mathcal{A}\}$. Show that $E(H)$ is the union of six 1-factors, $F_1, \ldots, F_6\}$, and set $F_0 = \{x_i y_i : 1 \leq i \leq n\}$. Finally, set

$$\begin{aligned}
\mathcal{C}_1 &= \{x_i x_j x_k : ijk \in \mathcal{A}\}, \\
\mathcal{C}_2 &= \{x_i y_j z_k : x_i y_j \in F_k, 0 \leq k \leq 6\}, \\
\mathcal{C}_3 &= \{x_i y_j y_k : ijk \in \mathcal{A} \setminus \mathcal{A}_0\}, \\
\mathcal{C}_4 &= \{y_i y_j y_k : ijk \in \mathcal{A}_0\}, \\
\mathcal{C} &= \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_5,
\end{aligned}$$

where $\mathcal{C}_5$ is a Steiner triple system on the 7-element set Z. Show that $\mathcal{C}$ is a Steiner triple system on U.

86. (Exercise 85 contd.) Deduce Kirkman's theorem that if n is of the form $6k+1$ or $6k+3$ then there is a Steiner triple system of order n.

III.7 Notes

The basic book on flows is L.R. Ford Jr. and D.R. Fulkerson, *Flows in Networks*, Princeton University Press, Princeton, 1962. It not only contains all the results mentioned in the chapter concerning flows and circulations, but also a number of applications to standard optimization problems.

The fundamental theorems of Menger, Hall, and Tutte are in K. Menger, Zur allgemeinen Kurventheorie, *Fund. Math.* **10** (1927) 96–115, P. Hall, On representatives of subsets, *J. London Math. Soc.* **10** (1935) 26–30, and W.T. Tutte, A factorization of linear graphs, *J. London Math. Soc.* **22** (1947) 107–111. The proof of Tutte's theorem we give is due to T. Gallai, Neuer Beweis eines Tutte'schen Satzes, *Magyar Tud. Akad. Közl.* **8** (1963) 135–139, and was rediscovered independently by I. Anderson, Perfect matchings in a graph, *J. Combinatorial Theory* Ser. B **10** (1971) 183–186 and W. Mader, Grad und lokaler Zusammenhang in endlichen Graphen, *Math. Ann.* **205** (1973) 9–11.

The stable matching theorem of Gale and Shapley and its variant concerning college admissions are from D. Gale and L.S. Shapley, College admissions and the stability of marriage, *American Mathematical Monthly* **69** (1962), 9–15.

The results in Exercises 8 and 9 are due to D. de Werra, Multigraphs with quasi-weak odd cycles, *J. Combinatorial Theory* Ser. B **23** (1977), 75–82.

A slightly simpler form of the result in Exercise 23 is due to W.T. Tutte; the proof indicated in the hint was found by C. Thomassen. An extensive survey of results concerning connectivity and matching can be found in Chapters I and II of B. Bollobás, *Extremal Graph Theory*, Academic Press, London and New York, 1978.

Stable matchings are discussed in D.E. Knuth, *Mariages Stables et Leur Relations avec d'Autres Problèmes Combinatoires*, Les Presses de l'Université de Montréal, Montréal, 1976, 106 pp., and in its slightly updated translation, *Stable Marriage and Its Relation to Other Combinatorial Problems – An Introduction to the Mathematical Analysis of Algorithms*, Amer. Math. Soc. 1997, xiii + 74 pp. For an account of the relationship among stable matchings, non-expansive networks and optimization, see Tomás Feder, *Stable Networks and Product Graphs*, Memoirs Amer. Math. Soc., Vol. **116**, 1995, x + 223 pp.

IV

Extremal Problems

Extremal problems are at the very heart of graph theory. Interpreting it broadly, extremal graph theory encompasses most of graph theory; in its narrow sense, it contains many of the deepest and most beautiful results of graph theory.

Of necessity, in this chapter we cannot take the broad view, so we shall concentrate on variants of the quintessential extremal problem, the *forbidden subgraph problem* : given a graph F, determine $\mathrm{ex}(n; F)$, the maximal number of edges in a graph of order n not containing F. Equivalently, how many edges guarantee that our graph contains F? For example, how many edges in a graph of order n force it to contain a path of length ℓ? A cycle of length at least ℓ? A cycle of length at most ℓ? A complete graph K_r?

More generally, an extremal question asks for the extreme values of certain graph parameters in various classes of graphs. For example, what is the maximal value of r for which there is a 2-connected r-regular graph of order n that is not Hamiltonian? Equivalently, how small a value of r guarantees that every 2-connected r-regular graph of order n is Hamiltonian? We shall not say much about these more general extremal questions, although occasionally we shall demand that our graphs be k-connected for some k or that their minimal degrees not be too small.

Before going into the details, it is appropriate to say a few words about terminology. If, for a given class of graphs, a certain graph parameter, say the number of edges or the minimal degree, is at most some number f, then the graphs for which equality holds are the *extremal graphs* of the inequality. As a trivial example, note that an acyclic graph of order n has at most $n - 1$ edges, and the extremal graphs are the trees of order n.

When we talk of extremal graphs, uniqueness is always understood *up to isomorphism.* Thus, a disconnected graph of order at least $n \geq 2d + 2$ and minimal degree at least $d \geq 0$ has at most $\binom{d+1}{2} + \binom{n-d-1}{2}$ edges, and $K_{d+1} \cup K_{n-d-1}$ is the unique extremal graph (see Exercise 2^-). Also, a graph of order n without odd cycles has at most $\lfloor n^2/4 \rfloor$ edges, and $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ is the only extremal graph.

In the forbidden subgraph problem a graph is *extremal* if it does not contain F and has $\mathrm{ex}(n; F)$ edges; the set of extremal graphs is $\mathrm{EX}(n; F)$. Thus we know from Mantel's theorem (Theorem I.2) that $\mathrm{ex}(n; K_3) = \lfloor n^2/4 \rfloor$ and, in fact, $\mathrm{EX}(n; K_3) = \{K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}\}$. Also, in the previous chapter, Tutte's factor theorem enabled us to solve a beautiful extremal problem: how many edges guarantee $k+1$ independent edges? In this case F consists of $k+1$ independent edges; that is, $F = (k+1)K_2$, and for $n \geq 2k+1$ the extremal graphs of $\mathrm{ex}(n; F)$ are $K_{2k+1} \cup \overline{K_{n-2k-1}}$ and/or $K_k + \overline{K_{n-k}}$ (see Exercises III.38–39).

The material in this chapter falls conveniently into two parts: the odd sections concern paths and cycles, while the even ones are about complete subgraphs. We have chosen to alternate the topics in order to have the simpler results first, as in most other chapters.

The first section is about paths and cycles (short and long) in graphs of large size. Among other results, we shall give a good bound on $\mathrm{ex}(n; P_\ell)$, the maximal number of edges in a graph of order n without a path of length ℓ. We shall also present some fundamental results about Hamilton cycles.

Extremal graph theory really started in 1941, when Turán, considerably extending Mantel's theorem, determined both the function $\mathrm{ex}(n; K_r)$ and the set $\mathrm{EX}(n; K_r)$. The second section is devoted to this fundamental theorem together with some related results.

When discussing $\mathrm{ex}(n; P_\ell)$ and $\mathrm{ex}(n; K_r)$, we mostly care about the case when n is large compared to ℓ and r. We get rather different problems if F and G have the same order. A prime example of these problems will be discussed in the third section, the problem of Hamilton cycles. Over the years considerable effort has gone into the solution of this problem, and in a certain rather narrow sense the present answers are satisfactory.

The fourth section is devoted to a deep and surprising theorem of Erdős and Stone, proved in 1946. The theorem, occasionally called the fundamental theorem of extremal graph theory, concerns $\mathrm{ex}(n; F)$, where F is a complete r-partite graph with t vertices in each class, but as an immediate corollary of this result one can determine $\lim_{n\to\infty} \mathrm{ex}(n; F)/n^2$ for every graph F.

The last two sections are about considerable new developments: Szemerédi's regularity lemma and its applications. In 1975, while proving his celebrated theorem on arithmetical progressions (see Section VI.4), Szemerédi discovered a beautiful result concerning the coarse structure of every graph. This theorem, *Szemerédi's regularity lemma*, is a vital tool in attacking numerous extremal problems. Once again, we do hardly more than point the way.

IV.1 Paths and Cycles

When looking for cycles in a graph, the most natural questions concern short cycles and long cycles. At most how large is the *girth*, the minimal length of a cycle? At least how large is the *circumference*, the maximal length of a cycle?

Let us see first what we can say about graphs with only a few more edges than vertices. A graph $G = G(n, n+1)$, that is, a graph with n vertices and $n+1$ edges, has girth $g(G) \le \lfloor 2(n+1)/3 \rfloor$. Indeed, G has at least two cycles, as its cyclomatic number is at least two. Now, if there are two edge-disjoint cycles, then $g(G) \le \frac{n+1}{2}$; otherwise, there are two vertices joined by three independent paths. Writing n_1, n_2 and n_3 for the lengths of these paths, we have $n_1 + n_2 + n_3 = n + 1$, and the three cycles formed by these three paths have lengths $n_1 + n_2$, $n_2 + n_3$, and $n_1 + n_3$. The sum of these three lengths is $2(n_1 + n_2 + n_3) \le 2(n+1)$, so G has at least one cycle of length at most $2(n+1)/3$. It is also easily seen that G need not contain a cycle of length less than $\lfloor 2(n+1)/3 \rfloor$.

Similarly, every graph $G(n, n+2)$ has girth at most $(n+2)/2$, and every $G(n, n+3)$ has girth at most $4(n+3)/9$ (see Exercises 12–14. Although this sequence can be continued for a few more values, the results become more and more complicated.

In looking for short cycles, it is more convenient to postulate that the minimal degree is large, rather than that the graph has many edges, so this is what we shall do now. In its natural formulation our first theorem gives a lower bound on the order of a graph in terms of the minimal degree and the *girth*, the length of a shortest cycle. Equivalently, the result gives an upper bound on the girth in terms of the order and minimal degree.

Theorem 1 *For $g \ge 3$ and $\delta \ge 3$ put*

$$n_0(g, \delta) = \begin{cases} 1 + \dfrac{\delta}{\delta - 2}\{(\delta - 1)^{(g-1)/2} - 1\} & \text{if } g \text{ is odd,} \\[2ex] \dfrac{2}{\delta - 2}\{(\delta - 1)^{g/2} - 1\} & \text{if } g \text{ is even.} \end{cases}$$

Then a graph G with minimal degree δ and girth g has at least $n_0(g, \delta)$ vertices.

Proof. Suppose first that g is odd, say $g = 2d+1$, $d \ge 1$. Pick a vertex x. There is no vertex z for which g contains two distinct z–x paths of length at most d, since otherwise G has a cycle of length at most $2d$. Consequently, there are at least δ vertices at distance 1 from x, at least $\delta(\delta - 1)$ vertices at distance 2, and so on, and at least $\delta(\delta - 1)^{d-1}$ vertices at distance d from x (Fig. IV.1).

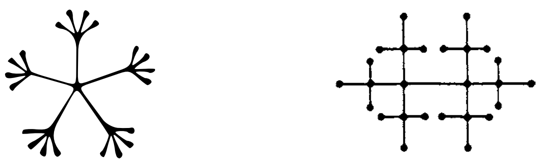


FIGURE IV.1. The cases $\delta = g = 5$ and $\delta = 4$, $g = 6$.

Thus

$$n \ge 1 + \delta + \delta(\delta - 1) + \cdots + \delta(\delta - 1)^{d-1},$$

as claimed.

Suppose now that g is even, say $g = 2d$. Pick two adjacent vertices, say x and y. Then there are $2(\delta - 1)$ vertices at distance 1 from $\{x, y\}$, $2(\delta - 1)^2$ vertices at distance 2, and so on, and $2(\delta - 1)^{d-1}$ vertices at distance $d - 1$ from $\{x, y\}$, implying the required inequality. □

Let G_0 be an *extremal graph* of Theorem 1, that is, a graph with parameters δ and g, for which equality holds. The proof above implies that G_0 is *regular of degree* δ; if $g = 2d + 1$, then G_0 has *diameter* d, and if $g = 2d$, then every vertex is within distance $d - 1$ of each pair of adjacent vertices. It is easily seen that $n_0(g, \delta)$ is also the *maximal* number for which there is a graph with *maximal degree* δ having the latter property (Exercise 4). We call G_0 a *Moore graph of degree* δ *and girth* g or, if $g = 2d + 1$, a *Moore graph of degree* δ *and diameter* d. In Chapter VIII we shall use algebraic methods to investigate Moore graphs. Here let us note only that the Heawood graph, the incidence graph of the Fano plane, shown in Fig. I.7, is a bipartite cubic graph of order 14 and girth 6, so it is a Moore graph of degree 3 and girth 6. Similarly, the Petersen graph, shown in Fig. V.11, is a Moore graph of degree 3 and diameter 2 (or girth 5).

Let us see now what we can say about *long* cycles and paths in a graph. Our first result in this direction is a theorem of Pósa, extending a fundamental theorem of Dirac from 1952. If a graph of order n is Hamiltonian, then its *circumference* is n, while the length of a longest path is $n - 1$. However, every non-Hamiltonian *connected* graph contains at least as long paths as the circumference of the graph. Indeed, if $C = x_1 x_2 \cdots x_\ell$ is a longest cycle and $\ell < n$ then there is a vertex y not on C that is adjacent to a vertex of C, say x_1. But then $y x_1 x_2 \cdots x_\ell$ is a path of length ℓ.

Theorem 2 *Let G be a connected graph of order $n \geq 3$ such that for any two non-adjacent vertices x and y we have*

$$d(x) + d(y) \geq k.$$

If $k = n$ then G is Hamiltonian, and if $k < n$ then G contains a path of length k and a cycle of length at least $(k + 2)/2$.

Proof. Assume that G is not Hamiltonian and let $P = x_1 x_2 \cdots x_\ell$ be a longest path in G. The maximality of P implies that the neighbours of x_1 and x_ℓ are vertices of P. As G does not contain a cycle of length ℓ, x_1 is not adjacent to x_ℓ. Even more, the path P cannot contain vertices x_i and x_{i+1} such that x_1 is adjacent to x_{i+1} and x_ℓ is adjacent to x_i, since otherwise $x_1 x_2 \cdots x_i x_\ell x_{\ell-1} \cdots x_{i+1}$ is a cycle of length ℓ (Fig. IV.2).

Consequently, the sets

$$\Gamma(x_1) = \{x_j : x_1 x_j \in E(G)\} \text{ and } \Gamma^+(x_\ell) = \{x_{i+1} : x_i x_\ell \in E(G)\}$$

are disjoint subsets of $\{x_2, x_3, \ldots, x_\ell\}$, and so

$$k \leq d(x_1) + d(x_\ell) = |\Gamma(x_1)| + |\Gamma^+(x_\ell| \leq \ell - 1 \leq n - 1.$$

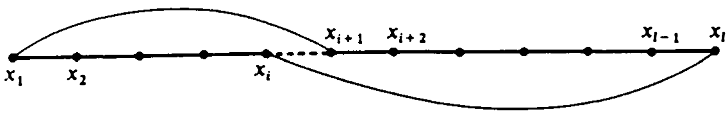


FIGURE IV.2. The construction of a cycle of length ℓ.

Now, if $k = n$ then this is a contradiction, so G is Hamiltonian. Also, if $k < n$, then this relation implies that G has a path of length $\ell - 1 \geq k$. This proves the first two assertions of the theorem.

Finally, the assertion about cycles is even simpler. Assume that $d(x_1) \geq d(x_\ell)$, so $d(x_1) \geq \lceil k/2 \rceil$. Put $t = \max\{i : x_1 x_i \in E(G)\}$. Then $t \geq d(x_1) + 1 \geq \lceil k/2 \rceil + 1$, and G contains the cycle $x_1 x_2 \cdots x_t$ of length t. $\square$

Theorem 2 contains Dirac's theorem: every graph of order $n \geq 3$ and minimal degree at least $n/2$ is Hamiltonian.

In Section 3 we shall make use of the proof of Theorem 2 to obtain detailed information about graphs without long cycles and paths. For the moment we confine ourselves to noting two of its consequences.

Theorem 3 *Let G be a graph of order n without a path of length $k (\geq 1)$. Then*

$$e(G) \leq \frac{k-1}{2} n.$$

A graph is an extremal graph (that is, equality holds for it) iff all its components are complete graphs of order k.

Proof. We fix k and apply induction on n. The assertion is clearly true if $n \leq k$. Assume now that $n > k$ and the assertion holds for smaller values of n.

If G is disconnected, then the induction hypothesis implies the result. Now, if G is connected, then it contains no K_k and, by Theorem 2, it has a vertex x of degree at most $(k-1)/2$. Since $G - x$ is not an extremal graph,

$$e(G) \leq d(x) + e(G - x) < \frac{k-1}{2} + \frac{k-1}{2}(n-1) = \frac{k-1}{2} n. \qquad \square$$

Theorem 4 *Let $k \geq 2$ and let G be a graph of order n in which every cycle has length at most k. Then*

$$e(G) \leq \frac{k}{2}(n-1).$$

A graph is extremal iff it is connected and all its blocks are complete graphs of order k. $\square$

The proof of this result is somewhat more involved than that of Theorem 3. Since a convenient way of presenting it uses "simple transforms" to be introduced in Section 3, the proof is left as an exercise (Exercise 37), with a detailed hint.

IV.2 Complete Subgraphs

What is $\mathrm{ex}(n; K_{r+1})$, the maximal number of edges in a graph of order n not containing a K_{r+1}, a complete graph of order $r+1$? If G is r-partite, then it does not contain a K_{r+1}, since every vertex class of G contains at most one vertex of a complete subgraph. Thus $\mathrm{ex}(n; K_{r+1})$ is at least as large as the maximal size of an r-partite graph of order n. In fact, there is a unique r-partite graph of order n that has maximal size. This graph is the *Turán graph* $T_r(n)$, the complete r-partite graph with n vertices and as equal classes as possible (see Fig. IV.3), so that if we order the classes by size and there are n_k vertices in the kth class, then $n_1 \le n_2 \le \cdots \le n_r \le n_1 + 1$. To see that this is the case, let G be an r-partite graph of order n and maximal size. Clearly, G is a complete r-partite graph. Suppose the classes are not as equal as possible; say there are m_1 vertices in the one class and $m_2 \ge m_1 + 2$ in another. Then, by transferring one vertex from the second class to the first, we would increase the number of edges by $(m_1+1)(m_2-1) - m_1m_2 = m_2 - m_1 - 1 \ge 1$. Note that the relations $n_1 \le n_2 \le \cdots \le n_r$ and $\sum_{i=1}^r n_i = n$ uniquely determine the n_i, and so $T_r(n)$ is unique. In fact, $n_i = \lfloor (n+i-1)/r \rfloor$ for $i = 1, \ldots, r$.

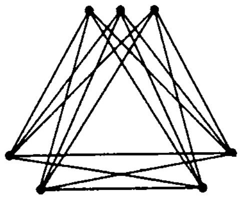


FIGURE IV.3. The Turán graph $T_3(7)$.

The number of edges in the Turán graph $T_r(n)$ is usually denoted by $t_r(n)$; thus, for example, $t_2(n) = \lfloor n^2/4 \rfloor$. Simple calculations show that

$$t_r(n) \ge \left(1 - \frac{1}{r}\right)\binom{n}{2}. \tag{1}$$

In fact, if $r \ge 1$ is fixed and $n \to \infty$, then

$$t_r(n) = \left(1 - \frac{1}{r} + o(1)\right)\binom{n}{2}.$$

Here and elsewhere, we use Landau's notation: $g = O(f)$ if g/f is bounded as $n \to \infty$, and $g = o(f)$ if $g/f \to 0$ as $n \to \infty$. In particular, $o(1)$ denotes a function tending to 0 as $n \to \infty$.

A fundamental theorem of Turán states that the trivial inequality $\mathrm{ex}(n; K_{r+1}) \ge t_r(n)$ is, in fact, an equality for every n and r. In proving this, somewhat as in the case of Hall's theorem, we have an embarrassment of riches: there are many beautiful ways of proving the theorem, since the Turán graph $T_r(n)$ is ideal for all kinds of induction arguments. Before getting down to some proofs, we observe some simple properties of $T_r(n)$ and, in general, of graphs of order n and size $t_r(n)$; indeed, after these observations, several proofs of Turán's theorem will be almost immediate.

Clearly, $\delta(T_r(n)) = n - \lceil n/r \rceil$ and $\Delta(T_r(n)) = n - \lfloor n/r \rfloor$, so the minimal degree of a Turán graph is at most one smaller than its maximal degree. In other words, given that $T_r(n)$ has n vertices and $t_r(n)$ edges, its degrees are as equal as possible: if $G = G(n, t_r(n))$, then $\delta(G) \leq \delta(T_r(n))$ and $\Delta(G) \geq \Delta(T_r(n))$. Also, if $x \in T_r(n)$ is a vertex of minimal degree, then $T_r(n) - x$ is precisely $T_r(n-1)$. If H is an $(r-1)$-partite graph of order $n-k$ and $H + \overline{K_k} \cong T_r(n)$, then, k is $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$ and $H \cong T_{r-1}(n-k)$. As a slight variant of this, we see that if $H = G(n-k, t_{r-1}(n-k))$ and $e(H + \overline{K_k}) = t_r(n)$, then k is $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$. Equivalently, $t_r(n) - k(n-k) > t_{r-1}(n-k)$ unless k is $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$.

We dignify the final observation by calling it a theorem.

Theorem 5 *Let G be a graph with n vertices and at least $t_r(n)$ edges, and let x be a vertex of maximal degree, say, $d(x) = n - k = \Delta(G)$. Set $W = \Gamma(x)$, $U = V(G) \setminus W$ and $H = G[W]$. Then $e(H) \geq t_{r-1}(n-k)$, and the inequality is strict unless $k = \lfloor n/r \rfloor$ and U is an independent set of vertices, each of degree $n-k$.*

Proof. As we noted above, $k \leq \lfloor n/r \rfloor$. Assume that $e(H) \leq t_{r-1}(n-k)$. Then

$$\begin{aligned} t_r(n) \leq e(G) &= e(H) + \frac{1}{2}\sum_{u \in U} d(u) + \frac{1}{2}e(U, W) \\ &\leq e(H) + k(n-k) \leq t_{r-1}(n-k) + k(n-k). \end{aligned}$$

Consequently, $k = \lfloor n/s \rfloor$, $e(U, W) = k(n-k)$, and so $G = H + \overline{K_k}$, as claimed. □

From here it is but a short step to connect $T_r(n)$ with complete subgraphs and thereby deduce the following extension of Turán's theorem.

Theorem 6 *Let G be a graph with n vertices and at least $t_r(n)$ edges. Consider the following simple algorithm for finding a complete subgraph of order $r+1$. Pick a vertex x_1 of maximal degree in $G_1 = G$, then a vertex x_2 of maximal degree in the subgraph G_2 of G_1 spanned by the neighbours of x_1, then a vertex x_3 of maximal degree in the subgraph G_3 of G_2 spanned by the neighbours of x_2 (in G_2), and so on, stopping with x_ℓ if it has no neighbours in G_ℓ. Then either G is a Turán graph $T_r(n)$, or else the procedure above constructs at least $r+1$ vertices, $x_1, x_2, \ldots, x_{r+1}$, which then span a complete subgraph.*

In particular, $\mathrm{ex}(n; K_{r+1}) = t_r(n)$, and $T_r(n)$ is the unique extremal graph.

Proof. We apply induction on r, noting that for $r = 1$ there is nothing to prove. Set $n - k = d(x_1) = \Delta(G)$. If $e(G_2) > t_{r-1}(n-k)$, then we are done by the induction hypothesis, since G_2 cannot be isomorphic to $T_{r-1}(n-k)$, and x_1, followed by the vertices $x_2, x_3, \ldots, x_{r+1}$ we find in G_2, gives the sequence as claimed. Otherwise, by Theorem 5, $k = \lfloor n/r \rfloor$, $e(G_2) = t_{r-1}(n-k)$, and $G = G_2 + \overline{K_k}$. Hence, by another application of the induction hypothesis to G_2, we see that either our procedure constructs $x_2, \ldots, x_{r+1}$, or else $G_2 \cong T_{r-2}(n-k)$ and $G_1 \cong T_{r-1}(n)$, as claimed. $\square$

In 1970 Erdős proved a beautiful result about *all* graphs containing no K_{r+1}, regardless of their number of edges, namely that the degree sequence of a graph without a K_{r+1} is dominated by the degree sequence of an r-partite graph. This result again implies Turán's theorem.

Theorem 7 *Let G be a graph with vertex set V that does not contain K_{r+1}, a complete graph of order r. Then there is an r-partite graph H with vertex set V such that for every vertex $z \in V$ we have*

$$d_G(z) \leq d_H(z).$$

If G is not a complete r-partite graph, then there is at least one vertex z for which the inequality above is strict.

Proof. We shall apply induction on r. For $r = 1$ there is nothing to prove, since G is the empty graph $\overline{K_n}$, which is 1-partite. Assume now that $r \geq 2$ and the assertion holds for smaller values of r.

Pick a vertex $x \in V$ for which $d_G(x)$ is maximal and denote by W the set of vertices of G that are joined to x. Then $G_0 = G[W]$ does not contain a K_r for otherwise with x it would form a K_{r+1}. By the induction hypothesis we can replace G_0 by an $(r-1)$-partite graph H_0 with vertex set W in such a way that $d_{G_0}(y) \leq d_{H_0}(y)$ for every $y \in W$ and strict inequality holds for at least one y unless G_0 is a complete $(r-1)$-partite graph. Add to H_0 the vertices in $V - W$ and join each vertex in $V - W$ to each vertex in W. To complete the proof let us check that the r-partite graph H obtained in this way has the required properties.

If $z \in U = V - W$, then $d_H(z) = d_H(x) = d_G(x) \geq d_G(z)$, and if $z \in W$, then $d_H(z) = d_{H_0}(z) + n - |W| \geq d_{G_0}(z) + n - |W| \geq d_G(z)$. Thus $d_G(z) \leq d_H(z)$ holds for every $z \in V$.

What can we say about G if $e(H) = e(G)$? Then $e(H_0) = e(G_0)$, so G_0 is a complete $(r-1)$-partite graph. Also, by counting the edges outside $G_0 = H_0$, we see that

$$\begin{aligned} 0 = e(G) - e(G_0) &= \sum_{u \in U} d_G(u) - e(G[U]) - |U||W| \\ &\leq |U||W| - e(G[U]) - |U||W| = -e(G[U]), \end{aligned}$$

implying that G is a complete r-partite graph. $\square$

In order to emphasize the importance of Turán's theorem, we state it once more, this time in its original form, as it was stated in 1940.

Theorem 8 *For $r, n \geq 2$ we have* $\mathrm{ex}(n; K_{r+1}) = t_r(n)$ *and* $\mathrm{EX}(n; K_{r+1}) = \{T_r(n)\}$. *In words, every graph of order n with more than $t_r(n)$ edges contains a K_{r+1}. Also, $T_r(n)$ is the unique graph of order n and size $t_r(n)$ that does not contain a K_{r+1}.*

Proof. The theorem is contained in Theorem 6, and it is also an immediate consequence of Theorem 7, since $T_r(n)$ is the unique r-partite graph of order n and maximal size.

Nevertheless, let us give two more proofs of the theorem itself, based again on the properties of $T_r(n)$.

3 *rd Proof.* For $r = 1$ there is nothing to prove, so fix $r \geq 2$ and apply induction on n. For $n \leq r + 1$ the assertion is trivial, so suppose that $n > r + 1$ and the theorem holds for smaller values of n.

Suppose G has n vertices, $t_r(n)$ edges, and it contains no K_{r+1}. As $T_r(n)$ is a maximal graph without a K_{r+1} (that is, no edge can be added to it without creating a K_{r+1}), the induction step will follow if we show that G is exactly a $T_r(n)$. Since the degrees in $T_r(n)$ differ by at most 1, we have

$$\delta(G) \leq \delta(T_r(n)) \leq \Delta(T_r(n)) \leq \Delta(G).$$

Let x be a vertex of G with degree $d(x) = \delta(G) \leq \delta(T_r(n))$. Then

$$e(G - x) = e(G) - d(x) \geq e(T_r(n - 1)),$$

so by the induction hypothesis $G_x = G - x$ is exactly a $T_r(n - 1)$.

A smallest vertex class of G_x contains $\lfloor (n - 1)/r \rfloor$ vertices, and the vertex x is joined to all but

$$n - 1 - \left(n - \left\lceil \frac{n}{r} \right\rceil\right) = \left\lfloor \frac{n-1}{r} \right\rfloor$$

vertices of G_x. Since x cannot be joined to a vertex in each class of G_x, it has to be joined to all vertices of G_x save the vertices in a smallest vertex class. This shows that $G = T_r(n)$, as required. □

4 *th Proof.* This time we apply induction on $n + r$. Assume that $2 \leq r < n$ and the assertion holds for smaller values of $r + n$. Fix a graph $G = G(n, t_r(n))$ without a K_{r+1}: as before, it suffices to prove that we must have $G \cong T_r(n)$. Since $t_r(n) > t_{r-1}(n)$, by the induction hypothesis G contains a K_r, say, with vertex set $W = \{x_1, x_2, \ldots, x_r\}$. Set $U = V(G) \setminus W$ and $H = G[U]$. Clearly, no vertex $x \in U$ sends r edges to W, so

$$\begin{aligned} e(H) &= e(G) - \binom{r}{2} - e(U, W) \\ &\geq t_{r-1}(n) - \binom{r}{2} - (n - r)(r - 1) = t_r(n - r). \end{aligned}$$

The second equality above follows from the fact that if we remove (the vertex set of) a K_r from $T_r(n)$ then we remove precisely $\binom{r}{2} + (n - r)(r - 1)$ edges, and we are left with a $T_r(n - r)$. Now, as H contains no K_{r+1}, by the induction hypothesis the inequality above implies that $H \cong T_r(n - r)$, and every vertex of H is joined to precisely $r - 1$ vertices of W. It is easily checked that this forces $G \cong T_r(n)$, as in the previous proof. □

The proofs above can easily be adapted to give a number of related results (cf. Exercises 18-23). Yet another proof of Turán's theorem will be given in Chapter VIII.

In a slightly different formulation, Turán's theorem gives a lower bound on the clique number of a graph with given order and size. A maximal complete subgraph of a graph is a *clique*, and the *clique number* $\omega(G)$ of a graph G is the maximal order of a clique in G. Simply, $\omega(G)$ is the maximal order of a complete subgraph of G. Now, Turán's theorem states that if a graph G has n vertices and $m \geq t_{r-1}(n)$ edges then $\omega(G) \geq r$, unless $m = t_{r-1}(n)$ and $G \equiv T_{r-1}(n)$.

Now let us turn to the *problem of Zarankiewicz*, which is the analogue of Turán's problem in bipartite graphs. Write $G_2(m, n)$ for a bipartite graph with m vertices in the first class and n in the second. What is the maximal size of a $G_2(m, n)$ if it does not contain a complete bipartite graph with s vertices in the first class and t in the second? This maximum is usually denoted by $z(m, n;\, s, t)$. The following simple lemma seems to imply a very good upper bound for the function $z(m, n;\, s, t)$.

Lemma 9 *Let m, n, s, t, k, r be non-negative integers, $2 \leq s \leq m$, $2 \leq t \leq n$, $0 \leq r < m$, and let $G = G_2(m, n)$ be an m by n bipartite graph of size $z = my = km + r$ without a $K_{s,t}$ subgraph having s vertices in the first class and t in the second. Then*

$$m\binom{y}{t} \leq (m-r)\binom{k}{t} + r\binom{k+1}{t} \leq (s-1)\binom{n}{t}. \tag{2}$$

Proof. Denote by V_1 and V_2 the vertex classes of G. We shall say that a t-set (i.e., a set with t elements) T of V_2 is *covered* by a vertex $x \in V_1$ if x is joined to every vertex in T. The number of t-sets covered by a vertex $x \in V_1$ is $\binom{d(x)}{t}$. Since the assumption on G is exactly that each t-set in V_2 is covered by at most $s-1$ vertices of V_1, we find that

$$\sum_{x \in V_1} \binom{d(x)}{t} \leq (s-1)\binom{n}{t}. \tag{3}$$

As $\sum_{x \in V_1} d(x) = z = my = km + r$, $0 \leq r < m$, and $f(u) = \binom{u}{t}$ is a convex function of u for $u \geq t$, inequality (3) implies (2). □

The proof of Lemma 9 is the simple but powerful double counting argument; as this is perhaps the most basic combinatorial argument, let us spell it out again, this time in terms of the edges of a bipartite graph H. One of the vertex classes of H is just V_1, but the other is $V_2^{(t)}$, the set of all t-subsets of V_2. In our new graph H, join $x \in V_1$ to $A \in V_2^{(t)}$ if in G the vertex x is joined to all t vertices of A. Now, counting from V_1, we see that

$$e(H) = \sum_{x \in V_1} \binom{d(x)}{t}.$$

On the other hand, as G contains no $K_{s,t}$, in H every vertex $A \in V_2^{(t)}$ has at most $s-1$ neighbours. Thus

$$e(H) \le (s-1)\binom{n}{t},$$

and the rest is simple algebra.

Theorem 10 *For all natural numbers m, n, s and t we have*

$$z(m,n;\, s,t) \le (s-1)^{1/t}(n-t+1)m^{1-1/t} + (t-1)m.$$

Proof. Let $G = G_2(m,n)$ be an extremal graph for the function $z(m,n;s,t) = my$ without a $K(s,t)$ subgraph. As $y \ge n$, inequality (2) implies

$$(y-(t-1))^t \le (s-1)(n-(t-1))^t m^{-1}. \qquad \square$$

The only advantage of Theorem 10 is that it is fairly transparent: for any particular choice of the parameters we are better off dealing with inequality (2). Thus, for example,

$$z(n,n;\, s,2) \le \frac{n}{2}(1 + (4(s-1)(n-1)+1)^{1/2}). \tag{4}$$

Indeed, with the notation of Lemma 9 we have $n\binom{y}{2} \le (s-1)\binom{n}{2}$. Hence

$$y^2 - y - (s-1)(n-1) \le 0,$$

implying that ny is at most the right-hand side of (4).

The method of proof of Lemma 9 also gives an upper bound for $\mathrm{ex}(n;\, K_2(t))$, the maximal number of edges in a graph of order n without a complete t by t bipartite subgraph.

Theorem 11 *Let n, s, t, k and r be non-negative integers, and let G be a graph of order* $z = ny/2 = \frac{1}{2}(kn+r)$, *containing no* $K_{s,t}$. *Then*

$$n\binom{y}{t} \le (n-r)\binom{k}{t} + r\binom{k+1}{t} \le (s-1)\binom{n}{t}.$$

Furthermore,

$$\mathrm{ex}(n, K_{s,t}) \le \frac{1}{2}(s-1)^{1/t}n^{2-1/t} + \frac{1}{2}(t-1)n.$$

Proof. As in Lemma 9, let us say that a t-set of the vertices is *covered* by a vertex x if x is joined to every vertex of the t-set. Since G does not contain a $K_{s,t}$, every t-set is covered by at most $s-1$ vertices. Therefore, if G has degree sequence $(d_i)_1^n$ then

$$\sum_1^n \binom{d_i}{t} \le (s-1)\binom{n}{t},$$

and the rest is as in Lemma 9 and Theorem 10. $\square$

In fact, there is no need to repeat the proof of Lemma 9 to prove Theorem 11: a simple, and general, "duplication" argument will do the job. Given a graph G with vertex set $V(G) = \{x_1, \dots, x_n\}$, construct a bipartite graph $H = D(G)$ as follows. Take two disjoint copies of $V(G)$, say $V_1 = \{x_1', \dots, x_n'\}$ and $V_2 = \{x_1'', \dots, x_n''\}$. The graph H has bipartition (V_1, V_2), and $x_i'x_j'' \in E(H)$ iff $x_ix_j \in E(G)$. Clearly, $e(H) = 2e(G)$; in fact, $d_G(x_i) = d_H(x_i') = d_H(x_i'')$ for every i. It is easily seen that if $K_{s,t} \not\subset G$, then $K_{s,t} \not\subset H$, and so $z(n, n; s, t) \geq 2\mathrm{ex}(n, K_{s,t})$.

What about the order of $z(n, n; t, t)$? We see from Theorem 10 that if $t \geq 2$ is fixed, then

$$z(n, n; t, t) \leq (t-1)^{1/t} n^{2-1/t} + (t-1)n. \tag{5}$$

Also, by Theorem 11,

$$\mathrm{ex}(n, K_{t,t}) \leq \frac{1}{2}(t-1)^{1/t} n^{2-1/t} + \frac{1}{2}(t-1)n. \tag{6}$$

It is very likely that (5) and (6) give the correct orders of the functions $z(n, n; t, t)$ and $\mathrm{ex}(n, K_{t,t})$, but this has been proved only for $t = 2$ and 3. In fact, it is rather hard to find nontrivial lower bounds for $z(m, n; s, t)$. In Chapter VII we shall use the probabilistic method to obtain a lower bound. Here we present an elegant result for $t = 2$, proved by Reiman in 1958, that indicates the connection between the problem of Zarankiewicz and designs, in particular projective spaces, and we shall conclude the section with some recent results for $t \geq 3$.

Theorem 12 *For $n \geq 1$, we have*

$$z(n, n; 2, 2) \leq \frac{1}{2} n\{1 + (4n-3)^{1/2}\},$$

and equality holds for infinitely many values of n. Furthermore,

$$\mathrm{ex}(n, C_4) \leq \frac{n}{4}(1 + \sqrt{4n-3}).$$

Proof. Since $2\mathrm{ex}(n, K_{s,t}) \leq z(n, n; s, t)$, the second inequality is immediate from the first. Moreover, the first inequality is just the case $s = 2$ of (4). In fact, the proof of Lemma 9 tells us a considerable amount about the graphs G for which equality is attained. We must have $d_1 = d_2 = \cdots = d_n = d$, and any two vertices in the second vertex class V_2 have degree d and any two vertices in V_1 have exactly one common neighbour. Also, precisely the same assertions hold with V_1 and V_2 interchanged.

Call the vertices in V_2 *points* and the sets $\Gamma(x)$, $x \in V_1$, *lines*. By the remarks above there are n points and n lines, each point is on d lines; and each line contains d points, there is exactly one line through any two points and any two lines meet in exactly one point. Thus we have arrived at the *projective plane* of order $d - 1$. Since the steps are easy to trace back, we see that equality holds for every n for which there is a projective plane with n points. In particular, equality holds for every $n = q^2 + q + 1$ where q is a prime power.

In conclusion, let us see the actual construction of G for the above values of n. Let q be a prime power and let $PG(2, q)$ be the projective plane over the field of order q. Let V_1 be the set of points and V_2 the set of lines. Then

$$|V_1| = |V_2| = q^2 + q + 1 = n.$$

Let G be the bipartite graph $G_2(n, n)$ with vertex classes V_1 and V_2 in which we join a point $P \in V_1$ to a line $\ell \in V_2$ by an edge iff the point P is on the line ℓ. (For $q = 2$ this gives us the Heawood graph, shown in Fig. I.7.) Then G has $n(q+1) = \frac{1}{2}n\{1 + (4n-3)^{1/2}\}$ edges, and it does not contain a quadrilateral. □

A variant of the construction above can be used to show that the bound for $\mathrm{ex}(n; K_{2,2}) = \mathrm{ex}(n; C_4)$ given in Theorem 11 is also essentially best possible.

The results for $K_{3,3}$ are almost as satisfactory as the results above for $K_{2,2} = C_4$. We see from (5) and (6) that $z(n, n; 3, 3) \le (2^{1/3} + o(1))n^{5/3}$ and $\mathrm{ex}(n, K_{3,3}) \le \frac{1}{2}(2^{1/3}+o(1))n^{5/3}$. By using an ingenious construction based on finite geometries, Brown showed in 1966 that $z(n, n; 3, 3) \ge (1 + o(1))n^{5/3}$ and so $\mathrm{ex}(n, K_{3,3}) \ge (\frac{1}{2} + o(1))n^{5/3}$. Thus (5) and (6) do give the correct orders for $K_{3,3}$. However, 30 years passed before it was proved that the constants (1 and 1/2) in the lower bounds of Brown are best possible. In 1996, Füredi gave the first substantial improvement on the simple upper bound in Theorem 10 when he proved that for $1 \le t \le s \le m$ we have

$$z(m, n; s, t) \le (s - t + 1)^{1/t} nm^{1-1/t} + tn + tm^{2-2/t}. \tag{7}$$

Combining (7) with the lower bound given by Brown, we see that $z(n, n; 3, 3) = (1 + o(1))n^{5/3}$ and $\mathrm{ex}(n, K_{3,3}) = (\frac{1}{2} + o(1))n^{5/3}$.

In spite of all these results, much remains to be done. It is very likely that (5) and (6) not only give the correct orders but $z(n, n; t, t) = (c_t + o(1))n^{2-(1/t)}$ and $\mathrm{ex}(n, K_{t,t}) = \frac{1}{2}(c_t + o(1))n^{2-(1/t)}$. Even more, perhaps Füredi's inequality (7) is essentially best possible and $c_t = 1$.

IV.3 Hamilton Paths and Cycles

A class of graphs is said to be *monotone* if whenever a graph L belongs to the class and M is obtained from L by adding to it an edge (but no vertex) then M also belongs to the class. Most theorems in graph theory can be expressed by saying that a monotone class $\mathcal{M}$ is contained in a monotone class $\mathcal{P}$. Of course, these classes are usually described in terms of graph invariants or subgraphs contained by them. For example, the simplest case of Turán's theorem, discussed in the previous section, states that the class $\mathcal{M} = \{G(n, m) : m > n^2/4\}$ is contained in $\mathcal{P} = \{G : G \text{ contains a triangle}\}$. It is worth noting that a class $\mathcal{P}$ of graphs is said to be a *property of graphs* if $L \in \mathcal{P}$ and $L \cong M$ imply $M \in \mathcal{P}$.

How should we go about deciding whether $\mathcal{M}$ is contained in $\mathcal{P}$? Bondy and Chvátal showed in 1976 that in some cases there is a simple and beautiful way of tackling this problem. Suppose we have a class $\mathcal{T}$ of triples (G, x, y), where G is a graph and x and y are non-adjacent vertices of G, such that if $(G, x, y) \in \mathcal{T}$ and

$G \in \mathcal{M}$, then G belongs to $\mathcal{P}$ iff $G^+ = G + xy$ does. This holds, for example, if $\mathcal{P}$ is the property of containing a K_r and $\mathcal{T} = \{(G, x, y) : |\Gamma(x) \cap \Gamma(y)| < r-2\}$. In this case G can be replaced by G^+. If G^+ also contains two non-adjacent vertices, say u and v, such that $(G^+, u, v) \in \mathcal{T}$, then we can repeat the operation: we can replace G^+ by $G^{++} = G^+ + uv$. Continuing in this way we arrive at a graph $G^* \supset G$ that belongs to $\mathcal{P}$ if G does and that is a *closure of G with respect to $\mathcal{T}$*; that is, it has the additional property that for no vertices $a, b \in G^*$ does $(G^*, a, b) \in \mathcal{T}$ hold. Thus it is sufficient to decide about these graphs $G^* \in \mathcal{M}$ whether or not they belong to $\mathcal{P}$.

Of course, the method above is feasible only if (i) the class $\mathcal{T}$ is simple enough, (ii) it is easy to show that G belongs to $\mathcal{P}$ iff G^+ does, and (iii) if we start with a graph $G \in \mathcal{M}$, then a graph $G^* \in \mathcal{M}$ is easily shown to belong to $\mathcal{P}$. In this section we give two examples due to Bondy and Chvátal that satisfy all these requirements: we shall give sufficient conditions for a graph to contain a Hamilton cycle or a Hamilton path. Because of the special features of these examples it will be convenient to use slightly different notation and terminology.

Let n and k be natural numbers and let $\mathcal{P}$ be a class of graphs of order n. We say that $\mathcal{P}$ is *k-stable* if whenever G is an arbitrary graph of order n, and x and y are non-adjacent vertices of G with $d(x) + d(y) \geq k$, then G has property $\mathcal{P}$ iff $G^+ = G + xy$ has it also. It is easily seen that for every graph G of order n there is a *unique minimal graph $G^* = C_k(G)$ containing G* such that

$$d_{G^*}(x) + d_{G^*}(y) \leq k - 1 \text{ for } xy \notin E(G^*).$$

In the notation of the previous paragraph, we shall take

$$\mathcal{T} = \{G, x, y) : |G| = n, \ xy \notin E(G), \ d(x) + d(y) \geq k\},$$

which is certainly simple enough, so (i) will be satisfied. It is also encouraging that $G^* = C_k(G)$ is unique. Almost by definition we have the following *principle of stability*: if $\mathcal{P}$ is a k-stable property of graphs of order n, then G *has property $\mathcal{P}$ iff $C_k(G)$ has it also.* We call $C_k(G)$ the *k-closure* of G.

Requirement (ii) is also satisfied, since the gist of the proof of Theorem 2 is that the property of containing a Hamilton cycle is n-stable and the property of containing a Hamilton path is $(n-1)$-stable. Indeed, if $d(x) + d(y) \geq n - 1$ whenever x and y are nonadjacent distinct vertices, then the graph is connected, so the proof of Theorem 2 can be applied. (In fact, this is exactly what motivated the notion of a k-closure.) By the stability principle we obtain the following reformulation of Theorem 2 in the case $k = n$ or $n - 1$.

Lemma 13 *A graph G is Hamiltonian iff $C_n(G)$ is, and G has a Hamilton path iff $C_{n-1}(G)$ does.* □

Depending on the amount of work we are able and willing to put in at this stage (cf. requirement (iii)), we obtain various sufficient conditions for a graph to be Hamiltonian. Of course, the case $k = n$ of Theorem 2 is obtained without any work, and so is the case $k = n - 1$, since the conditions imply immediately that $C_n(G) = K_n$ in the first case and $C_{n-1}(G) = K_n$ in the second, and K_n is

Hamiltonian if $n \geq 3$. In order to make better use of Lemma 13, we shall prove the following ungainly technical lemma.

Lemma 14 *Let G be a graph with vertex set $V(G) = \{x_1, x_2, \ldots, x_n\}$, whose k-closure $C_k(G)$ contains at most $t \leq n-2$ vertices of degree $n-1$. Then there are indices i, j, $1 \leq i < j \leq n$, such that $x_i x_j \in E(G)$ and each of the following four inequalities holds:*

$$
\begin{gathered}
j \geq \max\{2n - k - i, n - t\}, \\
d(x_i) \leq i + k - n, \qquad d(x_j) \leq j + k - n - 1, \\
d(x_i) + d(x_j) \leq k - 1.
\end{gathered} \tag{8}
$$

Remark. It is not assumed that the degree sequence $d(x_1), d(x_2), \ldots, d(x_n)$ of G is ordered in any way.

Proof. The graph $H = C_k(G)$ is not complete so, we can define two indices i and j as follows:

$$
\begin{aligned}
j &= \max\{\ell : d_H(x_\ell) \neq n-1\}, \\
i &= \max\{\ell : x_\ell x_j \notin E(H)\}.
\end{aligned}
$$

Then $x_i x_j \notin E(H)$, so

$$d_H(x_i) + d_H(x_j) \leq k - 1,$$

which implies the fourth inequality in (8). Each of the vertices

$$x_{j+1}, x_{j+2}, \ldots, x_n$$

has degree $n-1$ in H, so

$$n - j \leq t$$

and

$$n - j \leq \delta(H) \leq d_H(x_i).$$

The vertex x_j is joined to the $n-j$ vertices following it and to the $j-i-1$ vertices preceding it, so

$$d_H(x_j) \geq (n-j) + (j-i-1) = n - i - 1.$$

These inequalities enable us to show that the indices i, j, $1 \leq i < j \leq n$, satisfy the remaining three inequalities in (8). Indeed,

$$
\begin{aligned}
d_G(x_i) &\leq d_H(x_i) \leq k - 1 - d_H(x_j) \leq k - 1 - (n - i - 1) = i + k - n, \\
d_G(x_j) &\leq d_H(x_j) \leq k - 1 - d_H(x_i) \leq k - 1 - (n - j) = j + k - n - 1,
\end{aligned}
$$

and

$$i + j \geq (n - d_H(x_j) - 1) + (n - d_H(x_i)) \geq 2n - 1 - (k-1) = 2n - k,$$

completing the proof. □

Combining Lemma 13 and Lemma 14 (with $t = n - 2$ and $k = n$ or $n - 1$) we obtain the following theorem of Bondy and Chvátal, giving rather complicated but useful conditions for the existence of a Hamilton path or cycle.

Theorem 15 *Let G be a graph with vertex set $V(G) = \{x_1, x_2, \ldots, x_n\}$, $n \geq 3$. Let $\varepsilon = 0$ or 1 and suppose there are no indices i, j, $1 \leq i < j \leq n$, such that $x_i x_j \notin E(G)$ and*

$$j \geq n - i + \varepsilon,$$
$$d(x_i) \leq i - \varepsilon, \qquad d(x_j) \leq j - 1 - \varepsilon,$$
$$d(x_i) + d(x_j) \leq n - 1 - \varepsilon.$$

If $\varepsilon = 0$ then G has a Hamilton cycle, and if $\varepsilon = 1$ then G has a Hamilton path. □

An immediate consequence of this result is the following theorem of Chvátal.

Corollary 16 *Let G be a graph with degree sequence $d_1 \leq d_2 \leq \cdots \leq d_n$, $n \geq 3$, and let $\varepsilon = 0$ or 1. Suppose*

$$d_{n-k+\varepsilon} \geq n - k \text{ whenever } d_k \leq k - \varepsilon < \frac{1}{2}(n - \varepsilon).$$

If $\varepsilon = 0$ then G has a Hamilton cycle, and if $\varepsilon = 1$ then G has a Hamilton path. □

We draw the attention of the reader to Exercises 32– 33 which show that the assertions in the corollary above are in some sense best possible. In particular, if $d_1 \leq d_2 \leq \cdots \leq d_n$ is a graphic sequence such that

$$d_k \leq k < \frac{n}{2} \quad \text{and} \quad d_{n-k} < n - k,$$

then there is a graph G with vertex set $\{x_1, x_2, \ldots, x_n\}$ such that $d(x_i) \geq d_i$, $1 \leq i \leq n$, and G does not have a Hamilton cycle.

There is another customary way of showing that a graph has a Hamilton cycle or path. Let S be a longest x_0*-path* in G, that is, a longest path beginning at x_0 : $S = x_0 x_1 \cdots x_k$. Then $\Gamma(x_k) \subset \{x_0, x_1, \ldots, x_{k-1}\}$ since otherwise S could be continued to a longer path. If x_k is adjacent to x_j, $0 \leq j < k - 1$, then $S' = x_0 x_1 \cdots x_j x_k x_{k-1} \cdots x_{j+1}$ is another longest x_0-path. We call S' a *simple transform* of S. It is obtained from S by erasing the edge $x_j x_{j+1}$ and adding to it the edge $x_k x_j$. Note that if S' is a simple transform of S, then S is a simple transform of S' and S has exactly $d(x_k) - 1$ simple transforms. The result of a sequence of simple transforms is called a transform (see Fig. IV.4).

The theorem below is usually called *Pósa's lemma*: as we shall see in Chapter VII, it can be used to prove the existence of Hamilton cycles in random graphs. To present it, let L be the set of endvertices (different from x_0) of transforms of S and put $N = \{x_j \in S : x_{j-1} \in L \text{ or } x_{j+1} \in L\}$ and $R = V \setminus N \cup L$. Thus L is the collection of the last vertices of the transforms, N is the collection of their neighbours on S and R is the rest of the vertices.

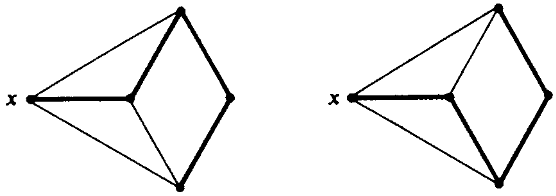


FIGURE IV.4. An x-path and a simple transform of it.

Theorem 17 *The graph G has no L–R edges.*

Proof. Recall that there is no edge between L and $V(G) \setminus V(S)$, since S is a longest x_0-path, so in particular $V(S) = V(P)$ for every transform P of S.

Suppose $x_i x_j \in E(G)$, where $x_i \in L$ and $x_j \in R$. Let S_i be a transform of S ending in x_i. Since at least one neighbour of x_j on S_i is the endvertex of a simple transform of S_i, x_j cannot have the same neighbours on S and S_i, since otherwise x_j would belong to N. However, when the edge $x_{j'}x_j$, $j' = j-1$ or $j+1$, is erased during a sequence $S \to S' \to S'' \to \cdots \to S_i$ of simple transformations, one of the vertices $x_{j'}, x_j$ is put into L and the other into N. Thus $x_j \in L \cup N = V(G) \setminus R$, contradicting our assumption. □

The theorems of this section are also obtained with the use of simple transforms: they are due to Thomason, who extended earlier results of Smith.

Theorem 18 *Let W be the set of vertices of even degree in a graph G and let x_0 be a vertex of G. Then there is an even number of longest x_0-paths ending in W.*

Proof. Let H be the graph whose vertex set is the set Σ of longest x_0-paths in G, in which $P_1 \in \Sigma$ is joined to $P_2 \in \Sigma$ if P_2 is a simple transform of P_1. Since the degree of $P = x_0 x_1 \cdots x_k \in \Sigma$ in H is $d(x_k) - 1$, the set of longest paths ending in W is exactly the set of vertices of odd degree in H. The number of vertices of odd degree is even in any graph, so the proof is complete. □

Theorem 19 *Let G be a graph in which every vertex has odd degree. Then every edge of G is contained in an even number of Hamilton cycles.*

Proof. Let $x_0 y \in E(G)$. Then in $G' = G - x_0 y$ only x_0 and y have even degree, so in G' there is an even number of longest x_0-paths that end in y. Thus either G has no Hamilton cycle that contains $x_0 y$ or it has a positive even number of them. □

The most striking case of Theorem 19 is that in a cubic graph every edge is contained in an even number of Hamilton cycles; in particular, for every edge of a Hamilton cycle there is another Hamilton cycle containing the edge.

IV.4 The Structure of Graphs

The Turán graph $T_r(n)$ does not contain a complete graph of order $r+1$, and by (1) it has at least $(1-\frac{1}{r})\binom{n}{2}$ edges. Therefore, a graph of order n and size at least $(1-\frac{1}{r})\binom{n}{2}$ need not contain a K_{r+1}. The main aim of this section is to prove a deep result of Erdős and Stone, published in 1946, that if $\varepsilon > 0$ is fixed then εn^2 more edges ensure not only a K_{r+1}, but a $K_{r+1}(t)$, a complete $(r+1)$-partite graph with t vertices in each class, with $t \to \infty$ as $n \to \infty$. The Erdős–Stone theorem is rightly called the fundamental theorem of extremal graph theory.

A considerably sharper result, giving the correct speed $\log n$ for $t \to \infty$, was published by Bollobás and Erdős in 1976; this is the theorem we shall prove. To be precise, in order to make the calculations more pleasant, we shall present only a weaker form of this result.

For $r = 1$ the problem is precisely the Zarankiewicz problem discussed in Section 2, but this time for rather dense graphs, with $t \to \infty$. What we want can be read out of (6), but as we shall be satisfied with an even simpler result, we run through the argument. We claim that if $\varepsilon > 0$ is fixed and n is large enough, then every graph G of order n and minimal degree at least εn contains a $K_2(t)$ with $t \geq \varepsilon \log n$.

To prove this, suppose G does not contain a $K_2(t)$. As before, we say that a set of t vertices is *covered* by a vertex x if x is joined to every vertex in the set. Every vertex of G covers at least $\binom{\varepsilon n}{t}$ sets of t vertices, and no set of t vertices is covered by t vertices. Therefore,

$$n\binom{\varepsilon n}{t} < t\binom{n}{t}.$$

This inequality is false for $t = \lceil \varepsilon \log n \rceil$ and n large, since then

$$\begin{aligned} t\binom{n}{t}\Big/ n\binom{\varepsilon n}{t} &\leq \frac{t}{n}\varepsilon^{-t}\left(1 - \frac{t}{\varepsilon n}\right)^{-t} < \frac{2t}{n}\varepsilon^{-t} \\ &< \frac{2t}{\varepsilon n} e^{\log(1/\varepsilon)\varepsilon \log n} \leq \frac{2t}{\varepsilon n} n^{1/\varepsilon} < 1. \end{aligned}$$

What we have just proved is the case $r = 1$ of the theorem below; this result is only slightly weaker than the form of the Erdős–Stone theorem to be given as Theorem 22.

Theorem 20 *Let $r \geq 1$ be an integer and let $\varepsilon > 0$. Then there is an integer $n_0 = n_0(r_1\varepsilon)$ such that if $|G| = n \geq n_0$ and*

$$\delta(G) \geq \left(1 - \frac{1}{r} + \varepsilon\right) n,$$

then $G \supset K_{r+1}(t)$, where

$$t \geq \frac{\varepsilon \log n}{2^{r-1}(r-1)!}.$$

Proof. We apply induction on r. As the case $r = 1$ was proved above, we proceed to the induction step. Let then $r \geq 2$ and let G be a graph with n vertices and minimal degree at least $(1 - \frac{1}{r} + \varepsilon)n$. Note that $0 < \varepsilon < 1/r$. Since

$$\delta(G) > \left(1 - \frac{1}{r} + \frac{1}{r(r-1)}\right)n,$$

by the induction hypothesis G contains a $K_r(T) = K$, say, with $|T| = \lceil d_r \log n \rceil$ vertices in one class, where $d_r = 2^{2-r}/r!$. Let U be the set of vertices in $G - K$ joined to at least $(1 - \frac{1}{r} + \frac{\varepsilon}{2})|K|$ vertices of K. We claim that

$$|U| \geq \varepsilon n.$$

To see this, note that the number f of edges between K and $G - K$ satisfies

$$|K|\{(1 - \frac{1}{r} + \varepsilon)n - |K|\} \leq f \leq |U||K| + (n - |U|)(1 - \frac{1}{r} + \frac{\varepsilon}{2})|K|,$$

that is,

$$\frac{r\varepsilon n}{n} \leq |U|(1 - \frac{r\varepsilon}{2}) + r|K|.$$

This implies that $|U| \geq r\varepsilon n/2 \geq \varepsilon n$ if n is large enough, so our claim is justified.

Set $t = \lceil \log n/2^{r-1}(r-1)! \rceil$. Then $t \leq \lceil (r\varepsilon/2)T \rceil$, so

$$\lceil (1 - \frac{1}{r} + \frac{\varepsilon}{2})|K| \rceil = \lceil (r-1)T + (\varepsilon r/2)T \rceil \geq (r-1)T + t.$$

Calling a subgraph H of G *covered* by a vertex x if x is joined to every vertex of H, this inequality shows that every vertex of U covers at least one $K_r(t)$ subgraph of K. In K there are only $\binom{T}{t}^r$ such subgraphs, so there is a set $W \subset U$,

$$|W| \geq |U| / \binom{T}{t}^r,$$

such that every vertex of W covers the same $K_r(t)$ subgraph of K. To complete the proof, all we have to check is that $|W| \geq t$. Now, $t/eT > \varepsilon/3$, and by Stirling's formula, $t \geq (t/e)^t$, so

$$\begin{aligned} |W| &\geq \varepsilon n \left(\frac{t}{eT}\right)^{tr} \geq \varepsilon n(\varepsilon/3)^{tr} \\ &> \varepsilon n(\varepsilon/3)^r \exp\{\log(\varepsilon/3) r\varepsilon \log n/2^{r-1}(r-1)!\}. \end{aligned}$$

Since $r \leq 2^{r-1}(r-1)!$ and $\log(\varepsilon/3)\varepsilon \geq \log(1/6)/2 > -1$, we have $|W| > t$, and we are done. □

The following observation enables us to weaken the condition above on the minimal degree to a condition on the size of a graph.

Lemma 21 *Let $c, \varepsilon > 0$. If n is sufficiently large, say $n > 3/\varepsilon$, then every graph of order n and size at least $(c + \varepsilon)\binom{n}{2}$ contains a subgraph H with $\delta(H) \geq c|H|$ and $|H| \geq \varepsilon^{1/2}n$.*

Proof. Let G be a graph of order $n > 3/\varepsilon$ and size $e(G) \geq (c+\varepsilon)\binom{n}{2}$. Note that in this case $0 < \varepsilon < \varepsilon + c \leq 1$. If the assertion fails then there is a sequence of graphs $G_n = G \supset G_{n-1} \supset \cdots \supset G_\ell$, $\ell = \lfloor \varepsilon^{1/2} n \rfloor$, such that $|G_j| = j$ and for $n \geq j > \ell$ the only vertex of G_j not in G_{j-1} has degree less than cj in G_j. Then

$$\begin{aligned} e(G_\ell) &> (c+\varepsilon)\binom{n}{2} - \sum_{j=\ell+1}^{n} cj = (c+\varepsilon)\binom{n}{2} - c\left\{\binom{n+1}{2} - \binom{\ell+1}{2}\right\} \\ &> \varepsilon\binom{n}{2} + c\binom{\ell+1}{2} - cn > \varepsilon\binom{n}{2} > \binom{\ell}{2}, \end{aligned}$$

since $0 < \varepsilon < 1$ and $n \geq 3/n$. This contradiction completes the proof. □

Putting together Theorem 20 and Lemma 21, we obtain a strengthening of the Erdős–Stone theorem of 1946, published by Bollobás and Erdős in 1973.

Theorem 22 *Let $r \geq 1$ be an integer and let $\varepsilon > 0$. Then there is an integer $n_0 = n_0(r, \varepsilon)$ such that if $|G| = n \geq n_0$ and*

$$e(G) \geq (1 - \frac{1}{r} + \varepsilon)\binom{n}{2},$$

then $G \supset K_{r+1}(t)$ for some $t \geq \varepsilon \log n/(2^{r+1}(r-1)!)$.

Proof. If $n > 3/\varepsilon$ then, by Lemma 21, G has a subgraph H with $|H| = h \geq \varepsilon^{1/2} n$ and $\delta(H) \geq (1 - \frac{1}{r} + \varepsilon/2)h$. Hence if n is sufficiently large then H contains a $K_{r+1}(t)$ with $t \geq \frac{\varepsilon}{2} \log h/(2^{r-1}(r-1)!) \geq \varepsilon \log n/(2^{r+1}(r-1)!)$, as claimed. □

The function $n_0(r, \varepsilon)$ appearing in Theorem 22 is not that large: one can check that $n_0(r, \varepsilon) = \max\{\lceil 3/\varepsilon \rceil, 100\}$ will do (see Exercise 58).

In a certain sense Theorem 22 is best possible: for every ε and r there is a constant d_r^* tending to 0 with ε such that the graph described in the theorem need not contain a $K_r(t)$ with $t = \lfloor d_r^* \log n \rfloor$. In fact, we shall see in Theorem VII.3 that if $0 < \varepsilon < \frac{1}{2}$ and $d_2^* > -2/\log(2\varepsilon)$, then for every sufficiently large n there is a graph $G(n, m)$ not containing a $K_2(t)$, where $m = \lfloor \varepsilon n^2 \rfloor$ and $t = \lfloor d_2^* \log n \rfloor$. This result will imply immediately (cf. Exercise VII.13) that if $r > 2$ and $0 < \varepsilon < \frac{1}{2}(r-1)^{-2}$ then any value greater than $-2/\log(2(r-1)^2\varepsilon)$ will do for d_r^*.

The fact that this example gives the correct speed for $d(\varepsilon, r)$ is a much deeper result: this was proved by Chvátal and Szemerédi in 1981, by making use of a very powerful tool, Szemerédi's regularity lemma, to be presented in the next section.

Since $d \log n \to \infty$ as $n \to \infty$, Theorem 22 has the following immediate corollaries. The first is a slightly weaker form of the original Erdős–Stone theorem.

Corollary 23 *Let $F = K_{r+1}(t)$, where $r \geq 1$ and $t \geq 1$. Then the maximal size of a graph of order n without a $K_{r+1}(t)$ is*

$$\mathrm{ex}(n; F) = (1 - \frac{1}{r})\binom{n}{2} + o(n^2).$$

□

Corollary 24 *Let $F_1, F_2, \ldots, F_\ell$ be non-empty graphs. Denote by $r+1$ the minimum of the chromatic numbers of the F_i, that is, let $r+1$ be the minimal number for which at least one of the F_i is contained in an $F = K_{r+1}(t)$ for some t. Then the maximal size of a graph of order n not containing any of the F_i is*

$$\mathrm{ex}(n; F_1, F_2, \ldots, F_\ell) = (1 - \frac{1}{r})\binom{n}{2} + o(n^2).$$

Proof. The Turán graph $T_r(n)$ does not contain any of the F_i so, by (1),

$$\mathrm{ex}(n; F_1, F_2, \ldots, F_\ell) \geq e(T_{r-1}(n)) = t_{r-1}(n) \leq (1 - \frac{1}{r})\binom{n}{2}.$$

Conversely, since, say $F_j \subset F = K_{r+1}(t)$ for some j and t,

$$\mathrm{ex}(n; F_1, F_2, \ldots, F_\ell) \leq \mathrm{ex}(n; F_j) \leq \mathrm{ex}(n; F) = (1 - \frac{1}{r})\binom{n}{2} + o(n^2). \qquad \square$$

Theorem 22 is the basis of a rather detailed study of the structure of extremal graphs, initiated by Erdős and Simonovits, giving us considerably more accurate results than Corollary 24. This theory is, however, outside the scope of our book.

The *density* of a graph G of order n is defined to be $e(G)/\binom{n}{2}$. The *upper density* of an *infinite* graph G is the supremum of the densities of arbitrarily large *finite* subgraphs of G. It is surprising and fascinating that not every value between 0 and 1 is the upper density of some infinite graph; in fact, the range of the upper density is a countable set.

Corollary 25 *The upper density of an infinite graph G is $1, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots,$ or 0. Each of these values is the upper density of some infinite graph.*

Proof. Let G_r be the complete r-partite graph with infinitely many vertices in each class. Since the density of $K_r(t)$ tends to $1 - \frac{1}{r}$ as t tends to ∞, the upper density of G_r is $1 - \frac{1}{r}$, proving the second assertion.

Now, let α be the upper density of G and suppose that

$$\alpha > 1 - \frac{1}{r-1},$$

where $r \geq 2$. Then there is an $\varepsilon > 0$ such that G contains graphs H_k of order n_k with $n_k \to \infty$ satisfying

$$e(H_k) \geq \frac{1}{2}\left(1 - \frac{1}{r-1} + \varepsilon\right) n_k^2.$$

By Theorem 20 each H_k contains a subgraph $K_r(t_k)$ with $t_k \to \infty$; the subgraphs $K_r(t_k)$ show that $\alpha \geq \frac{1}{r}$. $\square$

The results above give fairly satisfactory answers to the forbidden subgraph problem, provided that no forbidden subgraph is bipartite, and Erdős and Simonovits have proved several considerably stronger results. However, for a general bipartite graph F, the result $\mathrm{ex}(n; F) = o(n^2)$ is rather feeble, and for most bipartite graphs F we cannot even determine the exponent α of n for which

$$0 < \lim_{n\to\infty} \mathrm{ex}(n; F)/n^{\alpha} < \infty.$$

Also, we have only rudimentary results for hypergraphs, so much remains to be done.

IV.5 Szemerédi's Regularity Lemma

In 1975 Szemerédi proved one of the most beautiful results in combinatorics: every set of natural numbers with positive upper density contains arbitrarily long arithmetic progressions (see Section VI.4). A crucial step in the proof was an innocent looking lemma, which has turned out to be of vital importance in attacking a great variety of extremal problems. This lemma has come to be called Szemerédi's regularity lemma, although 'uniformity' rather than 'regularity' would be much closer to the mark. Roughly speaking, the lemma claims that the vertex set of *every* graph can be partitioned into boundedly many almost equal classes such that most pairs of classes are 'regular', in the sense that the number of edges between two subsets of the classes is about proportional to the possible number of edges between the subsets, provided that the subsets are not too small. Thus for a 'regular' pair of classes it does not happen that some two k-subsets span many edges while some others span few edges. In order to formulate the lemma precisely, we need some definitions and notation.

Given a graph $G = (V, E)$ and a pair (X, Y) of disjoint non-empty subsets of V, denote by $e(X, Y) = e_G(X, Y)$ the number of X–Y edges of G, and write $d(X, Y) = d_G(X, Y) = e(X, Y)/(|X||Y|)$ for the *density* of the X–Y edges of G. Call (X, Y) an *ε-uniform pair* if

$$|d(X^*, Y^*) - d(X, Y)| < \varepsilon$$

whenever $X^* \subset X$ and $Y^* \subset Y$ are such that $|X^*| \geq \varepsilon|X| > 0$ and $|Y^*| \geq \varepsilon|Y| > 0$. A partition $\mathcal{P} = (C_i)_{i=0}^{k}$ of the vertex set V is said to be an *equitable partition with exceptional class C_0* if $|C_1| = |C_2| = \cdots = |C_k|$. Finally, an *$\varepsilon$-uniform partition* is an equitable partition $(C_i)_{i=0}^{k}$ such that the exceptional class C_0 has at most εn vertices and, with the exception of at most εk^2 pairs, the pairs (C_i, C_j), $1 \leq i < j \leq k$, are ε-uniform.

Szemerédi's lemma (Theorem 29) states that every graph has an ε-uniform partition with a bounded number of classes. We begin the proof with two easy lemmas: the first concerns the densities $d(X, Y)$ and the second is a simple inequality. Readers are encouraged to skip the proofs.

Lemma 26 *Suppose that X and Y are disjoint sets of vertices of a graph G, and $X^* \subset X$ and $Y^* \subset Y$ are such that $|X^*| \geq (1-\gamma)|X| > 0$ and $|Y^*| \geq (1-\delta)|Y| > 0$. Then*

$$|d(X^*, Y^*) - d(X, Y)| \leq \gamma + \delta \qquad (9)$$

and

$$|d^2(X^*, Y^*) - d^2(X, Y)| < 2(\gamma + \delta). \tag{10}$$

Proof. Note that, rather crudely,

$$0 \le e(X, Y) - e(X^*, Y^*) \le (\gamma + \delta - \gamma\delta)|X||Y| < (\gamma + \delta)|X||Y|,$$

so

$$d(X, Y) - d(X^*, Y^*) \le \frac{e(X, Y) - e(X^*, Y^*)}{|X||Y|} < \gamma + \delta.$$

If G is replaced by its complement $\overline{G}$, then each density d changes to $1 - d$, so

$$d_G(X^*, Y^*) - d_G(X, Y) = d_{\overline{G}}(X, Y) - d_{\overline{G}}(X^*, Y^*)) < \gamma + \delta,$$

completing the proof of (9).

Inequality (10) is an immediate consequence of (9):

$$\begin{aligned}|d^2(X^*, Y^*) &- d^2(X, Y)| \\ &= |d(X^*, Y^*) + d(X, Y)||d(X^*, Y^*) - d(X, Y)| < 2(\gamma + \delta). \quad \square\end{aligned}$$

Lemma 27 *Let $(d_i)_{i=1}^s \subset \mathbb{R}$, $1 \le t < s$, $D = \frac{1}{s}\sum_{i=1}^s d_i$, and $d = \frac{1}{t}\sum_{i=1}^t d_i$. Then*

$$\frac{1}{s}\sum_{i=1}^s d_i^2 \ge D^2 + \frac{t}{s-t}(D-d)^2 \ge D^2 + \frac{t}{s}(D-d)^2.$$

In particular, if $t \ge \gamma s$ and $|D - d| \ge \delta > 0$, then

$$\frac{1}{s}\sum_{i=1}^s d_i^2 \ge D^2 + \gamma\delta^2.$$

Proof. With

$$e = \frac{1}{s-t}\sum_{i=t+1}^s d_i = \frac{sD - td}{s-t},$$

the convexity of the function x^2 implies that

$$\begin{aligned}\sum_{i=1}^s d_i^2 &= \sum_{i=1}^t d_i^2 + \sum_{i=t+1}^s d_i^2 \ge td^2 + (s-t)e^2 \\ &= td^2 + \frac{s^2D^2 - 2stdD + t^2d^2}{s-t} \\ &= sD^2 + \frac{st}{s-t}(D-d)^2.\end{aligned}$$

$\square$

Given an equitable partition $\mathcal{P} = (C_i)_{i=0}^k$ with exceptional class C_0, let us define the *square mean* of $\mathcal{P}$ as

$$q(\mathcal{P}) = \frac{1}{k^2} \sum_{1 \le i < j \le k} d^2(C_i, C_j).$$

As $d^2(C_i, C_j) \le 1$ for all i and j and the sum above has $\binom{k}{2}$ terms, we have $0 \le q(\mathcal{P}) < \frac{1}{2}$.

The final lemma, which is the cornerstone of the proof of the regularity lemma, claims that if $\mathcal{P}$ is not ε-uniform although C_0 is small enough, then there is a partition $\mathcal{P}' = (C_i')_{i=0}^{\ell}$, with ℓ a given function of k, such that $q(\mathcal{P}')$ is *appreciably larger* than $q(\mathcal{P})$ and $|C_0'|$ is only a *little larger* than $|C_0|$. To find an ε-uniform partition, all we need then is to repeatedly replace an appropriate starting partition by a partition appreciably increasing the square mean. The process ends after boundedly many steps in an ε-uniform partition.

Lemma 28 *Let G be a graph of order n with an equitable partition $V = \bigcup_{i=0}^k C_i$ of the vertex set with exceptional class C_0 and*

$$|C_1| = |C_2| = \cdots = |C_k| = c \ge 2^{3k+1}.$$

Suppose that the partition $\mathcal{P} = (C_i)_{i=0}^k$ is not ε-uniform, where $0 < \varepsilon < \frac{1}{2}$ and $2^{-k} \le \varepsilon^5/8$. Then there is an equitable partition $\mathcal{P}' = (C_i')_{i=0}^{\ell}$ with $\ell = k(4^k - 2^{k-1})$ and exceptional class $C_0' \supset C_0$ such that

$$|C_0'| \le |C_0| + \frac{n}{2^k}$$

and

$$q(\mathcal{P}') \ge q(\mathcal{P}) + \frac{\varepsilon^5}{2}.$$

Proof. For a pair (C_i, C_j) that is not ε-uniform, let $C_{ij} \subset C_i$ and $C_{ji} \subset C_j$ be sets showing that (C_i, C_j) is not ε-uniform: $|C_{ij}| \ge \varepsilon|C_i| = \varepsilon c$, $|C_{ji}| \ge \varepsilon|C_j| = \varepsilon c$, and

$$|d(C_{ij}, C_{ji}) - d(C_i, C_j)| \ge \varepsilon. \tag{11}$$

Furthermore, for an ε-uniform pair (C_i, C_j), set $C_{ij} = C_{ji} = \emptyset$.

Ideally, we would like to partition each C_i into a few (according to the statement of the lemma, into $4^k - 2^{k-1}$) sets C_h' of size d, say, such that each C_{ij} is the exact union of some of these sets C_h'. In this way, a large difference $|d(C_{ij}, C_{ji}) - d(C_i, C_j)|$ would guarantee, by Lemma 27, that the part of $q(\mathcal{P}')$ arising from $d^2(C_i, C_j)$, namely

$$\frac{d^2}{c^2} \sum \left\{ d^2(C_g', C_h') : \ C_g' \subset C_{ij}, C_h' \subset C_{ji} \right\},$$

is appreciably larger than $d^2(C_i, C_j)$. In turn, this would imply that $q(\mathcal{P}')$ is considerably larger than $q(\mathcal{P})$.

Although we cannot construct sets C'_h such that each C_{ij} is the exact union of some of these sets, we can come fairly close to it: we can achieve that each C_{ij} is almost the union of the sets C'_h it contains. We do this by considering, for each C_i, all the sets C_{ij} *at once*, and choosing the future C'_g sets (to be denoted by D_{ih}) such that they do not cut across any C_{ij}. The price we have to pay is simply that we cannot quite partition C_i into the sets C'_g, so we have to add the remainder to the "rubbish bin" C_0 to obtain a slightly larger exceptional set C'_0.

In order to carry out our plan, for each i, $1 \le i \le k$, consider the *atoms* of the algebra on C_i induced by the sets C_{ij}, $1 \le j \le k$, $j \ne i$. These atoms are the equivalence classes of the equivalence relation $\sim$ on C_i in which $x \sim y$ means that $x \in C_{ij}$ iff $y \in C_{ij}$. Note that C_i has at most 2^{k-1} atoms.

Set $d = \lfloor c/4^k \rfloor$, so that $d \ge 2^{k+1}$ and $4^k d \le c \le 4^k(d+1) - 1$, and put $H = 4^k - 2^{k-1}$. Let D_{ih}, $1 \le h \le H$, be pairwise disjoint d-subsets of C_i such that each D_{ih} is contained in some atom of C_i. It is possible to choose such sets D_{ih} since all but at most $d-1$ elements of each atom can be partitioned into d-subsets, and

$$Hd + 2^{k-1}(d-1) = 4^k d - 2^{k-1} < 4^k d \le c.$$

For $i, j = 1, \ldots, k, i \ne j$, set $\overline{C}_i = \bigcup_{h=1}^{H} D_{ih}$ and $\overline{C}_{ij} = \bigcup\{D_{ih} : D_{ih} \subset C_{ij}\} = \bigcup_{D_{ih} \subset C_{ij}} D_{ih}$. Our first aim is to show that $d(C_i, C_j)$ and $d(\overline{C}_i, \overline{C}_j)$ are close, and so are $d(\overline{C}_{ij}, \overline{C}_{ji})$ and $d(C_{ij}, C_{ji})$ if (C_i, C_j) is not ε-uniform, so that $d(\overline{C}_i, \overline{C}_j)$ and $d(\overline{C}_{ij}, \overline{C}_{ji})$ are almost as far from each other as $d(C_i, C_j)$ and $d(C_{ij}, C_{ji})$. Now,

$$\begin{aligned} \frac{|C_i \setminus \overline{C}_i|}{|C_i|} &\le \frac{4^k(d+1) - (4^k - 2^{k-1})d}{4^k(d+1)} = \frac{4^k + 2^{k-1}d}{4^k(d+1)} \\ &< \frac{1}{d} + \frac{1}{2^{k+1}} \le 2^{-k} \le \frac{\varepsilon^5}{8}. \end{aligned} \tag{12}$$

Consequently, by Lemma 26,

$$|d(\overline{C}_i, \overline{C}_j) - d(C_i, C_j)| \le \frac{\varepsilon^5}{4} \tag{13}$$

and

$$|d^2(\overline{C}_i, \overline{C}_j) - d^2(C_i, C_j)| \le \frac{\varepsilon^5}{2}. \tag{14}$$

Hence

$$\frac{1}{k^2} \sum_{1 \le i < j \le k} d^2(\overline{C}_i, \overline{C}_j) \ge \frac{1}{k^2} \sum_{1 \le i < j \le k} d^2(C_i, C_j) - \frac{\varepsilon^5}{4}. \tag{15}$$

Suppose (C_i, C_j) is not ε-uniform. Then, by (12),

$$\frac{|C_{ij} \setminus \overline{C}_{ij}|}{|C_{ij}|} \le \frac{|C_i \setminus \overline{C}_i|}{|C_{ij}|} \le \frac{\varepsilon^4}{8} \tag{16}$$

and

$$\begin{aligned}|\overline{C}_{ij}| &\ge |C_{ij}| - |C_i \setminus \overline{C}_i| \ge (\varepsilon - \varepsilon^5/8)|C_i| \ge (1-2^{-7})\varepsilon|C_i| \\ &\ge (1-2^{-7})\varepsilon|\overline{C}_i|.\end{aligned} \tag{17}$$

Lemma 26 and inequality (16) imply that

$$|d(\overline{C}_{ij}, \overline{C}_{ji}) - d(C_{ij}, C_{ji})| \le \frac{\varepsilon^4}{4}, \tag{18}$$

and so, by (11), (13) and (18), rather crudely,

$$\begin{aligned}|d(\overline{C}_{ij}, \overline{C}_{ji}) - d(\overline{C}_i, \overline{C}_j)| &\ge |d(C_{ij}, C_{ji}) - d(C_i, C_j)| \\ &\quad - |d(\overline{C}_{ij}, \overline{C}_{ji}) - d(C_{ij}, C_{ji})| - |d(\overline{C}_i, \overline{C}_j) - d(C_i, C_j)| \\ &> \varepsilon - \frac{\varepsilon^4}{4} - \frac{\varepsilon^5}{4} > \frac{15}{16}\varepsilon.\end{aligned} \tag{19}$$

Hence, if (C_i, C_j) is not ε-uniform then, by Lemma 27,

$$\begin{aligned}\frac{1}{H^2}\sum_{u=1}^{H}\sum_{v=1}^{H} d^2(D_{iu}, D_{jv}) &\ge d^2(\overline{C}_i, \overline{C}_j) + \frac{|C_{ij}|\,|C_{ji}|}{|\overline{C}_i|\,|\overline{C}_j|}\cdot\left(\frac{15}{16}\varepsilon\right)^2 \\ &\ge d^2(\overline{C}_i, \overline{C}_j) + \left((1-2^{-7})\varepsilon\right)^2\left(\frac{15}{16}\varepsilon\right)^2 \\ &\ge d^2(\overline{C}_i, \overline{C}_j) + \frac{3}{4}\varepsilon^4,\end{aligned} \tag{20}$$

since, by inequality (17), $|\overline{C}_{ij}||\overline{C}_{ji}| \ge \left((1-2^{-7})\varepsilon\right)^2|\overline{C}_i||\overline{C}_j|$.

Also, for every pair (C_i, C_j) we have

$$\begin{aligned}\frac{1}{H^2}\sum_{u=1}^{H}\sum_{v=1}^{H} d^2(D_{iu}, D_{jv}) &\ge \left\{\frac{1}{H^2}\sum_{u=1}^{H}\sum_{v=1}^{H} d(D_{iu}, D_{jv})\right\}^2 \\ &= d^2(\overline{C}_i, \overline{C}_j).\end{aligned} \tag{21}$$

All that remains is to rename the sets D_{ih} as C'_j, $1 \le j \le \ell$, and check that the obtained partition has the required properties. Thus, let $\{C'_1, \ldots, C'_\ell\} = \{D_{ih} : 1 \le i \le k, 1 \le h \le H\}$ and $C'_0 = V \setminus \bigcup_{j=1}^{\ell} C'_j$. Then $C'_0 \supset C_0$, with

$$|C'_0 \setminus C_0| = \sum_{i=1}^{k} |C_i \setminus \overline{C}_i| \le \frac{n}{2^k},$$

with the inequality following from (12). Finally, and most importantly, (20), (15), and the fact that there are at least εk^2 pairs (C_i, C_j) that are not ε-uniform imply that

$$q(\mathcal{P}') = \frac{1}{\ell^2} \sum_{1 \le i < j \le \ell} d^2(C'_i, C'_j) \ge \frac{1}{k^2} \sum_{1 \le i < j \le k} d^2(\overline{C}_i, \overline{C}_j) + \frac{3}{4}\, \varepsilon^4 \, \frac{\varepsilon k^2}{k^2}$$

$$\ge \frac{1}{k^2} \sum_{1 \le i < j \le k} d^2(C_i, C_j) - \frac{\varepsilon^5}{4} + \frac{3}{4}\varepsilon^5$$

$$\ge q(\mathcal{P}) + \frac{\varepsilon^5}{2},$$

as claimed. □

From Lemma 28 it is a short step to *Szemerédi's regularity lemma.* Due to its importance we call it a theorem.

Theorem 29 *For $m \in \mathbb{N}$ and $0 < \varepsilon < \frac{1}{2}$ there is an integer $M = M(\varepsilon, m)$ such that every graph of order at least m has an ε-uniform partition $(C_i)_{i=0}^k$ with $m \le k \le M$.*

Proof. Set $t = \lfloor \varepsilon^{-5} \rfloor$ and define $k_0, k_1, \dots, k_{t+1}$ by letting k_0 be the minimal integer satisfying $k_0 \ge m$ and $2^{-k_0} \le \varepsilon^5/8$, and setting $k_{i+1} = k_i(4^{k_i} - 2^{k_i - 1})$. We claim that $M = k_t 2^{3k_t + 2}$ will do.

Let G be a graph of order $n \ge m$. Partitioning the vertex set of G into n singletons and the empty set as the exceptional set, we obtain a 0-uniform partition $(C_i)_{i=0}^n$. Hence in proving our claim, we may assume that $n > M$.

Let $\mathcal{P}_0 = \left(C_i^{(0)}\right)_{i=0}^{k_0}$ be an equitable partition of the vertex set of G with exceptional class $C_0^{(0)}$ such that $|C_1^{(0)}| = \cdots = |C_{k_0}^{(0)}| = \lfloor n/k_0 \rfloor$ and $0 \le |C_0^{(0)}| \le n - k_0 \lfloor n/k_0 \rfloor < k_0 < \frac{\varepsilon}{2} n$. If $\mathcal{P}_0$ is ε-uniform, we are done. Otherwise, let $\mathcal{P}_1 = \left(C_i^{(1)}\right)_{i=0}^{k_1}$ be the partition guaranteed by Lemma 28, with $|C_0^{(1)}| \le |C_0^{(0)}| + n/2^{k_0} < \varepsilon n$. Once again, if $\mathcal{P}_1$ is ε-uniform, we are done; otherwise, let $\mathcal{P}_2 = \left(C_i^{(2)}\right)_{i=0}^{k_2}$ be the partition guaranteed by Lemma 28, with $|C_0^{(2)}| \le |C_0^{(0)}| + n(2^{-k_0} + 2^{-k_1}) < \varepsilon n$. Continuing in this way, we obtain an ε-uniform partition $\mathcal{P}_j = \left(C_i^{(j)}\right)_{i=0}^{k_j}$ for some j with $0 \le j \le t$.

Indeed, if $\mathcal{P}_j$ is not ε-uniform and $0 \le j \le t$ then $|C_i^{(j)}| \ge n/2k_j \ge 2^{3k_j + 1}$ and $2^{-k_j} \le \varepsilon^5/8$, so Lemma 28 guarantees a partition $\mathcal{P}_{j+1} = \left(C_i^{(j+1)}\right)_{i=0}^{k_{j+1}}$ with exceptional set $|C_0^{(j+1)}| \le |C_0^{(0)}| + n(2^{-k_0} + 2^{-k_1} + \cdots + 2^{-k_j}) < \varepsilon n$. However, $\mathcal{P}_{t+1}$ cannot exist since if it did exist then we would have

$$\frac{1}{2} > q(\mathcal{P}_{t+1}) = \frac{1}{k_{t+1}^2} \sum_{1 \le i < j \le t} d^2(C_i^{(t+1)}, C_j^{(t+1)}) \ge (t+1)(\varepsilon^5/2) > \frac{1}{2}.$$

This contradiction completes the proof. □

The regularity lemma has numerous reformulations: here we give two, leaving the easy proofs to the reader (see Exercise 60).

Theorem 29′ *For every $\varepsilon > 0$ and $m \in \mathbb{N}$ there is a natural number $M' = M'(\varepsilon, m)$ such that for every graph $G = (V, E)$ there is a partition $V = \bigcup_{i=1}^k C_i$ such that $m \le k \le M'$, $|C_1| \le |C_2| \le \cdots \le |C_k| \le |C_1| + 1$ and, with the exception of at most εk^2 pairs, the pairs (C_i, C_j), $1 \le i < j \le k$, are ε-uniform.* □

Theorem 29″ *For every $\varepsilon > 0$ and $m \in \mathbb{N}$ there is a natural number $M'' = M''(\varepsilon, m)$ such that for every graph $G = (V, E)$ there is a partition $V = \bigcup_{i=0}^k C_i$ such that $m \le k \le M''$, $|C_0| \le k - 1$, $|C_1| = |C_2| = \ldots = |C_k|$, and all but at most ε proportion of the pairs (C_i, C_j), $1 \le i < j \le k$, are ε-uniform.* □

The bound on $M(\varepsilon, m)$ given in the proof of Theorem 29 is enormous: unless m is immense compared to $1/\varepsilon$, it is about $2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$, where the height of the tower is about ε^{-5}. At first sight this seems to be extremely bad and far from the truth. However, in 1997 Gowers proved that $M(\varepsilon, 2)$ grows at least as a tower of 2s of height about $\varepsilon^{-1/16}$: the argument is a tour de force.

In fact, it would be a significant achievement to give reasonable estimates for a much finer function than $M(\varepsilon, m)$ or $M(\varepsilon, 2)$. Given a graph G, call a partition $V(G) = \bigcup_{i=0}^k V_i$ of the vertex set $(\beta, \gamma, \delta, \varepsilon)$-*uniform* if $|V_0| \le \beta|V(G)|$, $|V_1| \le |V_2| \le \cdots \le |V_k| \le |V_1| + 1$, and all but εk^2 of the pairs (V_i, V_j), $1 \le i < j \le k$, are such that if $W_i \subset V_i$, $W_j \subset V_j$, $|W_i| \ge \gamma|V_i|$, and $|W_j| \ge \gamma|V_j|$, then

$$|d(W_i, W_j) - d(V_i, V_j)| < \delta.$$

Let $M(\beta, \gamma, \delta, \varepsilon)$ be the minimal integer such that for every graph G there is a $(\beta, \gamma, \delta, \varepsilon)$-uniform partition $V(G) = \bigcup_{i=0}^k V_i$ with $2 \le k \le M(\beta, \gamma, \varepsilon, \delta)$. Determine the approximate order of $M(\beta, \gamma, \delta, \varepsilon)$ as the four variables tend to 0. This is very likely to be a tall order; as a consolation prize, one could try to determine the order of $M(\beta, \gamma, \delta, \varepsilon)$ as some variables are kept constant and the others tend to 0. For example, given some small values β_0 and ε_0, what can one say about $M(\beta_0, \gamma, \delta, \varepsilon_0)$, as γ and δ tend to 0?

IV.6 Simple Applications of Szemerédi's Lemma

The main use of a Szemerédi-type partition is that it guarantees the existence of certain subgraphs, even in graphs with not too many edges. Here is one of the standard ways of finding all small r-partite subgraphs.

Theorem 30 *Let $f \ge 2$, $r \ge 2$, $0 < \delta < 1/r$ and let $V_1, V_2, \ldots, V_r$ be disjoint subsets of vertices of a graph G. Suppose $|V_i| \ge \delta^{-f}$ for every i, and if $1 \le i < j \le r$ and $W_i \subset V_i$, $W_j \subset V_j$ satisfy $|W_i| \ge \delta^f |V_i|$ and $|W_j| \ge \delta^f |V_j|$, then $d(W_i, W_j) \ge \delta$. Then for all non-negative integers $f_1, \ldots, f_r$ with $\sum_{i=1}^r f_i = f$ there are sets $U_1 \subset V_1, \ldots, U_r \subset V_r$ with $|U_i| = f_i$ for $1 \le i \le r$, such that for $1 \le i < j \le r$ every vertex of U_i is joined to every vertex of U_j. In particular, G contains every r-partite graph on f vertices.*

Proof. Let us apply induction on f. For $f = 2$ the assertion is trivial, so suppose that $f \geq 3$ and the assertion holds for smaller values of f. We may assume that $f_1 \geq 1$.

For $2 \leq i \leq r$, let R_i be the set of vertices in V_1 joined to fewer than $\delta|V_i|$ vertices of V_i. Then $|R_i| < \delta^f|V_i|$, so $\left|\bigcup_{i=2}^r R_i\right| < (r-1)\delta^f|V_1| < |V_1|$. Hence there is a vertex $x \in V_1 \setminus \bigcup_{i=2}^r R_i$; set $V_1' = V_1 \setminus \{x\}$ and $V_i' = V_i \cap \Gamma(x)$ for $i = 2, \ldots, r$. Then $|V_1'| \geq \delta^{-f} - 1 \geq \delta^{-f+1}$ and $|V_1'| \geq (1-\delta^f)|V_1| \geq \delta|V_1|$; furthermore, $|V_i'| \geq \delta|V_i| \geq \delta^{-f+1}$ for $2 \leq i \leq r$. Also, if $W_i \subset V_i'$ and $W_j \subset V_j'$, $1 \leq i < j \leq r$, are such that $|W_i| \geq \delta^{f-1}|V_i'|$ and $|W_j| \geq \delta^{f-1}|V_j'|$, then $|W_i| \geq \delta^f|V_i|$ and $|W_j| \geq \delta^f|V_j|$, so $V_1', \ldots, V_r'$ satisfy the conditions for $0 < \delta < 1/r$ and $f-1$. Hence, by the induction hypothesis, there are sets $U_1' \subset V_1', \ldots, U_r' \subset V_r'$ with $|U_1'| = f_1 - 1$ and $|U_i'| = f_i$ for $2 \leq i \leq r$ such that for $1 \leq i < j \leq r$, every vertex of U_i' is joined to every vertex of U_j'. Clearly, the sets $U_1 = U_1' \cup \{x\}$, $U_2 = U_2'$, $\ldots$, $U_r = U_r'$ have the required properties. □

The proof above is very crude indeed, and even as it stands it shows that the restrictions on δ and V_i are unnecessarily severe and can be relaxed to $0 < \delta < (r-1)^{-1/2}$, $0 < \delta < 1/2$, and $|V_i| \geq \delta^{1-f}$.

More often than not, Theorem 30 is used in conjunction with a Szemerédi-type partition, as in the following immediate consequence of it.

Theorem 31 *Let $f \geq 2$, $r \geq 2$, $0 < \delta < 1/r$, and let $V_1, \ldots, V_r$ be disjoint subsets of vertices of a graph G. Suppose $|V_i| \geq \delta^{-f}$ for every i, and all pairs (V_i, V_j) are δ^f-regular, with density at least $\delta + \delta^f$. Then G contains every r-partite graph of order f.* □

As an application of Theorem 31, let us show that if F is a fixed subgraph with chromatic number $\chi(F) = r \geq 2$ and n is sufficiently large, then every graph of order n not containing F as a subgraph is close to a graph that does not contain K_r.

We set the scene in a little more generality than needed for the immediate application. Let $m \geq 2$, $\varepsilon > 0$, and $\delta > 0$ be given, and let $M = M''(\varepsilon, m)$ be as in Theorem 29″. For a graph G of order $n \geq M$, let $V(G) = \bigcup_{i=0}^k C_i$ be the vertex partition guaranteed by Theorem 29″; thus $m \leq k \leq M$, $|C_0| \leq k-1$, $|C_1| = |C_2| = \cdots = |C_k|$, and all but $\varepsilon\binom{k}{2}$ of the pairs (C_i, C_j), $1 \leq i < j \leq n$, are ε-uniform. Let $G[k;\ \varepsilon;\ d > \delta]$ be the union of the bipartite subgraphs of G spanned by (C_i, C_j) for the ε-regular pairs of density greater than δ. We call $G[k;\ \varepsilon;\ d > \delta]$ an $(m;\ \varepsilon;\ d > \delta)$-*piece* of G. For simplicity, we take the vertex set to be $\bigcup_{i=1}^k C_i$, so that $G[k;\ \varepsilon;\ d > \delta]$ is a k-partite graph with vertex-classes $|C_1| = \cdots = |C_k| = l$, such that $n - k + 1 \leq kl \leq n$. Furthermore, let $S[k;\ \varepsilon;\ d > \delta]$ be the graph on $[k]$ in which ij is an edge if and only if (C_i, C_j) is ε-uniform, with density more than δ. We call $S[k;\ \varepsilon;\ d > \delta]$ the *skeleton* of $G[k;\ \varepsilon;\ d > \delta]$.

Note that $G[k;\ \varepsilon;\ d > \delta]$ is not unique; we just pick one of the possible graphs and for $S[k;\ \varepsilon;\ d > \delta]$ take the skeleton it determines. Furthermore, these graphs need not be defined for every k: all we know is that they are defined for *some* k in the range $m \le k \le M$.

We define $G[k;\ \varepsilon;\ \delta_1 < d < \delta_2]$ and $S[k;\ \varepsilon;\ \delta_1 < d < \delta_2]$ analogously.

Theorem 32 *Let $0 < \varepsilon < 1$ and $0 < \delta < 1$ be real numbers, let $m \ge 2$ be an integer, and let $M = M''(\varepsilon, m)$ be as in Theorem 29″ of the previous section. Let G be a graph of order $n \ge M$, and let $H = G[k;\ \varepsilon;\ d > \delta]$ be an $(m;\ \varepsilon;\ d > \delta)$-piece of G. Then*

$$e(G) - e(H) < \left(\varepsilon + \delta + \frac{1}{m} + \frac{2M}{n}\right) n^2/2\,.$$

In particular, if $0 < \varepsilon \le \delta/2$, $m \ge 4/\delta$ and $n \ge 8M/\delta$ then

$$e(G) - e(H) < \delta n^2\,.$$

Proof. Let $V(G) = \bigcup_{i=0}^{k} C_i$ be the partition guaranteed by Theorem 29″ so that $|C_0| \le k-1$, $|C_1| = \cdots = |C_k|$, $m \le k \le M$, and H is the appropriate k-partite graph with classes $C_1, \ldots, C_k$. Clearly, $E(G) - E(H)$ consists of four types of edges: the edges incident with C_0, the edges joining vertices in the same class C_i, $1 \le i \le k$, the C_i-C_j edges with (C_i, C_j) not ε-uniform, $1 \le i < j \le k$ and the C_i-C_j edges with (C_i, C_j) ε-uniform with density at most δ, $1 \le i < j \le k$. Hence

$$\begin{aligned}
|E(G) - E(H)| &< kn + k\binom{n/k}{2} + \varepsilon\binom{k}{2}\left(\frac{n}{k}\right)^2 + \delta\binom{n}{2} \\
&< kn + \frac{n^2}{2k} + \frac{\varepsilon n^2}{2} + \frac{\delta n^2}{2} \\
&\le Mn + \frac{n^2}{2m} + \frac{\varepsilon n^2}{2} + \frac{\delta n^2}{2} \\
&\le \left(\varepsilon + \delta + \frac{2M}{n} + \frac{1}{m}\right)\frac{n^2}{2},
\end{aligned}$$

as claimed. □

Theorem 33 *For every $\varepsilon > 0$ and graph F, there is a constant $n_0 = n_0(\varepsilon, F)$ with the following property. Let G be a graph of order $n \ge n_0$ that does not contain F as a subgraph. Then G contains a set E' of less than εn^2 edges such that the subgraph $H = G - E'$ has no K_r, where $r = \chi(F)$.*

Proof. We may assume that $r \ge 2$, $0 < \varepsilon < 1/r$, and $f = |F| \ge 3$. Let $\delta = \varepsilon/2$ and $m \ge 8/\varepsilon = 4/\delta$.

Let $M = M''(\delta^f, m)$ be given by Szemerédi's lemma, as in Theorem 29″. We claim that $n_0 = \lceil 8M\delta^{-f}\rceil$ will do.

Indeed, let $H = G[k;\ \delta^f;\ d > \delta + \delta^f]$ be an $(m;\ \delta^f;\ d > \delta + \delta^f)$-piece of G, with skeleton $S = S[k;\ \delta^f;\ d > \delta + \delta^f]$. Then, by Theorem 32,

$$e(G) - e(H) < (\delta + \delta^f)n^2 < \varepsilon n^2.$$

Furthermore, by Theorem 30, S contains no K_r; therefore, neither does H. □

Note that Theorem 33 is a considerable extension of Corollary 23, which is essentially the original form of the Erdős–Stone theorem. Indeed, let $\varepsilon > 0$ and $t \geq 1$ be fixed, and let $n \geq n_0$ $(\varepsilon, K_r(t))$. Then every graph G of order n and size at least $t_{r-1}(n) + \varepsilon n^2$ edges contains a $K_r(t)$, since otherwise, by Theorem 33, by deleting fewer than εn^2 edges of G, we would get a graph H without a K_r. But as $e(H) > t_{r-1}(n)$, Turán's theorem implies that H does contain a K_r. Needless to say, as a proof of the Erdős–Stone theorem, this is far too heavy-handed.

Our final application of Szemerédi's lemma concerns a beautiful 'mixed' case of the quintessential extremal problem, that of determining $\mathrm{ex}(n; F_1, \ldots, F_k)$. We have studied $\mathrm{ex}(n; K_r)$, solved by Turán's theorem, and $\mathrm{ex}(n; K_{s,t})$, the problem of Zarankiewicz. What happens if we forbid *both* K_r and $K_{s,t}$? How large is $\mathrm{ex}(n; K_r, K_{s,t})$? In the case when r and s are fixed and $t = \lfloor cn \rfloor$ for some positive constant c, in 1988 Frankl and Pach gave an upper bound for this function. First we need a result of independent interest.

Theorem 34 *Let H be a k-partite graph with classes $C_1, \cdots, C_k$ where $|C_1| = \ldots = |C_k| = \ell$. Suppose there are q pairs (i, j), $1 \leq i < j \leq k$, with $E(C_i, C_j) \neq \emptyset$. Suppose also that $2 \leq s \leq t$, and G contains no $K_{s,t}$ with all s vertices in the same class C_i. Then*

$$2e(H) \leq (2q)^{1-1/s}\ell^{2-1/s}(t-1)^{1/s}k^{1/s} + 2q\ell s.$$

Proof. Except for the minor variation that not all pairs (C_i, C_j), $1 \leq i < j \leq k$, are joined by edges, we proceed much as in the standard estimate of the Zarankiewicz function $z(s, t)$. Write $d = 2e(G)/k(\ell)$ for the average degree of H. We may assume that $kd > 2q(s-1)$, since otherwise there is nothing to prove. For $x \in V(H)$ and $1 \leq i \leq k$, let $d_i(x)$ be the number of neighbours of x in C_i; also, let $P = \{(x, i) : x \in V(H), 1 \leq i \leq k, d_i(x) \geq 1\}$. Then, trivially,

$$|P| \leq 2q\ell.$$

Let us define a *claw* (or *s-claw*) of H as a star $K_{1,s}$ whose *base*, the set of s vertices in the second class, is contained in some class C_i. (In the usual estimate of the Zarankiewicz function, the base is allowed to be anywhere.) The vertex constituting the first class of a claw is the *centre* of the claw.

Since H contains no $K_{s,t}$, for every s-subset S of C_i there are at most $t-1$ claws with base S. Hence, writing N for the total number of claws in H,

$$N \leq (t-1)\sum_{i=1}^{k}\binom{|C_i|}{s} = (t-1)k\binom{\ell}{s}.$$

On the other hand, for each vertex x and class C_i, there are $\binom{d_i(x)}{s}$ claws with centre x and base in C_i, so

$$N = \sum_{x \in V(H)}\sum_{i=1}^{k}\binom{d_i(x)}{s} = \sum_{(x,i) \in P}\binom{d_i(x)}{s}.$$

Therefore,

$$\sum_{(x,i)\in P} \binom{d_i(x)}{s} \le (t-1)k\binom{\ell}{s}. \tag{22}$$

In order to give a lower bound for the left-hand side of (22), set

$$f_s(u) = \begin{cases} u(u-1)\cdots(u-s+1)/s! & \text{if } u \ge s-1, \\ 0 & \text{if } u \le s-1. \end{cases}$$

Then $f_s : \mathbb{R} \to \mathbb{R}$ is a convex function. As $\sum_{(x,i)\in P} d_i(x) = 2e(H) = k\ell d$, and $|P| \le 2q\ell$, the convexity of f_s implies that

$$2q\ell f_s\left(\frac{kd}{2q}\right) \le \sum_{(x,i)\in P} \binom{d_i(x)}{s}.$$

Since $\frac{kd}{2q} > s-1$, we have $f_s(\frac{kd}{2q}) = \binom{kd/2q}{s}$; recalling inequality (3) we find that

$$2q\ell\binom{kd/2q}{s} \le (t-1)k\binom{\ell}{s}.$$

But then, rather crudely,

$$2q\ell\left(\frac{kd}{2q} - (s-1)\right)^s \le (t-1)k\ell^s,$$

and so

$$2e(G) = k\ell d \le (2q)^{1-1/s}\ell^{2-1/s}(t-1)^{1/s}k^{1/s} + 2q\ell(s-1),$$

as claimed. □

We are ready to present the theorem of Frankl and Pach that we promised.

Theorem 35 *Let $r \ge 3$ and $s \ge 2$ be fixed integers, and c and γ positive constants. Then if n is sufficiently large and G is a graph of order n that contains neither K_r nor $K_{s,t}$, where $t = \lceil cn \rceil$, then*

$$e(G) \le c^{1/s}\left(\frac{r-2}{r-1}\right)^{1-1/s}\frac{n^2}{2} + \gamma n^2.$$

Proof. We may assume that $0 < \gamma < 1/2$ and $c < (r-2)/(r-1)$, since we do know that $e(G) \le t_{r-1}(n) \le \frac{r-2}{2(r-1)}n^2$.

Let $\delta = \gamma/2$, $m \ge 4/\delta$, and suppose that $n \ge 4Ms/\delta \ge 8M/\delta$, where $M = M''(\delta^r;\ m)$. Let G be a graph of order n containing neither K_r nor $K_{s,t}$. Let $H = G[k;\ \delta^r;\ d > \delta+\delta^r]$ be an $(m;\ \delta^r;\ d > \delta+\delta^r)$-piece of G with skeleton S. Then, by Theorem 32, $e(G)-e(H) < (\delta+\delta^r)n^2 < 2\gamma n^2/3$, and by Theorem 31, S does not contain a K_r. Hence, by Turán's theorem, $q = e(S) \le (r-2)k^2/2(r-1)$.

As H contains no $K_{s,t}$, by Theorem 34 we have, with $\ell = \lfloor n/k \rfloor$,

$$
\begin{aligned}
2e(H) &\le \left(\frac{r-2}{r-1}\right)^{1-1/s} (k\ell)^{2-1/s}(t-1)^{1/s} + k^2\ell s \\
&\le c^{1/s}\left(\frac{r-2}{r-1}\right)^{1-1/s} n^2 + Msn \\
&\le c^{1/s}\left(\frac{r-2}{r-1}\right)^{1-1/s} n^2 + \frac{\gamma n^2}{8}.
\end{aligned}
$$

Therefore,

$$e(G) < e(H) + \gamma n^2/2 + < e(H) + \gamma n^2,$$

as claimed. □

In fact, the upper bound in Theorem 35 is essentially best possible: if $r \ge 3$ and $s \ge 2$ are fixed integers and $0 < c \le (r-2)/(r-1)$, then

$$\lim_{n\to\infty} \mathrm{ex}(n; K_r, K_{s,t})\binom{n}{2}^{-1} = c^{1/s}\left(\frac{r-2}{r-1}\right)^{1-1/s},$$

where $t = \lfloor cn \rfloor$.

There are a great number of substantial applications of Szemerédi's regularity lemma. For example, in 1993 Komlós, Sárközy, and Szemerédi proved the following theorem, conjectured by Bollobás in 1978.

Theorem 36 *For every $\varepsilon > 0$ and $\Delta \ge 1$ there is an $n_0 = n_0(\varepsilon, \Delta)$ such that every graph of order n and minimal degree at least $(1+\varepsilon)n/2$ contains every tree of order n and maximal degree at most Δ.*

In fact, more is true: given $\varepsilon > 0$, if $c > 0$ is small enough, and n is large enough then every graph of order n and minimal degree at least $(1+\varepsilon)n/2$ contains every tree of order n and maximal degree at most cn. There are numerous related conjectures, the best known of which is the conjecture of Erdős and Sós from 1963: every graph of order n and size $\lfloor (k-1)n/2 \rfloor + 1$ contains every tree with k edges.

IV.7 Exercises

1.$^-$ Show that every graph with n vertices and minimal degree at least $\lfloor n/2 \rfloor$ is connected, but for every $n \ge 2$ there are disconnected graphs with minimal degree $\lfloor n/2 \rfloor - 1$.

2.$^-$ Let G be a graph of order $n \ge k+1 \ge 2$ and size at least $\binom{n}{2} - n + k$. Show that G is k-connected unless it has a vertex x of degree $k-1$ such that $G - x \cong K_{n-1}$.

3. Let $0 \le k \le n$. Show that an n by n bipartite graph without $k+1$ independent edges has size at most kn. Determine the unique extremal graph.

4. (i) Let G be a graph of order n, maximal degree $\Delta \ge 3$, and diameter d. Let $n_0(g, \delta)$ be as in Theorem 1. Prove that $n \le n_0(2d+1, \Delta)$, with equality iff G is Δ-regular and has girth $2d+1$.
(ii) Let G be a graph of order n, maximal degree $\Delta \ge 3$, and suppose every vertex is within distance $d-1$ of each pair of adjacent vertices. Prove that $n \le n_0(2d, \Delta)$, with equality iff G is Δ-regular and has girth $2d$.

5. Prove Theorem 4 for $k = 3$ and 4.

6. Show that a graph with n vertices and $m > 3(n-1)/2$ edges contains two vertices joined by three independent paths.

7. Prove that the maximal number of edges in a graph of order n without an even cycle is $\lfloor \frac{3}{2}(n-1) \rfloor$. Compare this with the maximal size of a graph without an odd cycle.

8. Show that a tree with $2k$ endvertices contains k edge-disjoint paths joining distinct endvertices.

9. Suppose x is not a cutvertex and has degree $2k$. Prove that there are k edge-disjoint cycles containing x. [Cf. Exercise 8.]

10. Show that if $\kappa(G) \ge 3$, then $G \supset TK_4$. Show that the same holds if $\delta(G) \ge 3$.

11. Deduce from the assertion in Exercise 10 that if $e(G) \ge 2|G| - 2$ then G contains a subdivision of K_4.

12. Recall that a graph of order n and size $n+1$ has girth at most $\lfloor \frac{2}{3}(n+1) \rfloor$. Show that a graph of order n and size $n+2$ has girth at most $\lfloor (n+2)/2 \rfloor$. Show also that both bounds are best possible. [*Hint.* Assuming that $\delta(G) \ge 3$, study the multigraph H with $\delta(H) \ge 3$ whose subdivision G is.]

13. Prove that for $k \ge 1$ the maximal girth of a graph of order $n = 9k - 3$ and size $9k$ is $4k$. What is the maximal girth of a graph of order n and size $n+3$?

14. Show that for every $k \ge 1$ there is a graph of order $16k - 4$, size $16k$ and girth $6k$. [*Hint.* Consider an octagon with the opposite vertices joined.]

15. Let $r \ge 1$. We say that the cycles $C_1, \ldots, C_r$ are *nested* if $V(G_1) \subset \cdots \subset V(G_r)$. Determine

$$\min\{n : K_n \text{ contains } r \text{ nested cycles}\}.$$

16. The *domination number* of a graph G is

$$\min\{|W| : W \subset V(G),\ W \cup \Gamma(W) = V(G)\}.$$

Show that if G has n vertices, then its domination number is at least $\lceil \sqrt{4n} \rceil - 1 - \Delta(G)$, and this inequality is best possible for every $n \ge 1$.

17. Show that the domination number of a graph of order n and minimal degree 2 is at most $\lfloor n/2 \rfloor$. Note that equality can be attained for every n. [*Hint.* Assuming, as we may, that every edge of our graph is incident with a vertex of degree 2, let $U = \{u \in V(G) : d(u) \geq 3\}$, and consider the partition $V(G) = U \cup W \cup Z$, where $W = \Gamma(U)$.]

18.$^-$ Show that a graph with n vertices and minimal degree $\lfloor (r-2)n/(r-1) \rfloor + 1$ contains a K_r.

19. Let G have $n \geq r+1$ vertices and $t_{r-1}(n) + 1$ edges.
(i) Prove that for every p, $r \leq p \leq n$, G has a subgraph with p vertices and at least $t_{r-1}(p) + 1$ edges.
(ii) Show that G contains two K_r subgraphs with $r-1$ vertices in common.

20.$^+$ Prove that for $n \geq 5$ every graph of order n with $\lfloor n^2/4 \rfloor + 2$ edges contains two triangles with exactly one vertex in common.

21.$^+$ Prove that if a graph with n vertices and $\lfloor n^2/4 \rfloor - \ell$ edges contains a triangle, then it contains at least $\lfloor n/2 \rfloor - \ell - 1$ triangles. [*Hint.* Let $x_1x_2x_3$ be a triangle and denote by m the number of edges joining $\{x_1, x_2, x_3\}$ to $V(G) - \{x_1, x_2, x_3\}$. Estimate the number of triangles in $G - \{x_1, x_2, x_3\}$ and the number of triangles sharing a side with $x_1x_2x_3$.]

22.$^+$ (i) Show that the edges of a graph of order n can be covered with not more than $\lfloor n^2/4 \rfloor$ edges and triangles.
(ii) Let G be a graph with vertices $x_1, \ldots, x_n$, $n \geq 4$. Prove that there is a set S, $|S| \leq \lfloor n^2/4 \rfloor$, containing non-empty subsets $X_1, X_2, \ldots, X_n$ such that x_ix_j is an edge of G if $X_i \cap X_j \neq \emptyset$.

23.$^-$ Let $1 \leq k \leq n$. Show that every graph of order n and size $(k-1)n - \binom{k}{2} + 1$ contains a subgraph with minimal degree k, but there is a graph of order n and size $(k-1)n - \binom{k}{2}$ in which every subgraph has minimal degree at most $k-1$. [*Hint.* Imitate the proof of Lemma 20.]

24.$^-$ Show that a graph of order n and size $(k-1)n - \binom{k}{2} + 1$ contains every tree of order $k+1$.

25. Let G be a graph of order n that does not contain a cycle with at least one of its diagonals. Prove that if $n \geq 4$, then G has at most $2n - 4$ edges.
Show that if $n \geq 6$ and G has $2n - 4$, edges then G is the complete bipartite graph $K(2, n-2)$. [*Hint.* Consider a longest path in G.]

26.$^-$ Let $k \geq 1$ and let G be a graph of order n without an odd cycle of length less than $2k + 1 \leq 5$. Prove that $\delta(G) \leq \lfloor n/2 \rfloor$ and $T_2(n)$ is the only extremal graph, unless $n = 2k + 1 = 5$, in which case there is another extremal graph, C_5.

27.$^+$ Let G be a graph of order n without an odd cycle of length less than $2k+1 \geq 5$. Prove that if G does not contain $\lceil n/2 \rceil$ independent vertices then $\delta(G) \leq 2n/(2k+1)$. Show that equality holds only for $n = (2k+1)/t$ and the extremal graphs are obtained from a cycle C_{2k+1} by replacing each vertex by t vertices, as in Fig. IV.5.

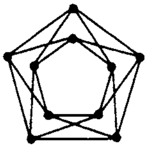


FIGURE IV.5. The graph $C_5(2)$.

28. Let $x_1, x_2, \dots, x_n$ be vectors of norm at least 1 in a Euclidean space. Prove that there are at most $\lfloor n^2/4 \rfloor$ unordered pairs i, j such that $|x_i + x_j| < 1$. [*Hint.* Show that if $|x_1| = |x_2| = |x_3| = 1$ then $|x_i + x_j| \geq 1$ for some i, j, $1 \leq i < j \leq 3$.]

29.$^+$ Let X and Y be independent identically distributed random variables taking values in a Euclidean space. Prove that $\mathbb{P}(|X+Y| \geq x) \geq \frac{1}{2}\mathbb{P}(|X| \geq x)^2$ for every $x \geq 0$.

30. Let $x_1, x_2, \dots, x_{3p} \in \mathbb{R}^2$ be such that $|x_i - x_j| \leq 1$. Prove that at most $3p^2$ of the distances $|x_i - x_j|$ are greater than $\sqrt{2}/2$. [*Hint.* Show that among any four of the points there are two within the distance $\sqrt{2}/2$ of each other.]

31. Recall that a maximal complete subgraph of a graph is a *clique* of the graph, and the *clique number* $\omega(G)$ of a graph G is the maximal order of a clique of G. Thus $\{x, y\}$ is the vertex set of a clique of G if $xy \in E(G)$ and no vertex of G is joined to both x and y. Show that, for every $n \geq 1$, there is a graph of order n with $\lceil n/2 \rceil$ cliques of different orders.

32. Show also that if G is a regular graph of order n then either $\omega(G) = n$ or else $\omega(G) \leq n/2$. Show also that if $n \geq 1$ and $1 \leq p \leq n/2$ then there is a regular graph G of order n with $\omega(G) = p$.

33.$^+$ We say that a set $W \subset V(G)$ *covers* the edges of a graph G if every edge of G is incident with at least one vertex in W. Denote by $\alpha_0(G)$ the minimal number of vertices covering the edges of G. Prove that if G has n vertices and m edges, then $\alpha_0(G) \leq 2mn/(2m+n)$, with equality iff $G = pK_r$ for some p and r, that is, iff each component of G is K_r for some r. [*Hint.* Note that $\delta_0(G) = n - \omega(\overline{G})$, and if $\omega(\overline{G}) = p$, then by Turán's theorem $e(\overline{G}) \leq t_p(n)$, so $m = \binom{n}{2} - t_p(n)$.]

34. The edge *clique-cover number* $\theta_e(G)$ of a graph G is the minimal number of cliques of G whose union is G. Call two vertices x, y *equivalent* if $xy \in E(G)$ and every $z \in V(G) \setminus \{x, y\}$ is joined to x iff it is joined to y. Check that if x and y are equivalent vertices then $\theta_+e(G) = \theta_e(G')$, where $G' = G \backslash \{y\}$. Prove that if G contains neither isolated vertices, nor equivalent vertices, then $\theta_e(G) \geq \log_2(n+1)$, where n is the order of G. [*Hint.* Let $K_1, \dots, K_m$

be cliques of G with $G = \bigcup_{i=1}^{m} K_i$. For $x \in V(G)$, let $I(x) = \{i : K_i$ contains an edge incident with $x\}$. Check that if $x, y \in V(G)$ and $x \neq y$, then $I(x) \neq I(y)$.]

35.$^+$ (Cf. Corollary 16.) Let $d_1 \leq d_2 \leq \cdots \leq d_n$ be a graphic sequence such that for some k,

$$d_k \leq k < \frac{n}{2} \quad \text{and} \quad d_{n-k} \leq n-k-1.$$

Show that there is a non-Hamiltonian graph G with vertex set $\{x_1, x_2, \ldots, x_n\}$ such that $d(x_i) \geq d_i$, $1 \leq i \leq n$ (cf. Fig. IV.6).

FIGURE IV.6. The graph $(K_2 \cup E_3) + K_3$ has no Hamilton cycle and $(K_2 \cup E_3) + K_2$ has no Hamilton path.

36.$^+$ (Cf. Corollary 16.) Let $d_1 \leq d_2 \leq \cdots \leq d_n$ be a graphic sequence such that for some k,

$$d_k \leq k-1 < \frac{1}{2}(n-1) \quad \text{and} \quad d_{n+1-k} \leq n-k.$$

Prove that there is a graph G with vertex set $\{x_1, x_2, \ldots, x_n\}$ such that $d(x_i) \geq d_i$, $1 \leq i \leq n$, and G does not contain a Hamilton path (cf. Fig. IV.6).

37. Prove that a non-Hamiltonian graph of order $n \geq 3$ has at most $\binom{n}{2} - (n-2)$ edges and there is a unique extremal graph.
Prove that a graph of order $n \geq 2$ without a Hamilton path has at most $\binom{n}{2} - (n-3)$ edges and $K_{n-1} \cup K_1$ is the unique extremal graph.

38. Given $\delta < n/2$, determine the maximal number of edges in a graph G of order n without a Hamilton cycle (path), provided that $\delta(G) = \delta$.

39.$^+$ Prove Theorem 4 by making use of simple transforms of a longest x_0-path $P = x_0x_1 \cdots x_t$. [*Hint.* Apply induction on n. If $\delta(G) \leq k/2$, the result follows by induction; otherwise, consider the set L of endvertices of simple transforms of P. Put $\ell = |L|$, $r = \max_{x \in L} d(x)$, and note that $\ell \geq r$ and the neighbours of each $x \in L$ are contained in $\{x_t, x_{t-1}, \ldots, x_{t-k+1}\}$. Deduce that $e(G) - e(G-L) \leq \ell(k-\ell) + \ell(r+\ell-k) \leq k\ell/2$ and complete the proof by applying the induction hypothesis to $G - L$.]

40. Let $1 < a_1 < a_2 < \cdots < a_k \leq x$ be natural numbers. Suppose no a_i divides the product of any two others. Prove that $k \leq \pi(x) + x^{2/3}$, where, as usual, $\pi(x)$ denotes the number of primes not exceeding x. [*Hint.* Put $V_1 = \{1, 2, \ldots, \lfloor x^{2/3} \rfloor\}$ and $V_2 = \{x : x^{2/3} \leq b \leq x$ and b is a prime$\}$. Show first that $a_i = b_ic_i$, where $b_i, c_i \in V = V_1 \cup V_2$. Let G be the graph (with

loops) with vertex set V whose edges (loops) are $b_i c_i$. Note that G does not contain a path of length 3.]

41.$^{++}$ Let $1 < a_1 < a_2 < \cdots < a_k \leq x$ be natural numbers. Suppose $a_i a_j \neq a_h a_\ell$ unless $\{i, j\} = \{h, \ell\}$. Prove that $k \leq \pi(x) + cx^{3/4}$ for some constant $c > 0$. [*Hint.* The graph G in the previous exercise contains no quadrilaterals; apply Theorem 8 to the bipartite subgraph of G with vertex classes V_1 and V_2. Recall the prime number theorem, namely that $(\pi(x) \log x)/x \to 1$ as $x \to \infty$.]

42.$^{+}$ Denote by $D_k(n)$ the maximal number of occurrences of the same positive distance among n points in $\mathbb{R}^k$. Prove that if $k \geq 2$ then

$$\lim_{n\to\infty} D_k(n)/n^2 = \frac{1}{2} - \frac{1}{2\lfloor k/2 \rfloor}$$

[*Hint.* (i) Note that if $\mathbf{x} \in \{\mathbf{z} \in \mathbb{R}^k;\ z_1^1 + z_2^2 = 1$ and $z_i = 0$ if $i > 2\}$ and $\mathbf{y} \in \{\mathbf{z} \in \mathbb{R}^k;\ z_3^2 + z_4^2 = 1$ and $z_i = 0$ if $i \neq 4$ or $4\}$, then $|\mathbf{x} - \mathbf{y}| = \sqrt{2}$.
(ii) Deduce from Theorem 20 that $D_k(n)$ is at least as large as claimed.]

43. Show that a graph of order $n \geq k(d+1)$ with at least $k \geq 2$ components and minimal degree at least d has at most

$$(k-1)\binom{d+1}{2} + \binom{n-(k-1)(d+1)}{2}$$

edges. What is the unique extremal graph?

44. By checking the details of the 'duplication' argument, show that $\mathrm{ex}(n, K_{s,t}) \leq \frac{1}{2} z(n, n; s, t)$.

45.$^{+}$ Show that if any $k+1$ vertices of a k-connected graph with at least 3 vertices span at least one edge, then the graph is Hamiltonian.

46. Let k and n be natural numbers. Show that every graph of order n and size greater than $k(n - (k+1)/2)$ contains a subgraph of minimal degree $k+1$. Show also that for every $m \leq k(n - (k+1)/2)$ there is a graph of order n and size m that has no subgraph of minimal degree at least $k+1$.

47. Let $X \subset \mathbb{R}^2$ with $|X| = n \geq 3$ and $\max\{d(x, y) : (x, y) \in X^{(2)}\} = 1$. Show that there are at most n pairs $(x, y) \in X^{(2)}$ with $d(x, y) = 1$, and this bound can be attained for every $n \geq 3$.
[*Hint.* Apply induction on n. For the proof of the induction hypothesis, set $E = \{(x, y) \in X^{(2)} : d(x, y) = 1\}$ and let G be the graph (X, E). Assuming that $|E| \geq n+1$, show that there is a subgraph $H \subset G$ with $\delta(H) \geq 2$ and $\Delta(H) \geq 3$, and make use of a vertex of degree at least 3 in H to arrive at a contradiction.]

48. Let $X = \{x_1, \ldots, x_n\}$ be a set of n points in the plane, with no three collinear, and let $G = (X, E)$ be a graph with $n+1$ edges. Show that there are edges $x_1 y_1, x_2 y_2 \in E$ such that the straight line segments $[x_1, y_2]$ and $[x_2, y_2]$ are

disjoint. Show also that the bound $n+1$ is best possible for every $n \geq 3$. [*Hint.* Imitate the proof of Exercise 47.]

49.$^+$ Show that an r-regular graph of order $2r+1$ is Hamiltonian. Show also that if $r \geq 2$, then our graph contains a triangle.

50.$^+$ (i) Prove that the maximal number of edges of a non-Hamiltonian graph of order $2n$ and minimal degree $n-1$ is $3n(n-1)/2$.
(ii) Determine $\mathrm{ex}(n; C_n)$.

51. Determine $\mathrm{ex}(n; P_k)$ for every n and k, where P_k is a path of length k.

52. Note that if G is a graph of order n, then $n - \alpha(G)$ is the minimal number of vertices representing all edges of G; i.e., $n - \alpha(G) = \min\{|R| : R \subset V(G), G - R \text{ has no edges}\}$. Here $\alpha(G)$ is the independence number, the maximal number of independent vertices, so that $\alpha(G) = \omega(\overline{G})$. Show that if G has no triangles then $e(G) \leq \alpha(G)(n - \alpha(G)) \leq n^2/4$.

53. Recall that the maximal number of edges in a graph of order n containing only even cycles is precisely $\lfloor n^2/4 \rfloor$. What is the maximum if every cycle-length is a multiple of 3? And if every cycle-length is a multiple of 4?

54.$^+$ Describe all 2-connected graphs that do not contain an odd cycle of length at least five.

55. Let G be a triangle-free graph of order n. Show that $\sum_{x \in V(G)} d(x)^2 \leq n^3/4$, with equality if and only if n is even and G is $T_2(n)$. [*Hint.* Recall the proof of Mantel's theorem from Chapter I.]

56.$^-$ For each $r \geq 3$, construct a graph of order $r+2$ that contains no K_r but is not $(r-1)$-partite.

57.$^-$ Let G be a graph of order n such that no set of $n-k$ vertices is independent (i.e., every set of $n-k$ vertices spans at least one edge) and no set of $k+1$ edges is independent (i.e., among any $k+1$ edges, there are two that share a vertex). Show that $e(G) \geq k+2$.

58. Show that for $n \geq 5$, the maximal number of edges of a triangle-free non-bipartite graph of order n is $\lfloor (n-1)^2/4 \rfloor + 1$. [*Hint.* Delete the vertex set of a shortest odd cycle.]

59. Let G be a triangle-free graph of order n and size $\lfloor n^2/4 \rfloor - m$. Show that G contains an induced bipartite subgraph of order at least $n - 8m/n$ (i.e., there is a set $W \subset V(G)$ such that $|W| \geq n - 8m/n$ and $G[W]$ is bipartite).

60. Given $r \geq 3$, determine the minimal order of a graph that is not $(r-1)$-partite and contains no K_r.

61. Let G be a graph of average degree $d > 0$, and let $r = \lceil d/4 \rceil$. Show that for some $k \geq r$, G contains a k by k bipartite graph with a 1-factor, in which every vertex in the first class has degree r.

62. Let $0 < c < c + \varepsilon \le 1$ and $\eta > (\varepsilon/(1-c))^{1/2}$. Show that if n is sufficiently large, then there is a graph of order n and size at least $(c+\varepsilon)\binom{n}{2}$ such that every subgraph H with $|H| \ge \eta n$ has minimal degree less than $c|H|$.

63.$^+$ Check the estimates in the proofs of Theorems 20 and 22 to show that in Theorem 22 we may take $n_0(r, \varepsilon) = \max\{\lceil 3/\varepsilon \rceil, 100\}$.

64.$^+$ Show that for all $\varepsilon_1 > \varepsilon_0 > 0$ there is an $\eta > 0$ such that if $X_0 \subset X_1$, $Y_0 \subset Y_1$, $X_1 \cap Y_1 = \emptyset$, $|X_1| \le (1+\eta)|X_0|$, $|Y_1| \le (1+\eta)|Y_0|$, and (X_0, Y_0) is ε_0-regular, then (X_1, Y_1) is ε_1-regular.

65. Deduce Theorems 29$'$ and 29$''$ from Theorem 29.

66.$^+$ Let $P_1, P_2, \dots, P_n$ be points in the unit square. Show that there are at least $(\lceil n/2 \rceil^2 + \lfloor n/2 \rfloor^2 - n)/2$ pairs (i, j), $1 \le i < j \le n$, with the distance $d(P_i, P_j)$ being at most 1.

67.$^+$ Let $P_1, P_2, \dots, P_n$ be points in the unit cube. Show that at least $n(n-7)/14$ pairs (P_i, P_j), $1 \le i < j \le n$, are at distance at most 1 from each other.

68.$^+$ Let $G = G(n, m)$ be triangle-free. Show that for some vertex $x \in G$ we have

$$e(G[W_x]) \le m - 4m^2/n^2,$$

where $W_x = \{y \in G\colon d(x, y) \ge 2\}$.
Show also that if n is even and $m = rn/2$ for some integer r then this inequality is best possible: for some graph $G = G(n, m)$ equality holds for every $x \in G$. [*Hint.* Imitate the proof of Theorem 2 in Chapter I.]

69. Deduce from Exercise 68 that if G is a triangle-free graph then $e(G[W_x]) \le n^2/16$ for some vertex x. Show also that if n is a multiple of 4 then this inequality is best possible.

70. Let G be a graph of size $\binom{k}{2} + 1$. Show that either k is even and $\Delta(G) = 1$, or else G has a subgraph of order $k+1$ without isolated vertices.

71. For $n \ge 1$, let $m(n)$ be the maximal integer m such that every graph of order $2n+1$ and size at most m is the union of a bipartite graph of maximal degree less than n. Check that $m(1) = 2$ and $m(2) = 7$, and prove that for $n \ge 3$ we have

$$m(n) = \binom{2n+1}{2} - \binom{n}{2} - 1.$$

[*Hint.* Make use of the result in Exercise 70.]

IV.8 Notes

There is an immense literature on extremal problems: here we shall give only the basic references.

The results concerning Hamilton cycles and paths presented in the chapter all originate in a paper of G.A. Dirac: Some theorems on abstract graphs, *Proc. London Math. Soc.* **2** (1952) 69–81. Theorem 2, L. Pósa's extension of Dirac's theorem, is from A theorem concerning Hamiltonian lines, *Publ. Math. Inst. Hungar. Acad. Sci.* **7** (1962) 225–226. The notion of the k-closure discussed in Section 3 was introduced in J.A. Bondy and V. Chvátal, A method in graph theory, *Discrete Math.* **15** (1976) 111–135; Corollary 14, characterizing minimal forcibly Hamiltonian degree sequences, is in V. Chvátal, On Hamilton's ideals, *J. Combinatorial Theory* Ser. B **12** (1972) 163–168. Theorem 17 is from L. Pósa, Hamilton circuits in random graphs, *Discrete Mathematics* **14** (1976) 359–364.

The special case of Theorem 19 concerning cubic (that is, 3-regular) graphs was proved by C.A.B. Smith, and was first mentioned in W.T. Tutte, On Hamiltonian circuits, *J. London Math. Soc.* **21** (1946) 98–101; the theorem itself is due to A.G. Thomason, Hamiltonian cycles and uniquely edge colourable graphs, *Annals of Discrete Math.* **3** (1978) 259–268.

The fundamental theorem of Turán concerning complete subgraphs is in P. Turán, On an extremal problem in graph theory (in Hungarian), *Mat. és Fiz. Lapok* **48** (1941) 436–452; its extension, Theorem 7, giving more information about the degree sequence of a graph without a K_r, is in P. Erdős, On the graph theorem of Turán (in Hungarian), *Mat. Lapok* **21** (1970) 249–251, and its algorithmic extension, Theorem 6, is from a forthcoming paper: B. Bollobás, Turán's theorem and maximal degrees, *J. Combinatorial Theory* Ser. B. The Erdős–Stone theorem is in P. Erdős and A.H. Stone, On the structure of linear graphs, *Bull. American Math. Soc* **52** (1946) 1087–1091; its extension, Theorem 22, is in B. Bollobás and P. Erdős, On the structure of edge graphs, *Bull. London Math. Soc.* **5** (1973) 317–321. The correct dependence of $K_{r+1}(t)$ on r and ε was determined by V. Chvátal and E. Szemerédi, Notes on the Erdős–Stone theorem, in *Combinatorial Mathematics* (Marseilles-Luminy, 1981), North-Holland Math. Studies, vol. **75**, North-Holland, Amsterdam, 1983, pp. 183–190.

Concerning the problem of Zarankiewicz, Theorem 12 is from I. Reiman, Über ein Problem von K. Zarankiewicz, *Acta Math. Acad. Sci. Hungar.* **9** (1958) 269–279, Brown's construction is from W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* **9** (1966) 281–285 and inequality (7), greatly improving Theorem 10, is from Z. Füredi, An upper bound on Zarankiewicz's problem, *Combinatorics, Probability and Computing* **5** (1996) 29–33. The elegant result in 34 is from A. Gyárfás, A simple lower bound on edge coverings by cliques, *Discrete Math.* **85** (1990) 103–104.

There are numerous papers of Erdős and Simonovits about forbidden subgraphs and the structure of graphs with many edges, including P. Erdős and M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hungar.* **1** (1966) 51–57, which can be found in Chapter VI of B. Bollobás, *Extremal Graph Theory*, Academic Press, London-New York-San Francisco, 1978; Chapter III of the same book concerns cycles and contains many more results than we could present in this chapter.

The two-volume treatise *Handbook of Combinatorics* (R.L. Graham, M. Grötschel, and L. Lovász, eds), North-Holland, Amsterdam, 1995, contains several review articles with numerous additional results, see J.A. Bondy, Basic graph theory: paths and circuits, in vol. I, pp. 3–110, and B. Bollobás, Extremal graph theory, in vol. II, pp. 1231–1292.

Szemerédi's regularity lemma is from E. Szemerédi, Regular partitions of graphs, in *Problèmes Combinatoires et Théorie des Graphes*, Colloq. Intern. C.N.R.S. **260**, Orsay, 1976, pp. 399–401. Theorem 35 and 36 are from P. Frankl and J. Pach, An extremal problem of K_r-free graphs, J. *Graph Theory* **12** (1988) 519–523, and J. Komlós, G.N. Sárközy, and E. Szemerédi, Proof of a packing conjecture of Bollobás, AMS Conference on Discrete Mathematics, DeKalb, Illinois, 1993, and *Combinatorics, Probability and Computing* **4** (1995) 241–255. For numerous exciting applications of the regularity lemma, see J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, in *Combinatorics, Paul Erdős is Eighty*, vol. 2 (D. Miklós, V.T. Sós, and T. Szőnyi, eds), *Bolyai Math. Soc.*, Budapest, 1996, pp. 295–352.

V

Colouring

We wish to arrange the talks in a congress in such a way that no participant will be forced to miss a talk they would like to hear: there are no undesirable clashes. Assuming a good supply of lecture rooms enabling us to hold as many parallel talks as we like, how long will the programme have to last? What is the smallest number k of time slots required? Let us reformulate this question in terms of graphs. Let G be the graph whose vertices are the talks and in which two talks are joined iff there is a participant wishing to attend both. What is the minimal value of k for which $V(G)$ can be partitioned into k classes, say $V_1, V_2, \ldots, V_k$, such that no edge joins two vertices of the same class? As in Section IV.4, we denote this minimum by $\chi(G)$ and call it the (*vertex*) *chromatic number* of G. The terminology originates in the usual definition of $\chi(G)$: a *proper colouring* or simply a *colouring* of the vertices of G is an assignment of colours to the vertices in such a way that adjacent vertices have distinct colours; $\chi(G)$ is then the minimal number of colours in a (vertex) colouring of G. Thus, for example, $\chi(K_k) = k$, $\chi(\overline{K_k}) = 1$, $\chi(C_{2k}) = 2$ and $\chi(C_{2k+1}) = 3$.

In general, it is difficult to determine the chromatic number of a graph. However, it is trivial that if $K_k \subset G$ then $\chi(G) \geq \chi(K_k) = k$. Putting this slightly differently,

$$\chi(G) \geq \omega(G), \tag{1}$$

where $\omega(G)$ is the *clique number* of G, the maximal order of a complete subgraph of G.

Let us remark here that we shall use real colours (red, blue, ...) only if there are few colours, otherwise the natural numbers will be our "colours". Thus a k-colouring of the vertices of G is a function $c : V(G) \to \{1, 2, \ldots, k\}$ such that each set $c^{-1}(j)$ is independent. The sets $c^{-1}(j)$ are the *colour classes* of the colouring.

Another scheduling problem goes as follows. Each of n businessmen wishes to hold confidential meetings with some of the others. Assuming that each meeting lasts a day and at each meeting exactly two businessmen are present, in how many days can the meetings be over? In this case one considers the graph H whose vertices correspond to the n businessmen and where two vertices are adjacent iff the two businessmen wish to hold a meeting. Then the problem above asks for the minimal number of colours in an edge-colouring of H, that is, in a colouring of the *edges* of H in such a way that no two adjacent edges have the same colour. This number, denoted by $\chi'(H)$, is the *edge-chromatic number* or *chromatic index* of H. Note that $\chi'(H)$ is exactly the chromatic number of the line graph of H:

$$\chi'(H) = \chi(L(H)). \tag{2}$$

In the first two sections of the chapter we shall present the basic results concerning colourings of vertices and edges. The chromatic numbers of graphs drawn on surfaces, especially on the plane, merit separate study. We shall devote much of Section 3 to planar graphs; we shall also discuss graphs on other surfaces, and we shall give a brief outline of the proof of the most famous result in graph theory, the four colour theorem.

If, instead of colouring every vertex with a colour from the same set $[k] = \{1, 2, \dots, k\}$, we demand that the colour of a vertex x be chosen from a special set or list $L(x)$ assigned to x, then we arrive at the concept of *list colouring*. How long do the lists have to be to guarantee that there is a proper colouring with this restriction? In terms of our example of talks in a congress, each speaker is available to talk only on a set of days $L(x)$: for how many days must each speaker be available to ensure that we can devise an appropriate programme? Some of the many beautiful results concerning list colourings will be presented in Section 4. As we shall see, list colourings are connected to the stable matchings we studied in Section III.5.

In the final section we shall prove the basic results concerning perfect graphs. A graph is *perfect* if for every induced subgraph of it we have equality in (1). These graphs have a surprisingly beautiful structure, and are important not only for their own sake but also because of their connections to optimization, linear programming and polyhedral combinatorics.

V.1 Vertex Colouring

In Section I.2 we noted the simple fact that a graph is bipartite iff it does not contain an odd cycle. Thus $\chi(G) \geq 2$ iff G contains an edge and $\chi(G) \geq 3$ iff G contains an odd cycle. For $k \geq 4$ we do not have a similar characterization of graphs with chromatic number at least k, though there are some complicated characterizations (cf. Exercises 30–34). Rather than asking for a characterization, let us lower our aim considerably, and ask for the most obvious reasons for a graph to have a large chromatic number. We have already noted one such reason, namely the existence of a large complete graph: this gave us inequality (1). After a moment's thought, another simple reason springs to mind: the absence of a large independent set. Indeed, if G does not contain $h+1$ independent vertices, then in every colouring of G at most h vertices get the same colour (every colour class has at most h vertices). Hence

$$\chi(G) \geq \max\{\omega(G), |G|/\alpha(G)\}, \tag{3}$$

where $\alpha(G)$, the *independence number* of G, is the maximal size of an independent set.

Although for many a graph G inequality (3) is very weak, it is a definite improvement on (1). Nevertheless, it is not too easy to see that $\omega(G)$ can be much smaller than $\chi(G)$. In fact, it is also not easy to see that we can have $\omega(G) = 2$ and $\chi(G)$ large, that is, that there are triangle-free graphs of large chromatic number (cf. Exercise 12). In Chapter VII we shall make use of random graphs and inequality (3) to show that there exist graphs with arbitrarily large chromatic number and arbitrarily large girth. The difficulty we encounter in finding such graphs shows that it would be unreasonable to expect a simple characterization of graphs with large chromatic number. Thus we shall concentrate on finding ways of colouring a graph with few colours.

How would one try to colour the vertices of a graph with colours $1, 2, \ldots,$ using as few colours as possible? A simple approach is as follows. Order the vertices, say $x_1, x_2, \ldots, x_n$, and then colour them one by one: give x_1 colour 1, then give x_2 colour 1 if $x_1x_2 \notin E(G)$ and colour 2 otherwise, and so on; colour each vertex with the smallest colour it can have at that stage. This so-called *greedy algorithm* does produce a colouring, but the colouring may (and usually does) use many more colours than necessary. Fig. V.1 shows a bipartite (i.e., 2-colourable) graph for which the greedy algorithm wastes two colours. However, it is easily seen (Exercise 3) that for every graph the vertices can be ordered in such a way that the greedy algorithm uses as few colours as possible. Therefore it is not surprising that it pays to investigate the number of colours needed by the greedy algorithm in various orders of the vertices.

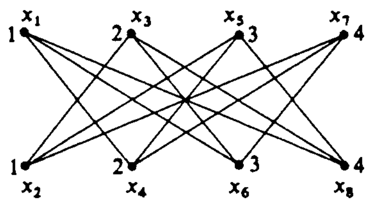


FIGURE V.1. In the order $x_1, x_2, \ldots, x_8$ the greedy algorithm needs four colours.

First, note that whatever order we take, the greedy algorithm uses at most $\Delta(G) + 1$ colours for colouring the vertices of a graph G. Indeed, when we come to colouring a vertex x of degree $d(x)$, at least one of the first $d(x) + 1$ colours has not been used for a neighbour of x, so at least one of these colours is available for x. This simple observation shows that what matters is not even the maximal degree but the maximal number of neighbours of a vertex we have coloured *before* we get to the vertex itself. From here it is but a short step to the following result.

Theorem 1 *Let $k = \max_H \delta(H)$, where the maximum is taken over all induced subgraphs of G. Then $\chi(G) \le k + 1$.*

Proof. The graph G itself has a vertex of degree at most k; let x_n be such a vertex, and put $H_{n-1} = G - \{x_n\}$. By assumption, H_{n-1} has a vertex of degree at most k. Let x_{n-1} be one of them and put $H_{n-2} = H_{n-1} - \{x_{n-1}\} = G - \{x_n, x_{n-1}\}$. Continuing in this way we enumerate all the vertices.

Now, the sequence $x_1, x_2, \ldots, x_n$ is such that each x_j is joined to at most k vertices preceding it. Hence the greedy algorithm will never need colour $k + 2$ to colour a vertex. □

In a somewhat more down-to-earth formulation, Theorem 1 says that a minimal $(k + 1)$-chromatic graph has minimal degree at least k: if $\chi(G) = k + 1$ and $\chi(H) \le k$ for every proper (induced) subgraph H of G then $\delta(G) \ge k$.

It is, of course, very easy to improve the efficiency of the greedy algorithm. If we already have a subgraph H_0 that we know how to colour with $\chi(H_0)$ colours, then we may start our sequence with the vertices of H_0, colour H_0 in an efficient way, and apply only then the algorithm to colour the remaining vertices. This gives us the following extension of Theorem 1.

Theorem 2 *Let H_0 be an induced subgraph of G and suppose every subgraph H satisfying $H_0 \subset H \subset G$, $V(H_0) \ne V(H)$, contains a vertex $x \in V(H) - V(H_0)$ with $d_H(x) \le k$. Then*

$$\chi(G) \le \max\{k + 1, \chi(H_0)\}.$$ □

In some cases the problem of colouring a graph can be reduced to the problem of colouring certain subgraphs of it. This happens if the graph is disconnected or has a cutvertex or, slightly more generally, contains a complete subgraph whose vertex set disconnects the graph. Then we may colour each part separately since, at worst by a change of notation, we can fit these colourings together to produce a colouring of the original graph, as shown in Fig. V.2.

As a rather crude consequence of Theorem 1 we see that $\chi(G) \le \Delta + 1$, where $\Delta = \Delta(G)$ is the maximal degree of G, since $\max_{H \subset G} \delta(H) \le \Delta(G)$. Furthermore, if G is connected and not Δ-regular, then clearly $\max_{H \subset G} \delta(H) \le \Delta - 1$, so $\chi(G) \le \Delta$. The following result, due to Brooks, takes care of the regular case.

Theorem 3 *Let G be a connected graph with maximal degree Δ. Suppose G is neither a complete graph nor an odd cycle. Then $\chi(G) \le \Delta$.*

Proof. We know already that we may assume without loss of generality that G is 2-connected and Δ-regular. Furthermore, we may assume that $\Delta \ge 3$, since a connected 2-regular 3-chromatic graph is an odd cycle.

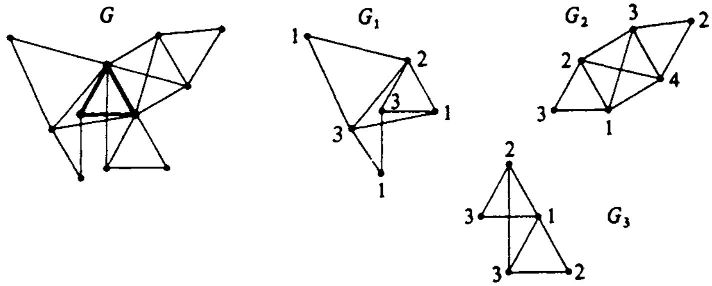


FIGURE V.2. The vertex set of the thick triangle disconnects G, and we find that $\chi(G) = \max\{\chi(G_1), \chi(G_2), \chi(G_3)\}$.

If G is 3-connected, let x_n be any vertex of G and let x_1, x_2 be two nonadjacent vertices in $\Gamma(x_n)$. Such vertices exist since G is regular and not complete. If G is not 3-connected, let x_n be a vertex for which $G - x_n$ is separable, and thus has at least two blocks. Since G is 2-connected, each endblock of $G - x_n$ has a vertex adjacent to x_n. Let x_1 and x_2 be such vertices belonging to different endblocks.

In either case, we have found vertices x_1, x_2 and x_n such that $G - \{x_1, x_2\}$ is connected, $x_1x_2 \notin E(G)$, but $x_1x_n \in E(G)$ and $x_2x_n \in E(G)$. Let $x_{n-1} \in V - \{x_1, x_2, x_n\}$ be a neighbour of x_n, let x_{n-2} be a neighbour of x_n or x_{n-1}, etc. Then the order $x_1, x_2, x_3, \ldots, x_n$ is such that each vertex other than x_n is adjacent to at least one vertex following it. Thus the greedy algorithm will use at most Δ colours, since x_1 and x_2 get the same colour and x_n, the only vertex with Δ neighbours preceding it, is adjacent to both. $\square$

Another colouring algorithm can be obtained by reducing the problem to colouring two other graphs derived from G. This reduction also enables us to obtain some information about the number of colourings of a graph with a given set of colours.

Let a and b be nonadjacent vertices of a graph G. Let G' be obtained from G by joining a to b, and let G'' be obtained from G by identifying a and b. Thus in G'' there is a new vertex (ab) instead of a and b, which is joined to the vertices adjacent to at least one of a and b (Fig. V.3).

These operations are even more natural if we start with the G': then G is obtained from G' by *cutting* or *deleting* the edge ab, and G'' is obtained from G'' by *fusing*, or *contracting*, ab.

The colourings of G in which a and b get *distinct* colours are in 1-to-1 correspondence with the colourings of G'. Indeed $c : V(G) \to \{1, 2, \ldots, k\}$ is a

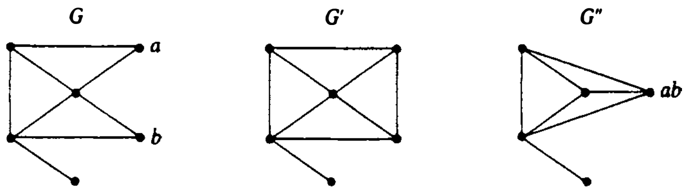


FIGURE V.3. The graphs G, G' and G''.

colouring of G with $c(a) \neq c(b)$ iff c is a colouring of G'. Similarly the colourings of G in which a and b get the *same* colour are in a 1-to-1 correspondence with the colourings of G''. In particular, if for a natural number x and a graph H we write $p_H(x)$ for the number of colourings of a graph H with colours $1, 2, \ldots, x$, then

$$p_G(x) = p_{G'}(x) + p_{G''}(x). \tag{4}$$

By definition $\chi(G)$ is the least natural number k for which $p_G(k) \geq 1$. Thus both the remarks above and relation (3) imply that

$$\chi(G) = \min\{\chi(G'), \chi(G'')\}. \tag{5}$$

The basic properties of $p_H(x)$ are given in our next result.

Theorem 4 *Let H be a graph with $n \geq 1$ vertices, m edges and k components. Then*

$$p_H(x) = \sum_{i=0}^{n-k} (-1)^i a_i x^{n-i},$$

where $a_0 = 1$, $a_1 = m$ and a_i is a positive integer for every i, $0 \leq i \leq n-k$.

Proof. We apply induction on $n + m$. For $n + m = 1$ the assertions are trivial so we pass to the induction step. If $m = 0$, we are again done, since in this case $k = n$ and, as *every map* $f : V(H) \to \{1, 2, \ldots, x\}$ is a colouring of H, we have $p_H(x) = x^n$. If $m > 0$ we pick two adjacent vertices of H, say a and b. Putting $G = H - ab$ we find that $G' = H$. Since $e(G) = m - 1$ and $|G''| + e(G'') \leq n - 1 + m$, by the induction hypothesis the assertions of the theorem hold for $p_G(x)$ and $p_{G''}(x)$. Note now that G'' has k components and G has at least k components. Therefore,

$$p_G(x) = x^n - (m-1)x^{n-1} + \sum_{i=2}^{n-k} (-1)^i b_i x^{n-i},$$

where b_i is a nonnegative integer for each i, and

$$p_{G''}(x) = x^{n-1} - \sum_{i=2}^{n-k} (-1)^i c_i x^{n-i},$$

where c_i is a positive integer for each i. Hence, by (3),

$$\begin{aligned} p_H(x) = p_{G'}(x) &= p_G(x) - p_{G''}(x) \\ &= x^n - mx^{n-1} + \sum_{i=2}^{n-k} (-1)^i (b_i + c_i) x^{n-i} \\ &= x^n - mx^{n-1} + \sum_{i=2}^{n-k} (-1)^i a_i x^{n-i}, \end{aligned}$$

where a_i is a positive integer for each i. $\square$

As a trivial consequence of Theorem 4, we see that $p_H(x)$ is a polynomial, so we are justified in calling it the *chromatic polynomial* of H. In fact, it is very easy to see from first principles that $p_H(x)$ is a polynomial in x with integer coefficients. Write $\pi_r(H)$ for the number of partitions of $V(H)$ into r non-empty independent sets. Then for every natural number x we have

$$p_H(x) = \sum_{r=1}^{n} \pi_r(H)(x)_r,$$

where $(x)_r = x(x-1)(x-2)\cdots(x-r+1)$ is the falling factorial. The coefficients of the chromatic polynomial have a fairly simple interpretation.

Theorem 5 *Let H be a graph with n vertices and edge set $E(H) = \{e_1, e_2, \dots, e_m\}$. Call a subset of $E(H)$ a broken cycle if it is obtained from the edge set of a cycle by deleting the edge of highest index. Then the chromatic polynomial of H is*

$$p_H(x) = \sum_{i=0}^{n-1} (-1)^i a_i x^{n-i},$$

where a_i is the number of i-subsets of $E(H)$ containing no broken cycle.

Proof. Let us apply induction on m. For $m = 0$ the assertion is trivial, so suppose that $m \geq 1$ and the assertion holds for smaller values of m. Let $e_1 = ab$ and, as before, set $G = H - ab$, so that $G' = G + ab = H$ and $G'' = G/ab$ satisfy (1).

With a slight abuse of notation, we identify not only $E(G) = \{e_2, e_3, \dots, e_m\}$, but also $E(G'')$, with a subset of $E(H)$. If an edge of $E(G'')$ comes from only one edge of $E(G)$, we keep its notation, and if an edge $(ab)x$ comes from two edges of G, say $e_i = ax$ and $e_h = bx$, then we denote $(ab)x$ by e_k, where $k = \max\{i, h\}$.

As (1) holds, to complete the induction step, all we have to check is that the number of i-subsets of $E(G')$ containing no broken cycle of G' is precisely the sum of the number of i-subsets of $E(G)$ containing no broken cycle and the number of $(i-1)$-subsets of $E(G'')$ containing no broken cycle. But this is a consequence of the following two simple assertions.

(1) Suppose $e_1 \notin F \subset E(G')$. Then F contains no broken cycle of G' iff F contains no broken cycle of G.

(2) Suppose $e_1 \in F \subset E(G')$. Then F contains no broken cycle of G' if $F - \{e_1\} \subset E(G'')$ and $F - \{e_1\}$ contains no broken cycle of G''. $\square$

As a by-product of Theorem 5, we see that the number of i-subsets of $E(H)$ containing no broken cycle is *independent* of the order imposed on $E(H)$—a fact which is far from obvious.

In general, Theorem 5 does not provide a practical method for determining the coefficients of the chromatic polynomial. However, if the graph has no short cycles then it does give us the first few coefficients.

Corollary 6 *Let H be a graph with n vertices, m edges, girth g and chromatic polynomial*

$$p_H(x) = \sum_{i=0}^{n} (-1)^i a_i x^{n-i}.$$

Then $a_i = \binom{m}{i}$ for $i \leq g-2$. Furthermore, if g is finite and H has c_g cycles of length g then $a_{g-1} = \binom{m}{g-1} - c_g$. □

The reduction $G \to \{G', G''\}$ also gives us a natural, although not very practical, algorithm for finding the chromatic number. Given a graph G, construct a sequence of graphs $G_0, G_1, \ldots$ as follows. Put $G_0 = G$. Having constructed G_i, if G_i is complete, terminate the sequence; otherwise, let G_{i+1} be G_i' or G_i''. The sequence has to end in a complete graph G_t, say of order $|G_t| = k$. A k-colouring of G_t can easily be lifted to a k-colouring of the original graph G, so $\chi(G) \leq k$. Equality (4) shows that if we construct all possible sequences from G then $\chi(G)$ is precisely the maximal order of a terminal graph.

There are other problems that can be tackled by the reduction $G \to \{G', G''\}$; a beautiful example is Exercise 15^+.

In Chapter X we shall return to this topic, when we study a substantial generalization of the chromatic polynomial, the *Tutte polynomial.* As we shall see, one of the most important properties of the Tutte polynomial is that it can be defined by the analogues of the cut and fuse operations for *multigraphs.*

V.2 Edge Colouring

In a colouring of the edges of a graph G, the edges incident with a vertex get distinct colours, so $\chi'(G)$, the edge-chromatic number, is at least as large as the maximal degree, $\Delta(G) = \max_x d(x)$:

$$\chi'(G) \geq \Delta(G). \tag{6}$$

At first sight it is somewhat surprising that this trivial inequality is, in fact, an equality for large classes of graphs, including the class of *bipartite graphs.* Indeed, Exercise 22 of Chapter III, which is an easy consequence of Hall's theorem, asserts that the edge set $E(G)$ of a bipartite graph G can be partitioned into $\Delta(G)$ classes of independent edges, that is, $\chi'(G) = \Delta(G)$.

Another trivial lower bound on $\chi'(G)$ follows from the fact that if G does not contain $\beta + 1$ independent edges, then each colour class has at most β edges, so we need at least $\lceil e(G)/\beta \rceil$ colour classes to take care of all the edges:

$$\chi'(G) \geq \lceil e(G)/\beta \rceil. \tag{7}$$

Proceeding as in the proof of Theorem I. 11, it is easy to show that if G is a complete graph of order at least 2 then equality holds in (7), that is, $\chi'(K^n) = n - 1$ if n is even, and $\chi'(K^n) = n$ if $n \geq 3$ is odd (Exercise 29).

How can one obtain an upper bound for $\chi'(G)$? Since each edge is adjacent to at most $2(\Delta(G) - 1)$ edges, Theorem 1 implies that

$$\chi'(G) \le 2\Delta(G) - 1.$$

Furthermore, if $\Delta(G) \ge 3$, then Brooks' theorem gives

$$\chi'(G) = \chi(L(G)) \le 2\Delta(G) - 2.$$

At first sight this inequality seems reasonably good. However, the following fundamental theorem of Vizing shows that this is not the case, because the edge-chromatic number is always very close to the maximal degree.

Theorem 7 *A graph G of maximal degree Δ has edge-chromatic number Δ or $\Delta + 1$.*

Proof. Let us assume that we have used $1, 2, \ldots, \Delta + 1$ to colour all but one of the edges. We are home if we can show that by recolouring some of the edges, we can colour this last edge as well with one of $1, 2, \ldots, \Delta + 1$.

We say that a colour is *missing at* a vertex z if no edge incident with z gets that colour. If z is incident with $d'(z) \le d(z) \le \Delta$ edges that have been coloured, then $\Delta + 1 - d'(z)$ colours are missing at z. In particular, at each vertex at least one colour is missing. Our aim is to move around the colours and the uncoloured edge in such a way that a colour will be missing at both endvertices of the uncoloured edge, enabling us to complete the colouring.

Let xy_1 be the uncoloured edge; let s be a colour missing at x and let t_1 be a colour missing at y_1. We shall construct a sequence of edges $xy_1, xy_2, \ldots$, and a sequence of colours $t_1, t_2, \ldots$ such that t_i is missing at y_i and xy_{i+1} has colour t_i. Suppose we have constructed $xy_1, \ldots, xy_i$ and $t_1, \ldots, t_i$. There is at most one edge xy of colour t_i. If $y \notin \{y_1, \ldots, y_i\}$, we put $y_{i+1} = y$ and pick a colour t_{i+1} missing at y_{i+1}, otherwise we stop the sequence. These sequences have to terminate after at most $\Delta(G)$ terms; let $xy_1, \ldots, xy_h$ and $t_1, \ldots, t_h$ be the complete sequences. Let us examine the two reasons that may have forced us to terminate these sequences.

(a) *No edge xy has colour t_h.* Then recolour the edges xy_i, $i < h$, giving xy_i colour t_i. In the colouring we obtain, every edge is coloured except xy_h. However, since t_h occurs neither at x nor at y_h, we may complete the colouring by assigning t_h to xy_h.

(b) *For some $j < h$ the edge xy_j has colour t_h.* To start with, recolour the edges xy_i, $i < j$, giving xy_i colour t_i. In this colouring the uncoloured edge is xy_j. Let $H(s, t_h)$ be the subgraph of G formed by the edges of colour s and t_h, where s is the original colour missing at x and t_h is missing at y_h. Each vertex of $H(s, t_h)$ is incident with at most 2 edges in $H(s, t_h)$ (one of colour s and the other of colour t_h), so the components of $H(s, t_h)$ are paths and cycles. Each of the vertices x, y_j and y_h has degree at most 1 in $H(s, t_h)$, so they cannot all belong to the same component of $H(s, t_h)$. Thus at least one of the following two cases has to hold.

(b1) *The vertices x and y_j belong to distinct components of $H(s, t_h)$.* In this case interchange the colours s and t_h in the component containing y_j. Then s is missing at both x and y_j, so we may complete the colouring by giving xy_j colour s.

(b2) *The vertices x and y_h belong to distinct components of $H(s, t_h)$.* Now continue the recolouring of the edges incident with x by giving xy_i colour t_i for each $i < h$, thereby making xy_h the uncoloured edge. This change does not involve edges of colours s and t_h, so $H(s, t_h)$ has not been altered. Now switch around the colours in the component containing y_h. This switch makes sure that s is missing at both x and y_h, so we can use s to colour the so far uncoloured edge xy_h. □

Note that the proof above gives an algorithm for colouring the edges with at most $\Delta + 1$ colours.

V.3 Graphs on Surfaces

There is no doubt that for well over a hundred years the best known problem in graph theory was the the four colour problem: prove that every plane graph is 4-colourable. After numerous false starts and partial results, the problem was solved in 1976 by Appel and Haken, relying on ideas of Heesch, when they proved that every plane graph can indeed be coloured with four colours. On the other hand, Euler's formula implies that every plane graph can be coloured with 6 colours. Indeed, by Theorem I.16, every plane graph of order n has at most $3n - 6$ edges and so its minimal degree is at most 5. Hence, by Theorem 1, the chromatic number is at most 6. Furthermore, with a little more work we can obtain the following stronger assertion.

Theorem 8 *Every plane graph is 5-colourable.*

Proof. Suppose the assertion is false and let G be a 6-chromatic plane graph with minimal number of vertices. As above, we know that G has a vertex x of degree at most 5. Put $H = G - x$. Then H is 5-colourable, say with colours $1, 2, \ldots, 5$. Each of these colours must be used to colour at least one neighbour of x, otherwise the missing colour could be used to colour x. Hence we may assume that x has 5 neighbours, say $x_1, x_2, \ldots, x_5$ in some cyclic order about x, and the colour of x_i is i, $i = 1, 2, \ldots, 5$. Denote by $H(i, j)$ the subgraph of H spanned by vertices of colour i and j.

Suppose first that x_1 and x_3 belong to distinct components of $H(1, 3)$. Interchanging the colours 1 and 3 in the component of x_1, we obtain another 5-colouring of H. However, in this 5-colouring both x_1 and x_3 get colour 3, so 1 is not used to colour any of the vertices $x_1, \ldots, x_5$. This is impossible because then x can be coloured 1.

Since x_1 and x_3 belong to the same component of $H(1, 3)$, there is an x_1–x_3 path P_{13} in H whose vertices are coloured 1 and 3. Analogously, H contains an x_2–x_4 path P_{24} whose vertices are coloured 2 and 4. However, this is impossible,

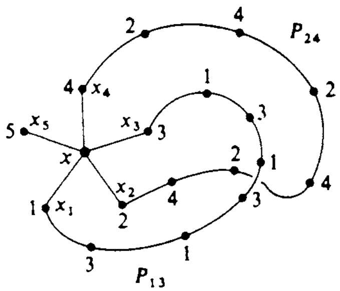


FIGURE V.4. The paths P_{13} and P_{24}.

since the cycle $x_1 P_{13} x_3$ of G separates x_2 from x_4 but P_{24} cannot meet this cycle (Fig. V.4). □

Clearly, not every plane graph is 3-colourable. Indeed, K_4 is planar and it does need 4 colours. Another 4-chromatic planar graph is obtained by joining all five vertices of a C_5 to a sixth vertex. Thus $\chi_0 = \max\{\chi(G) : G \text{ is planar}\}$ trivially satisfies $\chi_0 \geq 4$ and $\chi_0 \leq 5$, and the problem is to prove $\chi_0 \leq 4$.

Instead of a plane graph, we may wish to consider a graph drawn on a closed surface of arbitrary Euler characteristic. We shall see in a moment that, rather curiously, the plane is the exception: for every closed surface other than the plane, the problem is of an entirely different nature (and much easier).

We shall need very little about closed surfaces: in fact, all we need is their classification theorem and the Euler–Poincaré formula. For $p > 0$, let S_p be the closed surface obtained from a $4p$-gon by identifying pairs of sides, as in Fig. V.5(i), and for $q > 0$, let N_q be the closed surface obtained from a $2q$-gon by identifying pairs of sides, as in Fig. V.5(ii). Thus S_1 is the torus, N_1 is the projective plane and N_2 is the Klein bottle; also, let S_0 be the sphere. By the classification theorem, every closed surface is homeomorphic to precisely one of the *orientable* surfaces $S_0, S_1, \ldots$ or one of the *non-orientable* surfaces $N_1, N_2, \ldots$. For $p \geq 0$, the surface S_p has *genus* p and *Euler characteristic* $\chi = \chi(S_p) = 2(1-p)$, and for $q > 0$, the surface N_q has *genus* q and *Euler characteristic* $\chi = \chi(N_q) = 2-q$. It is rather unfortunate that χ is the standard symbol for both the Euler characteristic of a surface and the chromatic number of a graph. This conflict will occur only in this section and, hopefully, it will not lead to any confusion.

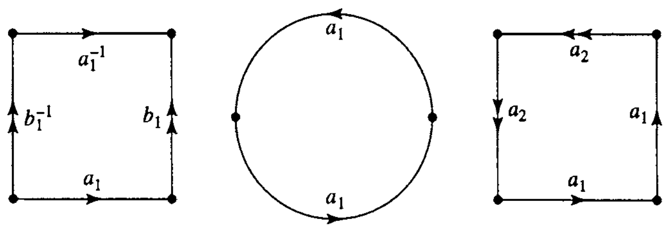


FIGURE V.5. The torus S_1, the projective plane N_1 and the Klein bottle N_2.

A *triangulation* of a surface is a drawing of a graph on the surface such that every face is a triangle. The Euler–Poincaré formula states that if a triangulation of a closed surface of Euler characteristic χ has α_0 vertices, α_1 edges and α_2 faces, then $\alpha_0 - \alpha_1 + \alpha_2 = \chi$. An immediate consequence of this is that if a graph G of order n is drawn on a surface of Euler characteristic χ, then

$$e(G) \leq 3n - 3\chi, \tag{8}$$

with equality iff G is a triangulation of the surface.

The following easy upper bound on the chromatic number of a graph drawn on a closed surface was obtained by Heawood in 1890.

Theorem 9 *The chromatic number of a graph G drawn on a closed surface of Euler characteristic $\chi \leq 1$ is at most*

$$h(\chi) = \lfloor (7 + \sqrt{49 - 24\chi})/2 \rfloor.$$

Proof. Let k be the chromatic number of G. We may and shall assume that G is a minimal graph of chromatic number k; otherwise, we may replace it by a subgraph. But then $\delta(G) \geq k - 1$, so all we need is that if, for $h = h(\chi)$, G has $n \geq h + 1$ vertices then its minimal degree is at most $h - 1$. Now, if $n \geq h + 1$ then $e(G) \leq 3n - 3\chi$ implies that

$$\delta(G) \leq 6 - 6\chi/(h + 1).$$

Hence if we had $\delta(G) \geq h$ then we would have

$$h \leq 6 - 6\chi/(h + 1),$$

that is,

$$h^2 - 5h + 6(\chi - 1) \leq 0.$$

But this would imply the contradiction

$$h \leq \frac{1}{2}(5 + \sqrt{49 - 24\chi}). \qquad \square$$

For a surface M, define its *chromatic number*, $s(M)$, as the maximum of the chromatic numbers of graphs drawn on M. Trivially, $s(S_g) \leq s(S_{g+1})$ since every graph that can be drawn on S_g can also be drawn on S_{g+1}; similarly, $s(N_g) \leq s(N_{g+1})$. The simple Theorem 9 states that if M is a surface of Euler characteristic χ then the chromatic number $s(M)$ is at most as large as the *Heawood bound* $h(\chi) = \lfloor (7 + \sqrt{49 - 24\chi})/2 \rfloor$.

When does equality hold? The following easy result shows that, for most values of χ, what matters is whether a complete graph can be drawn on a surface.

Theorem 10 *Let $\chi \leq 0$, $h = h(\chi) = \lfloor (7 + \sqrt{49 - 24\chi})/2 \rfloor$, and let G be a minimal h-chromatic graph drawn on a surface of Euler characteristic χ. If $\chi \neq -1, -2$ or -7 then $G = K_h$.*

Proof. All we shall use is inequality (8): a graph of order n drawn on a surface of Euler characteristic χ has at most $3(n - \chi)$ edges.

Suppose $G \neq K_h$. Then $n \geq h+2$. Furthermore, if $n = h+2$ then, as claimed by Exercise 38,

$$e(G) = \binom{h+2}{2} - 5,$$

which is easily checked to be greater than $3(h+2-\chi)$. Hence $n \geq h+3$. Our graph G is a minimal h-chromatic graph, so $\delta(G) \geq h-1 \geq 6$ and, by Brooks' theorem, G is not $(h-1)$-regular. Therefore

$$e(G) > \frac{n(h-1)}{2},$$

and so

$$n(h-1)+1 \leq 6(n-\chi). \tag{9}$$

Since $h \geq 1$, inequality (9) has to hold for $n = h+3$, that is,

$$h^2 - 4h - 20 + 6\chi \leq 0.$$

This implies that

$$h \leq 2 + \sqrt{24 - 6\chi}. \tag{10}$$

Simple calculations show that (10) fails for $\chi \leq -20$, and it is easily checked that for $-19 \leq \chi \leq 0$ inequality (10) fails unless $\chi = -1, -2$ or -7. □

In fact, Theorem 10 holds without any exceptions: this can be proved by using a slightly better bound on the size of a minimal h-chromatic graph of order n.

From Theorem 10, it is easy to determine the chromatic number of a surface of small genus other than the sphere.

Theorem 11 *The torus, the projective plane and the Klein bottle have chromatic numbers $s(S_1) = 7$, $s(N_1) = 6$ and $s(N_2) = 6$.*

Proof. The Euler characteristics of these surfaces are $\chi(N_1) = 1$ and $\chi(S_1) = \chi(N_2) = 0$, therefore Theorem 9 implies that $s(N_1) \leq 6$ and $s(S_1), s(N_2) \leq 7$. Fig. V.6 shows that K_6 triangulates N_1 and K_7 triangulates S_1, so $s(N_1) = 6$, $s(S_1) = 7$ and $6 \leq s(N_2) \leq 7$.

Our problem is then to decide whether the chromatic number of the Klein bottle is 6 or 7. We know from Theorem 10 that $s(N_2) = 7$ iff K_7 can be drawn on N_2, and so K_7 triangulates N_2. To complete the proof, we shall show that K_7 triangulates a unique closed surface, the torus, so that $s(N_2) = 6$.

Suppose then that we have a triangulation by K_7 of a closed surface (of Euler characteristic 0). Then every vertex of K_7 is on the boundary of six triangular faces, and the third sides of these triangles form a 6-cycle. Writing $0, 1, \ldots, 6$ for the vertices, we may assume that the 6-cycle 'surrounding' 0 is 123456. Then vertex 1 is surrounded by $602x \cdot y$, vertex 2 by $301x \cdot \cdot$, and so on (see Fig. V.7). But then x has to be 4 or 5: by symmetry, we may assume that it is 4. Having made this choice, everything else is determined: looking at the neighbourhoods of 1 and 6, namely the cycles $y6024\cdot$ and $501y \cdot \cdot$, we see that $y = 3$, then we

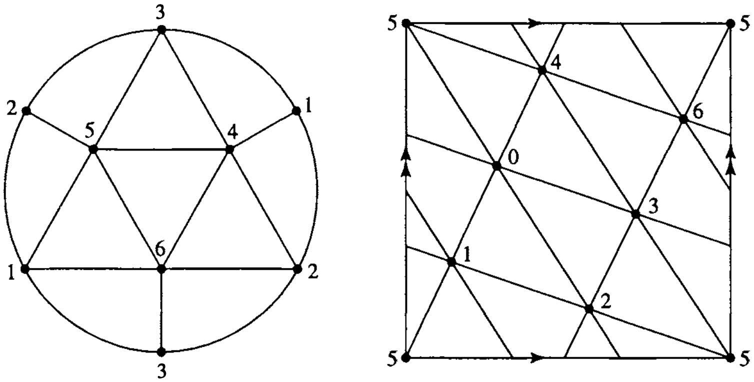


FIGURE V.6. Triangulations of the projective plane N_1 by K_6, and of the torus S_1 by K_7.

get $z = 2$, $u = 1$, and so on, as shown in Fig. V.7. What we have proved is that *if* K_7 triangulates a surface then this triangulation is unique (up to reflection) and is as in Fig. V.7. But this labelling is easily seen to be consistent and to give a triangulation of the torus. (As it happens, we already know that K_7 triangulates the torus, but in this proof we were forced to find that triangulation.) In particular, K_7 cannot be drawn on the Klein bottle, so $s(N_2) = 6$, and we are done. □

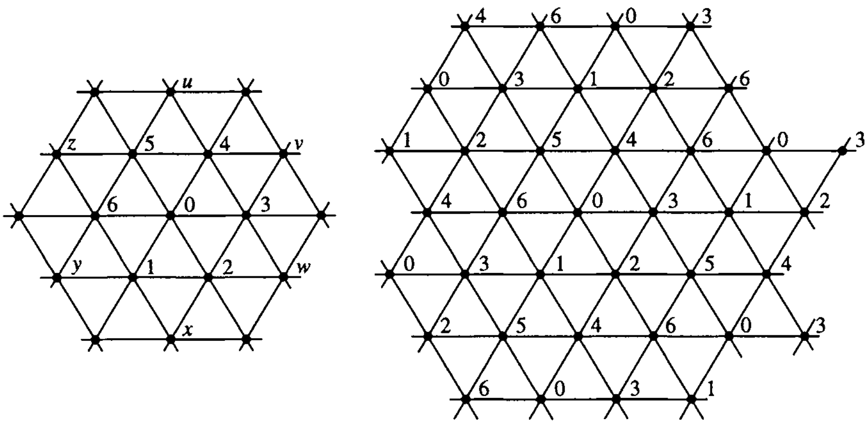


FIGURE V.7. The start of a triangulation given by K_7, and the labelling of the entire triangular lattice.

In fact, the Heawood bound $h(\chi)$ in Theorem 9 is best possible for every closed surface other than the Klein bottle: if M is a closed surface of Euler characteristic $\chi \leq 1$ and M is not the Klein bottle, then $s(M) = h(\chi)$. Although this was claimed by Heawood in 1890, his proof was incorrect, and the assertion became known as Heawood's conjecture. The first correct proof of Heawood's conjecture was found by Ringel and Youngs only over 75 years later. Note that the difficulty in proving this deep result lies in finding a drawing of a single fixed graph, $K_{h(\chi)}$, on a surface of Euler characteristic $\chi \leq 1$. What we have to do for $\chi \leq -1$ is rather similar to the proof of Theorem 11: we have to find a 'consistent colouring' of a triangular tessellation of the hyperbolic plane in which every vertex has degree $h(\chi) - 1$. On the other hand, in order to solve the four colour problem one has to show that *every* plane graph can be coloured with four colours. Thus the difficulty in solving the four colour problem has almost nothing to do with the problem of determining $s(M)$ for $\chi(M) \leq 1$.

For fear of upsetting the balance of the book, we shall say only a few words about the solution of the four colour problem. We saw in Section I.4 that a plane graph G determines a *map* $M = M(G)$ consisting of the plane graph G and the countries determined by the plane graph. A colouring of a map is a colouring of the countries such that no two countries sharing an edge in their boundaries get the same colour. The original form of the four colour problem, as posed by Francis Guthrie in 1852, asked for a proof of the assertion that every plane map can be coloured with four colours. His teacher, de Morgan, circulated the problem amongst his colleagues, but it was first made popular in 1878 by Cayley, who mentioned it before the Royal Society. Almost at once "proofs" appeared, by Kempe in 1879 and by Tait in 1890. Heawood's refutation of Kempe's proof was published in 1890, though he modified the proof to obtain the five colour theorem. Tait's paper also contained false assumptions, which prompted Petersen to observe in 1891 that the four colour theorem is equivalent to the conjecture that every 2-connected cubic planar graph has edge chromatic number three (Exercise 28^+). Contributions to the solution since the turn of the century include Birkhoff's introduction of the chromatic polynomial and works by various authors giving lower bounds on the order of a possible counterexample. In 1943 Hadwiger made a deep conjecture containing the four colour theorem as a special case: if $\chi(G) = k$, then G is contractible to K_k (see Exercises 16–18).

In hindsight, the most important advance was made by Heesch. The problem was at last solved by Appel and Haken in 1976, making use of a refinement of Heesch's method and fast computers. The interested reader is referred to some papers of Appel and Haken, to the book of Saaty and Kainen, and a recent paper of Robertson, Sanders, Seymour and Thomas for a detailed explanation of the underlying ideas of the proof. All we have room for is a few superficial remarks.

What makes the *five colour* theorem true? The following two facts: (i) a minimal 6-chromatic plane graph cannot contain a vertex of degree at most 5, and (ii) a plane graph has to contain a vertex of degree at most 5. We can go a step further and ask why (i) and (ii) hold. A look at the proof shows that (i) is proved by making a good use of the paths P_{ij}, called *Kempe chains* after Kempe, who used them in his false proof of 1870, and (ii) follows immediately from Euler's formula $n - e + f = 2$.

The attack on the four colour problem initiated by Heesch goes along similar lines. A *configuration* is a connected cluster of vertices of a plane graph together with the degrees of the vertices. A configuration is *reducible* if no minimal 5-chromatic plane graph can contain it and a set of configurations is *unavoidable* if every plane graph contains at least one configuration belonging to the set. In order to prove that every plane graph is 3-colourable, one sets out to find an *unavoidable set of reducible configurations*. How should one show that a configuration is reducible? Replace the cluster of vertices by a smaller cluster, 4-colour the obtained smaller graph and use Kempe chains to show that the 4-colouring can be "pulled back" to the original graph. How should one show that a set of configurations is unavoidable? Make extensive use of Euler's formula. Of course, one may always assume that the graph is a maximal plane graph. Assigning a *charge* of $6 - k$ to a vertex of degree k, Euler's formula guarantees that the total charge is 12. Push charges around the vertices according to some *discharging rules*, that is, transfer some charge from a vertex to some of its neighbours, until it transpires that the plane graph has to contain one of the configurations.

Looking again at the five colour theorem, we see that the proof was based on the fact that the configurations consisting of single vertices of degree at most 5 form an unavoidable set of configurations (for the five colour theorem).

The simplistic sketch above does not indicate the difficulty of the actual proof. In order to rectify this a little, we mention that Appel and Haken needed over 1900 reducible configurations and more than 300 discharging rules to complete the proof. Furthermore, we invite the reader to prove the following two simple assertions.

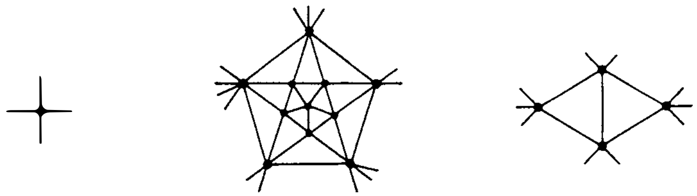


FIGURE V.8. Three reducible configurations; in the last two examples the outer vertices may have arbitrary degrees.

1. The configurations in Fig. V.8 are reducible.
2. Let G be a maximal planar graph of order at least 25 and minimal degree 5. Call a vertex a *major* vertex if its degree is at least 7, otherwise, call it *minor*. Then G contains one of the following:
   (a) a minor vertex with 3 consecutive neighbours of degree 5,
   (b) a vertex of degree 5 with minor neighbours only,
   (c) a major vertex with at most one neighbour of degree at least 6.

For twenty years, the Appel and Haken proof was neither simplified, nor thoroughly checked, as in addition to the huge program, the proof requires that some 1400 graphs be put into the computer by hand. Recently, however, Robertson, Sanders, Seymour and Thomas produced their version of the proof, with an unavoidable set of 'only' 633 reducible configurations, and with 'only' 32 discharging rules. This proof is considerably easier to check, since the immense task of checking unavoidability by hand is replaced by a formally written proof, which can be read and verified by a computer in a few minutes.

V.4 List Colouring

Recall that a graph is k-colourable iff to every vertex x we can assign a colour $c(x) \in [k] = \{1, 2, \dots, k\}$ such that adjacent vertices get distinct colours. Now suppose that to every vertex x of a k-colourable graph we assign a paint-box or *list* $L(x)$ of k colours. Is it possible to assign to each vertex one of the colours from its own list such that adjacent vertices get distinct colours? At first sight, it seems trivial that such an assignment is always possible, since *"surely the worst case is when the lists are identical, as that maximizes the chances of a conflict."*

However, this first impression is clearly misleading. For example, let G be the complete three by three bipartite graph $K_{3,3}$ with vertex classes $V_1 = \{x_1, x_2, x_3\}$ and $V_2 = \{y_1, y_2, y_3\}$, and let $L(x_i) = L(y_i) = \{1, 2, 3\} - \{i\}, i = 1, 2, 3$. Then in any colouring of the vertices from these lists, at least two colours must be used to colour V_1, and at least two to colour V_2, so there is bound to be an edge joining two vertices of the same colour (see Fig. V.9). This realization leads to an important variant of the chromatic number, the *list-chromatic number.*

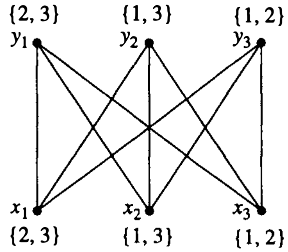


FIGURE V.9. The graph $K_{3,3}$ with lists of size 2 assigned to the vertices, without a proper colouring from the lists.

Given a graph G and a map L assigning to each vertex a set $L(x)$, an L-colouring of G is a *proper* colouring c of the vertices such that $c(x) \in L(x)$ for every $x \in V(G)$. The *list-chromatic number* $\chi_\ell(G)$ of G is the minimal integer k such that G has an L-colouring whenever $|L(x)| \geq k$ for every $x \in V(G)$. Clearly, $\chi_\ell(G) \geq \chi(G)$ for every graph G, since $\chi(G)$ is the minimal integer k such that G has an L-colouring when $L(x) = [k]$ for every $x \in V(G)$.

The example above shows that we may have $\chi_\ell(G) \geq 3$ and $\chi(G) = 2$. In fact, it is easily seen that for every $k \geq 2$ there is a bipartite graph G with $\chi_\ell(G) > k$. Indeed, writing $A^{(k)}$ for the set of all k-subsets of a set A, let G be the complete bipartite graph with vertex classes $V_1 = \{x_1, x_2, \dots, x_{2k-1}\}^{(k)}$ and $V_2 = \{y_1, y_2, \dots, y_{2k-1}\}^{(k)}$. Also, for $\mathbf{x} = \{x_{i_1}, x_{i_2}, \dots x_{i_k}\} \in V_1$ and $\mathbf{y} = \{y_{i_1}, y_{i_2}, \dots y_{i_k}\} \in V_2$, set $L(\mathbf{x}) = L(\mathbf{y}) = \{i_1, i_2, \dots, i_k\}$. Then G is bipartite and has no L-colouring, since in any L-colouring we would have to use at least k colours to colour V_1 and at least k colours to colour V_2, so we would have two adjacent vertices with the same colour. Hence $\chi(G) = 2$ and $\chi_\ell > k$. On the other hand, the greedy algorithm shows that $\chi_\ell(G) \le \Delta(G) + 1$ for every graph G.

Our aim in this section is to prove two beautiful results, due to Thomassen and Galvin, claiming that under certain circumstances the list-chromatic number is not much larger than the chromatic number. These theorems strengthen considerably two of our rather simple earlier results. As the proofs are short and very elegant, the reader may be surprised to learn that much effort had gone into proving these results before Thomassen and Galvin found their ingenious proofs.

We start with Thomassen's theorem, strengthening Theorem 8 by claiming that the list-chromatic number of a planar graph is at most 5. The proof below is a striking example of the admirable principle that it is frequently *much easier* to prove an appropriate generalization of an assertion than the original clean assertion. In this case the generalization concerns list-colourings of almost maximal planar graphs, with *varying* list sizes. To be precise, call a plane graph a *near-triangulation* if the outer face is a cycle and all the inner faces are triangles.

As in a maximal plane graph of order at least 4 *every* face is a triangle, the following result is clearly stronger than the assertion that every planar graph has list-chromatic number at most 5.

Theorem 12 *Let G be a near-triangulation with outer cycle $C = x_1x_2\cdots x_k$, and for each $x \in V(G)$ let $L(x)$ be a list of colours assigned to x, such that $L(x_1) = \{1\}$, $L(x_2) = \{2\}$, $|L(x)| \ge 3$ for $3 \le i \le k$, and $|L(x)| \ge 5$ for $x \in V(G - C)$. Then G has an L-colouring.*

Proof. Let us apply induction on the order of G. For $|G| = 3$ the assertion is trivial, so suppose that $|G| > 3$ and the assertion holds for graphs of order less than $|G|$. We shall distinguish two cases, according to whether C contains a 'diagonal' from x_k or not.

(i) First suppose that G contains a 'diagonal' x_kx_j, $2 \le j \le k-2$, of C. Then we can apply the induction hypothesis to the graph formed by the cycle $x_kx_1x_2\cdots x_j$ and its interior and *then*, having fixed the colours of x_k and x_j, to the cycle $x_kx_jx_{j+1}\cdots x_{k-1}$ and its interior, to find an L-colouring of G.

(ii) Now suppose that G contains none of the edges x_kx_j, $2 \le j \le k-2$. Let the neighbours of x_k be $x_{k-1}, y_1, y_2, \ldots, y_\ell$ and x_1, in this order, so that $x_kx_{k-1}y_1$, $x_ky_1y_2, \ldots, x_ky_\ell x_1$ are internal faces of our plane graph (see Fig. V.10).

Let a and b be colours in $L(x_k)$, distinct from 1. Our aim is to use one of a and b to colour x_k, having coloured the rest of the graph. To this end, let $L'(x) = L(x)$ if $x \notin \{y_1, \ldots, y_\ell\}$ and $L'(y_i) = L(y_i) - \{a, b\}$ for $1 \le i \le l$. Then, by the induction hypothesis, the graph $G' = G - x_k$, with outer cycle $x_1x_2\cdots x_{k-1}y_1y_2\cdots y_\ell$, has an L'-colouring. Extend this L'-colouring of G' to an L-colouring of G by assigning a or b to x_k such that x_k and x_{k-1} get distinct colours. □

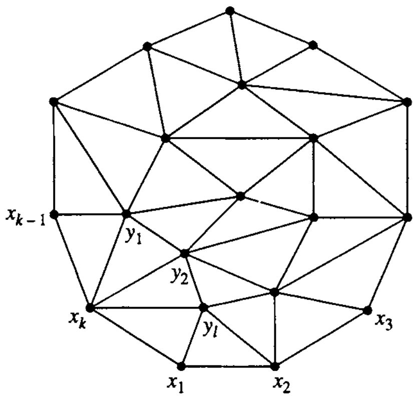


FIGURE V.10. The second case in the proof of Theorem 12.

Theorem 12 is not only considerably stronger than Theorem 8, the five colour theorem, but it is also best possible: as shown by Voigt, there are planar graphs of list-chromatic number exactly 5.

Our next aim is to prove Galvin's theorem concerning list-colourings of the *edges* of a bipartite graph. Suppose that for every edge $e \in E(G)$ of a graph G, we are given a list $L(e)$ of colours. An *L-edge-colouring* of G is a proper edge-colouring λ of G such that $\lambda(e) \in L(e)$ for every $e \in E(G)$. For a function $f : E(G) \to \mathbb{N}$, we say that G is *f-edge-choosable* if G has an L-edge-colouring whenever $|L(e)| \geq f(e)$ for every $e \in E(G)$. The minimal k such that G is k-edge-choosable is called the *list-edge-chromatic number* of G, or the *list-chromatic index* of G, or the *edge-choosability number* of G, and is denoted by $\chi_l'(G)$ or $ch(G)$. To make this terminology a little less cumbersome, we shall frequently omit the word *edge* when there is no danger of confusion, so we shall talk of L-colourings and f-choosable graphs.

As we shall make use of the existence of a stable matching, we shall follow the conventions used in Section III. 5. Let G be a bipartite graph with bipartition (V_1, V_2) and a certain assignment of preferences. For $e = aA \in E(G)$ let $t_G(e)$ be the sum of the number of vertices the vertex a prefers to A and the number of vertices the vertex A prefers to a. We call $t_G : E(G) \to \mathbb{Z}^+ = \{0, 1, \ldots\}$ the *total function* of the assignment of preferences.

Note that if H is a subgraph of G and $E_0 \subset E(H)$ then

$$t_G(e) - t_{G-E_0}(e) = t_H(e) - t_{H-E_0}(e) \tag{11}$$

for every edge $e \in E(H) - E_0$. (Needless to say, the preferences in subgraphs of G are taken as in G.) Furthermore, a matching M in H is stable iff

$$t_H(e) - t_{H-M}(e) \geq 1 \tag{12}$$

for every edge $e \in E(H) - M$.

After all this preparation it is easy to state and prove a result that will readily imply that $\chi_\ell'(G) = \chi'(G)$ for every bipartite graph G.

Theorem 13 *Let G be a bipartite graph with total function t_G given by a certain assignment of preferences. Then G is $(t_G + 1)$-choosable.*

Proof. We apply induction on the size of G. If $E(G) = \emptyset$, there is nothing to prove, so suppose $E(G) \neq \emptyset$ and the assertion holds for graphs of smaller size.

Let us fix an assignment of preferences for G. For each edge $e \in E(G)$, let $L(e)$ be a set of $t_G(e) + 1$ natural numbers. We have to show that the edges of G have an L-colouring.

Let $I \neq \emptyset$ be the set of edges whose lists contain a certain colour i, and let $H = (V(G), I)$ be the subgraph of G with edge-set I. By Theorem III.15, the graph H contains a stable matching M. Let $G' = G - M$, and for $e \in E(G')$ set $L'(e) = L(e) - \{i\}$. We claim that

$$|L'(e)| \geq t_{G'}(e) + 1 \tag{13}$$

for every $e \in E(G')$. Indeed, if $e \notin I$ then $L'(e) = L(e)$ so this is clearly the case. Also, if $e \in I - M = E(H) - M$ then, by relations (11) and (12),

$$t_G(e) - t_{G'}(e) = t_H(e) - t_{H'}(e) \geq 1,$$

so

$$|L'(e)| = |L(e)| - 1 \geq t_G(e) \geq t_{G'}(e) + 1,$$

proving (13).

By the induction hypothesis, G' has an L'-colouring; colouring the edges of M by i, we get an L-colouring of the edges of G. □

From here it is but a short step to Galvin's theorem.

Theorem 14 *The list-chromatic index of a bipartite graph equals its chromatic index.*

Proof. Let G be a bipartite graph with bipartition (V_1, V_2), and let $\lambda : E(G) \to [k]$ be an edge-colouring of G, where k is the chromatic index of G. Define preferences on G as follows: let $a \in V_1$ prefer a neighbour A to a neighbour B iff $\lambda(aA) > \lambda(aB)$, and let $A \in V_2$ prefer a neighbour a to a neighbour b iff $\lambda(aA) < \lambda(bA)$. Note that the total function defined by this assignment of preferences is at most $k - 1$ on every edge, since if $\lambda(aA) = j$ then a prefers at most $k - j$ of its neighbours to A, and A prefers at most $j - 1$ of its neighbours to a. Hence, by Theorem 11, G is k-choosable. □

As we noted in Section 2, the chromatic index of a bipartite graph equals its maximal degree, so Theorem 14 can be restated as

$$\chi'_\ell(G) = \chi'(G) = \Delta(G)$$

for every bipartite graph G.

It is easily seen that the result above holds for bipartite multigraphs as well (see Exercise 52); indeed, all one has to recall is that every bipartite multigraph contains a stable matching.

We know that, in general, $\chi_\ell(G) \neq \chi(G)$ even for planar graphs, although we do have equality for the line graphs of *bipartite* graphs. Recall that the *line graph* of a graph $G = (V, E)$ is $L(G) = (E, F)$, where $F = \{ef : e, f \in E, e$ and f are adjacent$\}$. Indeed, it is conjectured that we have equality for all line graphs, in other words, $\chi'_\ell(G) = \chi'(G)$ for *all* graphs. Trivially,

$$\chi'_\ell(G) = \chi_\ell(L(G)) \leq \Delta((L(G)) + 1 \leq 2\Delta(G) - 1,$$

but it is not even easy to see that

$$\chi'_\ell(G) \leq (2 - 10^{-10})\Delta(G)$$

if $\Delta(G)$ is large enough. In fact, in 1996 Kahn proved that if $\varepsilon > 0$ and $\Delta(G)$ is large enough then

$$\chi'_\ell(G) \leq (1 + \varepsilon)\Delta(G).$$

Even after these beautiful results of Galvin and Kahn, we seem to be far from a proof of the full conjecture that $\chi'_\ell(G) = \chi'(G)$ for every graph.

V.5 Perfect Graphs

In the introduction to this chapter we remarked that perhaps the simplest reason why the chromatic number of a graph G is at least k is that G contains a k-clique, a complete graph of order k. The observation gave us the trivial inequality (1), namely that $\chi(G)$ is at least as large as the clique number $\omega(G)$, the maximal order of a complete subgraph of G.

The chromatic number $\chi(G)$ can be considerably larger than $\omega(G)$; in fact, we shall see in Chapter VII that, for all k and g, there is a graph of chromatic number at least k and girth at least g. However, here we shall be concerned with graphs at the other end of the spectrum: with graphs all whose induced subgraphs have their chromatic number equal to their clique number. These are the so-called *perfect* graphs. Thus a graph G is perfect if $\chi(H) = \omega(H)$ for every induced subgraph H of G, including G itself. Clearly, bipartite graphs are perfect, but a triangle-free graph containing an odd cycle is not perfect since its clique number is 2 and its chromatic number is at least 3. It is less immediate that the complement of a bipartite graph is also perfect. This is perhaps the first result on perfect graphs, proved by Gallai and König in 1932, although the concept of a perfect graph was only explicitly defined by Berge in 1960. Recall that the complement of a graph $G = (V, E)$ is $\overline{G} = (V, V^{(2)} - E)$. Although $\omega(\overline{G})$ is $\alpha(G)$, the independence number of G, in order to have fewer functions, we shall use $\omega(\overline{G})$ rather than $\alpha(G)$.

Theorem 15 *The complement of a bipartite graph is perfect.*

Proof. Since an induced subgraph of the complement of a bipartite graph is also the complement of a bipartite graph, all we have to prove is that if $G = (V, E)$ is a bipartite graph then $\chi(\overline{G}) = \omega(\overline{G})$.

Now, in a colouring of $\overline{G}$, every colour class is either a vertex or a pair of vertices adjacent in G. Thus $\chi(\overline{G})$ is the minimal number of vertices and edges of G, covering all vertices of G. By Corollary III.10, this is precisely the maximal number of independent vertices in G, that is, the clique number $\omega(\overline{G})$ of $\overline{G}$. □

For our next examples of perfect graphs, we shall take line graphs and their complements.

Theorem 16 *Let G be a bipartite graph with line graph $H = L(G)$. Then H and $\overline{H}$ are perfect.*

Proof. Once again, all we have to prove is that $\chi(H) = \omega(H)$ and $\chi(\overline{H}) = \omega(\overline{H})$.

Clearly, $\omega(H) = \Delta(G)$ and $\chi(H) = \chi'(G)$. But as G is bipartite, $\chi'(G) = \Delta(G)$ (see the beginning of Section 2), so $\chi(H) = \Delta(G) = \omega(H)$.

And what is $\chi(\overline{H})$? The minimal number of vertices of G covering all the edges. Finally, what is $\omega(\overline{H})$? The maximal number of independent edges of G. By Corollary III.10, these two quantities are equal. □

Yet another class of perfect graphs can be obtained from partially ordered sets. Given a partially ordered set $P = (X, <)$, its *comparability graph* is $C(P) = (X, E)$, where $E = \{xy \in X^{(2)} : x < y \text{ or } y < x\}$.

Theorem 17 *Comparability graphs and their complements are perfect.*

Proof. Once again, it suffices to show that if P is a partially ordered set then for $H = C(P)$ we have $\chi(H) = \omega(H)$ and $\chi(\overline{H}) = \omega(\overline{H})$.

To see the first equality, for $x \in P$ let $r(x)$, the *rank* of x, be the maximal integer r for which P contains a chain of r elements, with maximal element x. Then for $k = \max_r r(x)$ the map $r : P \to [k]$ gives a k-colouring of H, and a chain of size k gives a k-clique.

The second equality is deeper. Indeed, $\chi(\overline{H})$ is the minimal number of chains into which P can be partitioned, and $\omega(\overline{H})$ is precisely the maximal number of elements in an antichain. Therefore the equality $\chi(\overline{H}) = \omega(\overline{H})$ is none other than Dilworth's theorem, Theorem III.12. □

It does not take much to notice that, in all the examples above, the complement of a perfect graph is also perfect. In fact, the cornerstone of the theory of perfect graphs, the *perfect graph theorem*, claims that this holds without exception, not only for the examples above. This fundamental result was proved by Lovász and Fulkerson in the early 1970s; although the proof below is relatively simple, it needs a little preparation.

Lemma 18 *A necessary and sufficient condition for a graph G to be perfect is that for every induced subgraph $H \subset G$ there is an independent set of vertices, I, such that*

$$\omega(H - I) < \omega(H).$$

That is, a graph is perfect iff every induced subgraph H has an independent set meeting every clique of H of maximal order $\omega(H)$.

Proof. The necessity holds with plenty to spare. Indeed, let H be a graph with $k = \chi(H) = \omega(H)$, and let I be a colour class of a k-colouring of H. Then $\omega(H - I) \leq \chi(H - I) = \chi(H) - 1 < \omega(H)$.

The sufficiency of the condition will be proved by induction on $\omega(G)$. For $\omega(G) = 1$ there is nothing to prove, so suppose that $\omega(G) > 1$ and the assertion holds for smaller values of the clique number. Let H be an induced subgraph of G and I an independent set with $\omega(H - I) < \omega(H)$. By the induction hypothesis, we can colour $H - I$ with $\omega(H - I)$ colours; colouring the vertices of I with a new colour, we obtain a colouring of H with $\omega(H - I) + 1 \leq \omega(H)$ colours. Thus $\chi(H) \leq \omega(H)$, and we are done. □

The next result needed in the proof of the perfect graph theorem we shall give is of interest in its own right, as it enables one to construct large families of perfect graphs. In order to state it, we need the notion of *substitution.*

Let G be a graph with vertex-set $V(G) = [n] = \{1, \ldots, n\}$, and let $G_1, \ldots, G_n$ be vertex-disjoint graphs. Let $G^* = G[G_1, \ldots, G_n]$ be obtained from $\bigcup_{i=1}^{n} G_i$ by joining all vertices of G_i to all vertices of G_j whenever $ij \in E(G)$. We say that G^* is obtained from G by *substituting* $G_1, \ldots, G_n$ for the vertices or by *replacing* the vertices of G by $G_1, \ldots, G_n$. Note that if we replace the vertices of G one by one with the graphs $G_1, \ldots, G_n$, we get the same graph G^*.

We are ready to state the *replacement theorem* for perfect graphs.

Theorem 19 *A graph obtained from a perfect graph by replacing its vertices by perfect graphs is perfect.*

Proof. As we may replace the vertices one by one, it suffices to prove that if a vertex x of a perfect graph G is replaced by a perfect graph G_x then the resulting graph G^* is perfect. Furthermore, since every induced subgraph of G^* is of precisely the same form (obtained from a perfect graph by replacing one of its vertices by a perfect graph), by Lemma 18 it suffices to show that G^* itself contains an independent set of vertices meeting every clique of G^* with $\omega(G^*)$ vertices.

Having identified our task, let us get on with the job. By Lemma 18, the graph G_x has an independent set I such that $\omega(G_x - I) < \omega(G_x)$. Colour G with $\omega(G)$ colours, and let W_x be the colour class containing x. Then $J = I \cup (W_x - x)$ is an independent set in G^*. We claim this set J will do for the independent set. Let K be a clique of G^* with $\omega(G^*)$ vertices, and let us show that J meets K.

Note that either K is a clique in $G - x$, or it is the union of a clique of G_x of order $\omega(G_x)$ and a clique of $G[\Gamma(x)]$. Now, if K is a clique in $G - x$ then, as it has $\omega(G*) \geq \omega(G)$ vertices, it meets every colour class of G in our $\omega(G)$-colouring, including W_x, so $K \cap J = K \cap W_x \neq \emptyset$. On the other hand, if K meets G_x then K meets I, as the part of K in G_x is an $\omega(G_x)$-clique of G_x. Hence J does meet K as claimed. □

After all this preparation, we are ready to prove the perfect graph theorem of Lovász and Fulkerson.

Theorem 20 *The complement of a perfect graph is perfect.*

Proof. Let us prove the theorem by induction on the order n of our perfect graph. For $n = 1$ there is nothing to prove, so suppose that $n > 1$ and the theorem holds for perfect graphs of order less than n. In order to prove the induction step, by Lemma 18 all we need is that if G is a perfect graph of order n, then $\overline{G}$ contains an independent set I such that $\omega(G - I) < \omega(G)$. Translating this into an assertion about G, all we need is that G contains a complete graph K such that $\alpha(G - K) < \alpha(G)$.

Suppose then that this fails, that is, for every complete subgraph K of G, there is an independent set I_K with $\alpha(G)$ vertices that is disjoint from K. As we wish to count, let us put this slightly differently: if $K_1, K_2, \dots, K_t$ are all the complete subgraphs of G then, for every r, $1 \le r \le t$, there is an independent set I_r with $\alpha(G)$ vertices, which is disjoint from K_r.

For a vertex x of G, denote by $i(x)$ the number of independent sets I_r containing x. Let G^* be obtained from G by substituting a complete graph of order $i(x)$ for every vertex x. We know from the replacement theorem, Theorem 19, that G^* is perfect. But is it?

First, let us give an upper bound for the clique nummber $\omega(G^*)$. Every complete subgraph of G^* is obtained from a complete subgraph of G by substituting at most $i(x)$ vertices for each vertex x. Hence, there is an r, $1 \le r \le t$, such that

$$\omega(G^*) = \sum_{x \in K_r} i(x).$$

But

$$\sum_{x \in K_r} i(x) = \sum_{x \in K_r} \sum_{x \in I_s} 1 = \sum_{s=1}^{t} |K_r \cap I_s| \le t - 1,$$

since $|K_r \cap I_s| \le 1$ for all r and s, and $|K_r \cap I_r| = 0$. Therefore,

$$\omega(G^*) \le t - 1.$$

And what about $\chi(G^*)$? By the construction of G^*,

$$|G^*| = \sum_{x \in G} i(x) = \sum_{r=1}^{t} |I_r| = t\alpha(G),$$

and as G^* is obtained from G by substituting complete graphs for the vertices, $\alpha(G^*) = \alpha(G)$. Consequently,

$$\chi(G^*) \ge \frac{|G^*|}{\alpha(G^*)} = t.$$

Thus $\omega(G^*) < \chi(G^*)$, contradicting the fact that G^* is perfect, and so completing the proof of the theorem. □

There is another beautiful proof of the perfect graph theorem or, to be precise, of a slight extension of the perfect graph theorem, suggested by the trivial inequality (2). Indeed, if H is an induced subgraph of a perfect graph then, by (2),

$$\omega(H) = \chi(H) \geq |H|/\alpha(H) = |H|/\omega(\overline{H}),$$

so that

$$|H| \leq \omega(H)\omega(\overline{H}). \tag{14}$$

Hajnal and Simonovits conjectured that this trivial necessary condition for a graph to be perfect is also sufficient, namely that a graph is perfect if, and only if, (14) holds for every induced subgraph H. This conjecture was proved by Lovász in 1972, and in 1996 Gasparian found a shorter proof of it. Note that the perfect graph theorem is an immediate consequence of this result.

Let us turn to yet another characterization of perfect graphs, indicating the connection between perfect graphs and linear programming. First we need a variant of the independence number of a graph. Identifying a set with its characteristic function, an independent set of vertices of a graph G is naturally identified with a function $f : V(G) \to \{0, 1\}$ such that $\sum_{v \in K} f(v) \leq 1$ for every clique $K \subset G$. The clique number $\alpha(G)$ is the maximum of $\sum_{v \in K} f(v)$ over all such functions.

If we allow f to take any value between 0 and 1 (or just any non-negative value), then we get the *fractional independence number* $\alpha^*(G)$ of G:

$$\alpha^*(G) = \max \sum_{v \in G} f(v),$$

where the maximum is over all functions $f : V(G) \to [0, 1]$ such that $\sum_{v \in K} f(v) \leq 1$ for every clique $K \subset G$. Another beautiful result of Lovász is that a graph is perfect if, and only if, $\alpha^*(H) = \alpha(H)$ for every induced subgraph H.

Having seen several classes of perfect graphs, what about graphs that are not perfect? We noted earlier that every triangle-free non-bipartite graph is imperfect. But what about a characterization of perfect graphs in terms of forbidden induced subgraphs? As an induced subgraph of a perfect graph is perfect, it would suffice to characterize *critically imperfect graphs*, that is, imperfect graphs whose every induced proper subgraph is perfect. Examples of such graphs are the odd cycles of length at least 5 and, by the perfect graph theorem, the complements of these graphs.

Rather surprisingly, no other minimal examples are known. Indeed, the so called *perfect graph conjecture*, proposed by Berge in 1960, claims that these are the only examples: a graph G is perfect if, and only if, neither G nor its complement $\overline{G}$ contains an induced odd cycle of length at least 5. Equivalently, the odd cycles of length at least 5 and their complements are the only critically imperfect graphs.

Clearly, the perfect graph theorem would be an immediate consequence of the perfect graph conjecture. However, in spite of much effort, we do not seem to be close to a proof of this conjecture.

V.6 Exercises

1.$^-$ Show that a graph G has at least $\binom{\chi(G)}{2}$ edges.

2. For each $k \geq 3$ find a bipartite graph with vertices $x_1, x_2, \ldots, x_n$ for which the greedy algorithm uses k colours. Can this be done with $n = 2k-2$? Show that it cannot be done with $n = 2k-3$.

3.$^-$ Given a graph G, order its vertices in such a way that the greedy algorithm uses only $k = \chi(G)$ colours.

4.$^-$ Order the vertices of a graph G according to their degrees, so that $V(G) = \{x_1, x_2, \ldots, x_n\}$ and $d(x_1) \geq d(x_2) \geq \cdots$. Show that in this order the greedy algorithm uses at most $\max_i \min\{d(x_i)+1, i\}$ colours, and so if k is the maximal natural number for which $k \leq d(x_k)+1$ then $\chi(G) \leq k$.

5. Deduce from Exercise 4 that if G has n vertices then

$$\chi(G) + \chi(\overline{G}) \leq n+1.$$

6.$^-$ Show that $\chi(G) + \chi(\overline{G}) \geq 2\sqrt{n}$.

7. Let $G = (V, E)$ be a graph of maximal degree 3. Show that for some partition $V = V_1 \cup V_2$ both $G[V_1]$ and $G[V_2]$ consist of independent edges and vertices.

8. Let d, d_1 and d_2 be nonnegative integers with $d_1 + d_2 = d - 1$. Prove that if $\Delta(G) = d$ then the vertex set $V(G)$ of G can be partitioned into two classes, say $V(G) = V_1 \cup V_2$, such that the graphs $G_i = G[V_i]$ satisfy $\Delta(G_i) \leq d_i$, $i = 1, 2$. [*Hint.* Consider a partition $V(G) = V_1 \cup V_2$ for which $d_1 e(G_2) + d_2 e(G_1)$ is minimal.]

9. (Exercise 8 contd.) Let now $d, d_1, d_2, \ldots, d_r$ be nonnegative integers with $\sum_1^r (d_i + 1) = d + 1$. Prove that if $\Delta(G) = d$ then there is a partition $V(G) = \bigcup_1^r V_i$ such that the graphs $G_i = G[V_i]$ satisfy $\Delta(G_i) \leq d_i$, $i = 1, 2, \ldots, r$.

10. Given natural numbers r and t, $2r \leq t$, the *Kneser graph* $K_t^{(r)}$ is constructed as follows. Its vertex set is $T^{(r)}$, the set of r-element subsets of $T = \{1, 2, \ldots, t\}$, and two vertices are joined iff they are disjoint subsets of T. Fig. V.11 shows $K_5^{(2)}$, the so called *Petersen graph.* Prove that $\chi(K_t^{(r)}) \leq t - 2r + 2$, $\chi(K_5^{(2)}) = 3$ and $\chi(K_6^{(2)}) = 4$.

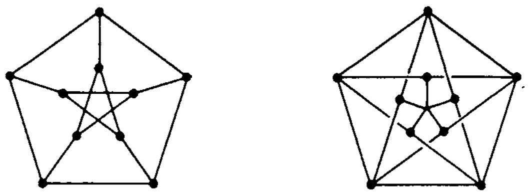


FIGURE V.11. The Petersen graph and the Grötzsch graph.

11. Check that the *Grötzsch graph*, shown in Fig. V.11, has girth 4 and chromatic number 4. Show that there is no graph of order 10 with girth at least 4 and chromatic number 4.

12.$^{++}$ Try to construct a triangle-free graph of chromatic number 1526 without looking at Chapters VI or VII.

13.$^{+}$ Show that there is a unique graph G_0 of order n and size $m = \lfloor n^2/4 \rfloor$ such that if G is also of order n and size m then

$$p_G(x) \leq p_{G_0}(x)$$

whenever x is sufficiently large.

14.$^{-}$ Find graphs G and H of order n and the same size such that $\chi(G) < \chi(H)$ but $p_G(x) < p_H(x)$ if x is sufficiently large.

15.$^{+}$ Given a connected graph G containing at least one cycle, define a graph H on the set S of all spanning trees of G by joining T_1 to T_2 iff $|E(T_1) \setminus E(T_2)| = 1$. (Cf. simple transforms of an x-path in Section IV. 3.) Imitate the proof of the fact that $p_H(x)$ is a polynomial (Theorem 4) and the proof suggested in Exercise 12 to show that H is not only Hamiltonian, but every edge of it is contained in a Hamilton cycle.

16. Let x be a vertex of a graph G and, for $r \geq 0$, let G_r be the subgraph of G induced by the vertices at distance r from x. (Thus G_r is the 'sphere' of radius r about x.) Show that $\chi(G)$ is at most $\chi(G_r) + \chi(G_{r+1})$ for some r.

17.$^{+}$ Recall from Chapter I that a graph G has a subgraph *contractible* to a graph H with vertex set $\{y_1, \ldots, y_k\}$ if G contains vertex disjoint connected subgraphs $G_1, \ldots, G_k$ such that, for $i \neq j$, there is an edge $y_iy_j \in E(H)$ iff G has a G_i–G_j edge; in notation, $G \succ H$ or $H \prec G$.
Prove that for every natural number p there is a minimal integer $c(p)$ such that every graph with chromatic number at least $c(p)$ has a subgraph contractible to K_p. By making use of the result in the previous exercise, show that $c(1) = 1$, $c(2) = 2$ and $c(n+1) \leq 2c(n) - 1$ for $n \geq 2$.

18.$^{+}$ *Hadwiger's conjecture* states that $c(p) = p$ for every p. Prove this for $p \leq 4$.

19. Can you show that for every $p \geq 1$ there is an integer $\delta(p)$ such that every graph of minimal degree at least $\delta(p)$ is contractible to K_p?

20. Let G be obtained from a 3-connected graph by adding to it a vertex x and 3 edges incident with x. Show that G is contractible to K_5^-, that is, to a complete graph of order 5 from which an edge has been deleted.

21. Prove that if $\chi(G) \geq 5$ then either $K_5 \prec G$ or $K_5^- \prec G - x$ for every $x \in V(G)$.

22. Show that the truth of Hadwiger's conjecture for $p = 5$ implies the four colour theorem.

23. Show that a planar map $M = M(G)$ can be 2-coloured iff every vertex of G has even degree. [*Hint.* If every vertex of G has even degree then G is a union of edge-disjoint cycles. For another solution, apply induction on the number of edges, and delete the edges of a cycle forming the boundary of a face of $M(G)$.]

24.$^-$ Let $M = M(G)$ be a triangular map, that is, a map in which every country has three sides. Show that M is 3-colourable unless $G = K_4$.

25.$^-$ Prove that a map $M = M(G)$ is 4-colourable if G has a Hamilton cycle.

26.$^-$ For each plane graph G construct a cubic plane graph H such that if $M(H)$ is 4-colourable then so is $M(G)$.

27.$^-$ According to Tait's conjecture every 3-connected cubic plane graph has a Hamilton cycle. (i) Show that Tait's conjecture implies the four colour theorem. (ii) By examining the graph in Fig. V.12 disprove Tait's conjecture.

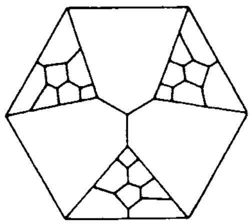


FIGURE V.12. Tutte's counterexample to Tait's conjecture.

28.$^+$ Let G be a cubic plane graph. Show that G is 3-edge-colourable iff $M(G)$ is 4-colourable. [*Hint.* Let $1, a, b$ and c be the elements of the Klein four-group $C_2 \times C_2$, so that $a^2 = b^2 = c^2 = 1$. Colour the edges with a, b and c, and the countries with 1, a, b and c.]

29. Find the edge chromatic number of K_n.

30.$^-$ Show that every cubic Hamiltonian graph has at least three Hamilton cycles.

31. Suppose the cubic graph G has exactly one edge-colouring with $\chi'(G)$ colours, up to a permutation of the colours. Show that $\chi'(G) = 3$ and that G has exactly 3 Hamilton cycles.

32.$^+$ Let $P_{n,k}$ be obtained from two vertex-disjoint n-cycles, $v_1 v_2 \cdots v_n$ and $w_1 w_2 \cdots w_n$, say, by joining v_i to w_{i+k}, with suffices computed modulo n. Show that $P_{9,2}$ is *uniquely 3-edge-colourable* (cf. Exercise 31); that is, up to a permutation of the colours it has a unique 3-edge-colouring. Show also that if $n \geq 2$ then $P_{6n+3,2}$ is not uniquely 3-edge-colourable, and it has exactly three Hamilton cycles.

33. Let $n = 2^p$. Show that K_{n+1} is not the union of p bipartite graphs but K_n is. Deduce that if there are $2^p + 1$ points in the plane then some three of them determine an angle of size at least $\pi(1 - (1/p))$.

34. Let $\chi(G) = k$. What is the minimal number of r-chromatic graphs whose union is G?

35.$^-$ Show that a k-chromatic graph can be oriented in such a way that a longest directed path has k vertices.

36. Prove the following theorem of Roy and Gallai. If a graph G can be oriented in such a way that no directed path contains more than k vertices then $\chi(G) \leq k$. [*Hint.* Omit a minimal set of edges to destroy all directed cycles. For a vertex x let $c(x)$ be the maximal number of vertices on a directed path in the new graph starting at x. Check that c is a proper colouring.]

37. Let G be a graph of maximal degree at most 2, without a triangle and without three independent edges, such that for any two vertices there is an edge incident with neither of them. Show that $G = C_5 \cup \overline{K_{n-5}}$.

38. A graph G is said to be *k-critical* if $\chi(G) = k$ and $\chi(H) < k$ for every proper subgraph H of G. Note that K_2 is the only 2-critical graph and the odd cycles are the only 3-critical graphs. Show that if $G \neq K_k$ is k-critical then $|G| \geq k + 2$. Deduce from the previous exercise that if G is a k-critical graph with $k + 2$ vertices then $k \geq 3$ and $G = C_5 + K_{k-3}$. In particular, $e(G) = \binom{k+2}{2} - 5$.

39. Let G_1 and G_2 be vertex disjoint graphs, containing edges $x_1y_1 \in E(G_1)$ and $x_2y_2 \in E(G_2)$. The *Hajós sum* $G = (G_1, x_1y_1) + (G_2, x_2y_2)$ of the pairs (G_1, x_1y_1) and (G_2, x_2y_2) is obtained from $G_1 \cup G_2$ by identifying x_1 and x_2, deleting the edges x_1y_1, x_2y_2, and adding the edge y_1y_2 (see Fig.V.13). Check that $\chi(G) \geq \min\{\chi(G_1), \chi(G_2)\}$. [In fact, Hajós proved in 1961 that $\{G : \chi(G) \geq k\}$ is precisely the smallest class of graphs containing K_k that is closed under Hajós sums and the trivial operations of adding edges and identifying non-adjacent vertices.]

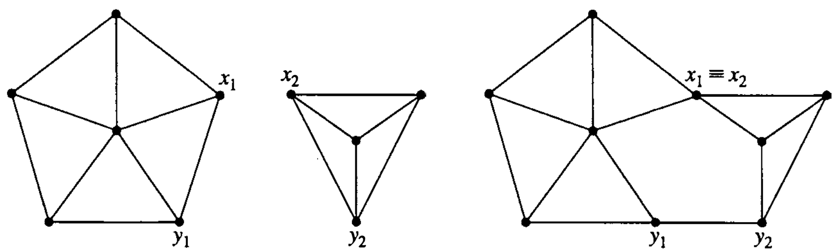


FIGURE V.13. The Hajós sum $(G_1, x_1y_1) + (G_2, x_2y_2)$ of a wheel and a complete graph.

40.$^+$ (Exercise 39 contd.) Let $\mathcal{H}_k$ be the smallest collection of (isomorphism classes of) graphs such that (1) $K_k \in \mathcal{H}_k$, (2) if $H \in \mathcal{H}_k$ and $G \supset H$ then $G \in \mathcal{H}_k$,

(3) if $H \in \mathcal{H}_k$ and G is obtained from H by identifying two nonadjacent vertices, then $G \in \mathcal{H}_k$, (4) if $G_1, G_2 \in \mathcal{H}_k$ and G is the Hajós sum of G_1 and G_2 then $G \in \mathcal{H}_k$. Prove that $\mathcal{H}_k$ is precisely the class $\mathcal{G}_k$ of graphs of chromatic number at least k. [*Hint.* The result in Exercise 39 implies that $\mathcal{H}_k \subset \mathcal{G}_k$. Assume that the converse inclusion is false and let $G \in \mathcal{G}_k \setminus \mathcal{H}_k$ be a counterexample of minimal order and maximal size. Then G cannot be a complete q-partite graph, so it contains vertices a, b_1 and b_2 such that $b_1b_2 \in (G)$ but $ab_1, ab_2 \in E(G)$. Let $G_1 = G + ab_1$ and $G_2 = G + ab_2$. Then G_1 and G_2 are not counterexamples so belong to $\mathcal{H}_k$. Find out how G can be obtained from copies of G_1 and G_2 by the allowed operations.]

41. (Exercises 39 and 40 contd.) Show that, for $k \geq 3$, the Hajós sum of two k-critical graphs is again k-critical.

42. Show that, for $k \geq 3$ and $\ell \geq 1$, there is a k-critical graph of order $n = (k-1)\ell + 1$ and size $\ell\binom{k}{2} - 1$.

43. Show that a 4-critical graph with 7 vertices has at least 11 edges, and this bound is best possible.

44. Let k be a natural number. Prove that an infinite graph is k-colourable iff every finite subgraph of it is. [*Hint.* Apply Tychonov's theorem as in Exercise III.31.]

45. Check that the chromatic polynomial of a tree T of order n is
$$p_T(x) = x(x-1)^{n-1}.$$
Deduce that the chromatic polynomial of a forest F of order n and size m is
$$p_F(x) = x^{n-m}(x-1)^m.$$
Use Corollary 6 to deduce the same assertion.

46. Let e be a bridge of a graph G. Show that $p_G(x) = \frac{x-1}{x} p_{G-e}(x)$.

47. Let G be a connected graph with blocks $B_1, B_2, \ldots, B_\ell$. Show that
$$p_G(x) = x^{-\ell+1} \prod_{i=1}^{\ell} p_{B_i}(x).$$

48. Let $G = G_1 \cup G_2$, with $H = G_1 \cap G_2$ being a complete graph K^r. Show that
$$p_G(x) = \frac{p_{G_1}(x)p_{G_2}(x)}{(x)_r} = \frac{p_{G_1}(x)p_{G_2}(x)}{p_H(x)}.$$

49. Show that if G is a connected graph of order n then $(-1)^{n-1}p_G(x) > 0$ for all x, $0 < x < 1$.

50. Show that $|p_G(-1)|$ is the number of acyclic orientations of G.

51. Let us assign a list $L(x)$ of two colours to every vertex x of an odd cycle. Show that there is an L-colouring unless we assign the same set to every vertex.

52. Check that Theorem 14 holds for bipartite multigraphs as well.

53. A graph is said to be *triangulated* if every cycle of length at least 4 has a diagonal, that is, if the graph contains no induced cycle of length at least 4. Show that a connected graph G is triangulated iff whenever S is a minimal set of vertices such that $G \setminus S$ is disconnected then $G[S]$ is complete.

54. A vertex whose neighbours induce a complete graph is said to be *simplicial*. Show that every non-empty triangulated graph has at least two simplicial vertices. Deduce that a graph G is triangulated iff its vertices have an enumeration $x_1, x_2, \ldots, x_n$ such that each x_k is a simplicial vertex of $G[\{x_1, \ldots, x_k\}]$.

55.$^+$ An *interval graph* has vertex set $\{I_1, \ldots, I_n\}$, where each I_j is an interval $[a_j, b_j] \subset \mathbb{R}$, and two intervals I_j and I_h are adjacent if they meet. Show that every interval graph is triangulated, and its complement is a comparability graph.
Without making use of the perfect graph theorem, show that interval graphs and their complements are perfect.

56. Given a permutation π of $[n] = \{1, 2, \ldots, n\}$, the *permutation graph* $G(\pi)$ has vertex set $[n]$, with ij an edge if π switches the order of i and j. Thus, for $i < j$, we join i to j iff $\pi(j) < \pi(i)$. Without making use of the perfect graph theorem, show that permutation graphs are perfect.

57.$^{++}$ To appreciate the depth of Theorem 14, try to give a direct proof of the assertion that the list-chromatic index of the complete k by k bipartite graph is k. If you fail (and it would be a wonderful achievement if you did not), try to prove it for $k = 2$, 3 and 4.

58. Grünbaum conjectured in 1970 that for all $k \geq 2$ and $g \geq 3$ there are (k, k, g)-*graphs*, that is, k-chromatic k-regular graphs of girth at least g. Show that the graph in Fig.V.14, constructed by Brinkman, is a (4, 4, 5)-graph.

59. Fill in the details in the proof of Theorem 10.

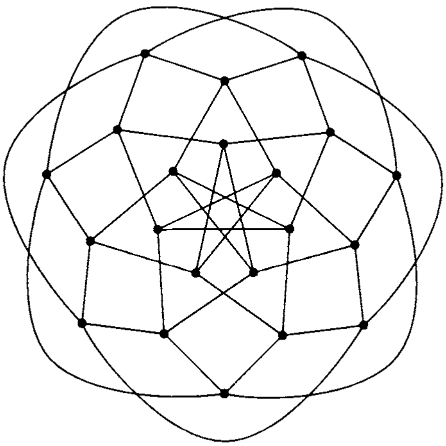


FIGURE V.14. The Brinkman graph.

60. Let $n \geq 4$. Show that if K_n triangulates a closed surface then $n \not\equiv 2 \pmod 3$. Deduce the converse from the Ringel–Youngs theorem, namely that $s(M) - h(\chi) = \lfloor (7 + \sqrt{49 - 24\chi})/2 \rfloor$ for every closed surface M of Euler characteristic $\chi < 0$.

61. Let G be a graph of order $2n$ such that for every $S \subset V(G)$, the graph $G - S$ has at most $|S|$ odd components. Show that $\chi(\overline{G}) \leq n$, and we can have equality for every $n \geq 1$.

62. Check from first principles that the complement of an odd cycle of length at least 5 is imperfect.

63.$^+$ For $\alpha, \omega \geq 2$, call a graph G an *(α, ω)-graph* if it has $\alpha\omega + 1$ vertices and for every $v \in G$ the graph $G - v$ can be partitioned into α cliques, each of order ω, and also into ω independent sets, each with α vertices. Recalling that a graph G is perfect iff every induced subgraph $H \subset G$ satisfies $|H| \leq \alpha(H)\omega(H)$, show that every critically imperfect graph is an (α, ω)-graph for some α, $\omega \geq 2$.

64.$^+$ Let $\alpha, \omega \geq 2$ be integers, and let G be the $(\omega - 1)$st power of an $(\alpha\omega + 1)$-cycle $C_{\alpha\omega+1}$. Thus $V(G) = \mathbb{Z}_{\alpha\omega+1}$ and $ij \in E(G)$ if

$$E(G) = \{ij : i - j = \pm 1, \pm 2, \ldots, \pm(\omega - 1).$$

Show that G is an (α, ω)-graph. Is G critically imperfect?

65. For $k \geq 1$, let G_k be the graph with vertex set

$$V = [2^k + 1]^{(2)} = \{1, 2, \ldots, 2^k + 1\}^{(2)}$$

in which $\{a, b\} \in V$ is joined to $\{b, c\} \in V$ whenever $a < b < c$. Thus G_k has $\binom{2^k+1}{2}$ vertices and $\sum_{\ell=0}^{2^k} \ell(2^k - \ell)$ edges. Prove that G_k is triangle-free and $\chi(G_k) = k + 1$.

66. In 1947, Tutte constructed a sequence $G_3, G_4, \ldots$ of triangle-free graphs as follows. Let G_3 be an odd cycle with at least 5 vertices. Having constructed G_k with n_k vertices, set $m_k = k(n_k - 1) + 1$ and $n_{k+1} = \binom{m_k}{n_k} n_k + m_k$. Let W be a set of m_k vertices, and for each $\alpha \in W^{(n_k)}$, i.e. each n_k-subset α of W, let G_α be a copy of G_k, with the sets W and $V(G_\alpha)$, $\alpha \in W^{(n_k)}$, all disjoint. Let G_{k+1} be obtained from $\bigcup_\alpha G_\alpha \cup W$ by adding, for each α, a complete matching from α to $V(G_\alpha)$. Thus $|G_{k+1}| = n_{k+1}$. Show that each G_k is triangle-free and $\chi(G_k) = k$.

67.$^+$ Let G be the *infinite* graph whose vertex set is $\mathbb{R}^2$ and in which two points are joined if their distance is 1. Show that $4 \leq \chi(G) \leq 7$.

68.$^+$ Show that the chromatic number of a triangle-free graph drawn on a surface of Euler characteristic $E \leq 0$ is at most $(5 + \sqrt{25 - 16E})/2$.

69. Let G be a triangle-free graph with vertex set $\{x_1, x_2, \dots, x_n\}$. Construct a graph G' from G by adding to it $n + 1$ vertices, $x'_1, x'_2, \dots, x'_n$ and y, and joining each x'_i to the vertices in $\Gamma_G(x_i) \cup \{y\}$. (Thus x'_i 'duplicates' x_i, and y is joined to the other new vertices.) Show that G' is triangle-free and $\chi(G') = \chi(G)$. Use this construction to exhibit triangle-free graphs G_3, G_4 and G_5, with $\chi(G_k) = k$.

70. Let G be the graph of order $2n + 1 \geq 5$ obtained from $K_{n,n}$ by subdividing an edge by a vertex. Show that $\chi'(G) = \Delta(G) + 1 = n + 1$, but $\chi'(G - e) = \Delta(G - e) = n$ for every edge e of G.

71. Show that there is no plane graph G such that
(1) every face other than $x_1 x_2 \cdots x_\ell$ is a triangle,
(2) all degrees not on this face are even, and
(3) all degrees $d(x_1), \dots, d(x_{m-1})$ are odd, where $m = \lfloor \ell/2 \rfloor$, and $d(x_m)$ is odd iff ℓ is odd.

72. Let G be a triangulation of the plane, with all degrees even. Show that $\chi(G) = 3$. [*Hint.* Pick a vertex x_1. Let V_1 be the smallest set of vertices such that (1) $x_1 \in V_1$, and (2) if $x \in V_1$ and both xyz and $x'yz$ are faces then $x' \in V_1$. Use the result in Exercise 71 to check that V_1 is an independent set. Clearly, $G - V_1$ is a collection of even cycles.]

73. Let G be a cubic plane graph. Prove that the map $M(G)$ is 3-colourable iff each country has an even number of sides.

74. Show that the only vertex-critical 3-chromatic graphs are the odd cycles: if $\chi(G) = 3$ and $\chi(G - x) = 2$ for every vertex x then $G = C_{2k+1}$ for some $k \geq 1$.

75. Let G be the graph on $\mathbb{Z}_{17}$ with i joined to j iff $i - j \in \{\pm 2, \pm 6, \pm 7, \pm 8\}$. Show that G is a vertex-critical 5-chromatic graph with a critical edge: $\chi(G) = 5$, $\chi(G - x) = 4$ for every vertex x, and $\chi(G - xy) = 5$ for every edge xy.

76.$^+$ Prove that the chromatic number of a triangle-free graph of size m is at most $2m^{1/3} + 1$. [*Hint.* Apply induction on m, making use of Exercise 68 of Chapter IV.]

V.7 Notes

Theorem 3 is in R.L. Brooks, On colouring the nodes of a network, *Proc. Cambridge Phil. Soc.* **37** (1941) 194–197, and Vizing's theorem, Theorem 5, is in V.G. Vizing, On an estimate of the chromatic class of a p-graph (in Russian), *Diskret. Analiz* **3** (1964) 23–30. A detailed account of results concerning colouring graphs on surfaces, culminating in the proof of Heawood's conjecture by Ringel and Youngs' can be found in *Map Color Theorem*, Grundlehren der math. Wiss. 209, Springer-Verlag, Berlin, 1974. The use of the discharging procedure in attacking the four colour problem is described in H. Heesch, Untersuchungen zum Vierfarbenproblem, B-I-Hochschulskripten 810/810a/810b, Bibliographisches Institut, Mannheim, Vienna, Zürich, 1969. The first proof of the four colour theorem is in K. Appel and W. Haken, Every planar map is four colourable, Part I: discharging, *Illinois J. of Math.* **21** (1977) 429–490 and K. Appel, W. Haken and J. Koch, Every planar map is four colourable, Part II: reducibility, *Illinois J. of Math.* **21** (1977) 491–567. A history of the four colour problem and a digest of its proof are provided by T.L. Saaty and P.C. Kainen, *The Four Color Problem, Assaults and Conquest*, McGraw-Hill, New York, 1977.

The recent proof mentioned in the text is in N. Robertson, D. Sanders, P.D. Seymour and R. Thomas, The four-colour theorem, *J. Combinatorial Theory, Ser. B* **70** (1997) 2–44.

The theorems of Thomassen and Galvin were published in C. Thomassen, Every planar graph is 5-choosable, *J. Combinatorial Theory*, Series B **62** (1994) 180–181, and F. Galvin, The list chromatic index of a bipartite multigraph, *J. Combinatorial Theory*, Series B **63** (1995) 153–158. Our presentation of Galvin's theorem is based on T. Slivnik, A short proof of Galvin's theorem on the list-chromatic index of a bipartite multigraph, *Combinatorics, Probability and Computing* **5** (1996) 91–94. Voigt's construction of a planar graph of list-chromatic number 5 is in M. Voigt, List colourings of planar graphs, *Discrete Mathematics* **120** (1993) 215–219. Kahn's breakthrough in the list-colouring problem, briefly mentioned at the end of Section 4, is in J. Kahn, Asymptotically good list-colourings, *J. Combinatorial Theory* A **73** (1996) 1–59.

The perfect graph theorem was proved in L. Lovász, Normal hypergraphs and the perfect graph conjecture, *Discrete Math.* **2** (1972) 253–267, and D.R. Fulkerson, Blocking and anti-blocking pairs of polyhedra, *Math. Programming* **1** (1971) 168–194; for an excellent review of the theory and application of perfect graphs see L. Lovász, Perfect graphs, in *Selected Topics in Graph Theory 2* (L.W. Beineke and R.J. Wilson, eds), Academic Press, London, 1983, pp. 55–87. The simple proof of the sufficiency of condition (13) we mentioned in the text is in G. S. Gasparian, Minimal imperfect graphs: a simple approach, *Combinatorica* **16** (1996) 209–212.

The partition results of Exercises 8 and 9 are in L. Lovász, On decomposition of graphs, *Studia Sci. Math., Hungar.* **1** (1966) 237–238.

The result in Exercise 36 is due to B. Roy,, Nombre chromatique et plus longs chemins d'un graphe, *Rev. AFIRO* **1** (1967) 127–132, and T. Gallai, On directed paths and circuits, in *Theory of Graphs* (P. Erdős and G. Katona, eds), Academic Press, New York, 1968, 115–118, and the results in Exercises 39–41 are from G. Hajós, Über eine Konstruktion nicht n-färbbarer Graphen, *Wiss. Zeitschr. Martin Luther Univ. Halle-Wittenberg,* Math.-Natur. Reihe **10** (1961) 116–117.

Colouring is a naturally appealing part of graph theory, and the subject has a vast literature. Many of the fundamental results are due to G.A. Dirac; for these and other results see Chapter V of B. Bollobás, *Extremal Graph Theory*, Academic Press, London and New York, 1978.

An excellent comprehensive book on colourings is T.R. Jensen and B. Toft, *Graph Coloring Problems*, Wiley-Interscience, New York, 1995, xix+295 pp., and a relevant recent review article is B. Toft, Colouring, stable sets and perfect graphs, in *Handbook of Combinatorics* vol. I (R.L. Graham, M. Grötschel and L. Lovász, eds), North-Holland, Amsterdam, 1995, pp. 233–288.

VI
Ramsey Theory

In a party of six people there is always a group of three who either all know each other or are all strangers to each other. If the edges of the complete graph on an infinite set N are coloured red or blue then for some infinite set $M \subset N$ all the edges joining vertices of M get the same colour. Both of these assertions are special cases of a theorem published by Ramsey in 1930. The original theorems of Ramsey have been extended in many directions, resulting in what has come to be called *Ramsey theory*: a rich theory expressing the deep mathematical principle, vastly extending the pigeon-hole principle, that no matter how we partition the objects of a 'large' structure into a 'few' classes, one of these classes contains a 'large' subsystem. While Dirichlet's pigeon-hole principle guarantees that we have 'many' objects in the same class, without any condition on their relationship to each other, in Ramsey theory we look for a large substructure in the same class: we do not only want infinitely many red edges, say, but we want all the edges joining vertices of an infinite set to be red. Or, in the first example, we do not only want three pairs of acquaintances, but we want these three acquaintances to 'form a triangle', to be the three pairs of acquaintances belonging to three people.

The quintessential result of Ramsey theory dealing with richer mathematical structures than graphs is van der Waerden's theorem, predating the theorems of Ramsey, which states that given k and p, if W is a large enough integer and we partition the set of the first W natural numbers into k classes, then one of the classes contains an arithmetic progression with p terms.

Ramsey theory is a large and beautiful area of combinatorics, in which a great variety of techniques are used from many branches of mathematics, and whose results are important not only in graph theory and combinatorics, but in set theory, logic, analysis, algebra, and geometry as well. In order to demonstrate this, we shall go well beyond graph theory to present several striking and deep results, including the Erdős–Rado canonical theorem, extending Ramsey's original theorem to infinitely many colours; Shelah's theorem, extending the Hales-Jewett theorem (which itself extends van der Waerden's theorem); and the theorems of Galvin, Prikry, and Hindman about Ramsey properties of infinite sequences. Nevertheless, we shall hardly do more than scratch the surface of modern Ramsey theory.

VI.1 The Fundamental Ramsey Theorems

We shall consider partitions of the *edges* of graphs and hypergraphs. For the sake of convenience a partition will be called a *colouring*, but one should bear in mind that a colouring in this sense has nothing to do with the edge colourings considered in Chapter V. Adjacent edges may have the same colour and, indeed, our aim is to show that there are large subgraphs all of whose edges have the same colour. In a 2-colouring we shall often choose red and blue as colours; a subgraph is *red* (*blue*) if all its edges are red (blue).

As we shall see, given a natural number s, there is an integer $R(s)$ such that if $n \geq R(s)$ then every colouring of the edges of K_n with red and blue contains either a red K_s or a blue K_s. The assertion about a party of six people claims precisely that $R(3) = 6$ will do. In order to show the existence of $R(s)$ in general, for any s and t, we define the *Ramsey number* $R(s, t)$ as the smallest value of n for which every red–blue colouring of K_n yields a red K_s or a blue K_t. In particular, $R(s, t) = \infty$ if there is no such n such that in every red–blue colouring of K_n there is a red K_s or a blue K_t. It is obvious that

$$R(s, t) = R(t, s)$$

for every $s, t \geq 2$ and

$$R(s, 2) = R(2, s) = s,$$

since in a red–blue colouring of K_2 either there is a blue edge or else every edge is red. The following result, due to Erdős and Szekeres, states that $R(s, t)$ is finite for every s and t, and at the same time it gives a bound on $R(s, t)$. Although qualitatively it is a special case of Ramsey's original theorem, the bound it gives is considerably better than that given by Ramsey.

Theorem 1 *The function $R(s,t)$ is finite for all $s, t \geq 2$. If $s > 2$ and $t > 2$ then*

$$R(s, t) \leq R(s-1, t) + R(s, t-1) \tag{1}$$

and

$$R(s, t) \leq \binom{s+t-2}{s-1}. \tag{2}$$

Proof. As we shall prove (1) and (2), it will follow that $R(s, t)$ is finite.

(i) When proving (1) we may assume that $R(s-1, t)$ and $R(s, t-1)$ are finite. Let $n = R(s-1, t) + R(s, t-1)$ and consider a colouring of the edges of K_n

with red and blue. We have to show that in this colouring there is either a red K_s or a blue K_t. To this end, let x be a vertex of K_n. Since $d(x) = n - 1 = R(s-1, t) + R(s, t-1) - 1$, either there are at least $n_1 = R(s-1, t)$ red edges incident with x or there are at least $n_2 = R(s, t-1)$ blue edges incident with x. By symmetry we may assume that the first case holds. Consider a subgraph K_{n_1} of K_n spanned by n_1 vertices joined to x by red edges. If K_{n_1} has a blue K_t, we are done. Otherwise, by the definition of $R(s-1, t)$, the graph K_{n_1} contains a red K_{s-1} which forms a red K_s with x.

(ii) Inequality (2) holds if $s = 2$ or $t = 2$ (in fact, we have equality since $R(s, 2) = R(2, s) = s$). Assume now that $s > 2$, $t > 2$ and (2) holds for every pair (s', t') with $2 \le s' + t' < s + t$. Then by (1) we have

$$\begin{aligned} R(s, t) &\le R(s-1, t) + R(s, t-1) \\ &\le \binom{s+t-3}{s-2} + \binom{s+t-3}{s-1} = \binom{s+t-2}{s-1}. \end{aligned} \qquad \square$$

It is customary to distinguish *diagonal Ramsey numbers* $R(s) = R(s, s)$ and *off-diagonal Ramsey numbers* $R(s, t)$, $s \ne t$. It is not surprising that the diagonal Ramsey numbers are of greatest interest, and they are also the hardest to estimate. Re calling that a graph is *trivial* if it is either complete or empty, the diagonal Ramsey number $R(s)$ is the minimal integer n such that every graph of order n has a trivial subgraph of order s.

We see from Theorem 1 that

$$R(s) \le \binom{2s-2}{s-1} \le \frac{2^{2s-2}}{\sqrt{s}}. \tag{3}$$

Although the proof above is very simple, the bound (3) was hardly improved for over 50 years. The best improvement is due to Thomason, who in 1988 proved that

$$R(s) \le \frac{2^{2s}}{s} \tag{4}$$

if s is large. Although the improvement over (3) is small, this is a hard result, and we shall not prove it. In Chapter VII we shall show that $R(s)$ does grow exponentially: $R(s) \ge 2^{s/2}$. It is widely believed that there is a constant c, perhaps even $c = 1$, such that

$$R(s) = 2^{(c+o(1))s},$$

but this is very far from being proved.

The result easily extends to colourings with any finite number of colours: given k and $s_1, s_2, \dots, s_k$, if n is sufficiently large, then every colouring of K_n with k colours is such that for some i, $1 \le i \le k$, there is a K_{s_i} coloured with the i th colour. (The minimal value of n for which this holds is usually denoted by $R_k(s_1, \dots, s_k)$.) Indeed, if we know this for $k-1$ colours, then in a k-colouring of K_n we replace the first two colours by a new colour. If n is sufficiently large (depending on $s_1, s_2, \dots, s_k$) then either there is a K_{s_i} coloured with the i th colour for some i, $3 \le i \le k$, or else for $m = R(s_1, s_2)$ there is a K_m coloured with the new colour. In other words, in the original colouring this K_m is coloured with the first two (original) colours. In the first case we are home, and in the second, for $i = 1$ or 2 we can find a K_{s_i} in K_m coloured with the i th colour. This shows that

$$R_k(s_1, \ldots, s_k) \le R_{k-1}(R(s_1, s_2), s_3, \ldots, s_k).$$

In fact, Theorem 1 also extends to hypergraphs, that is, to colourings of the set $X^{(r)}$ *of all r-tuples* of a finite set X with k colours. This is one of the theorems proved by Ramsey. We now turn our attention to this.

Denote by $R^{(r)}(s, t)$ the minimal value of n for which every red–blue colouring of $X^{(r)}$ yields a red s-set or a blue t-set, provided that $|X| = n$. Of course, a set $Y \subset X$ is called *red* (*blue*) if every element of $Y^{(r)}$ is red (blue). Note that $R(s, t) = R^{(2)}(s, t)$. As in the case of Theorem 1, the next result not only guarantees that $R^{(r)}(s, t)$ is finite for all values of the parameters (which is certainly not at all obvious at first), but also gives an upper bound on $R^{(r)}(s, t)$. The proof is an almost exact replica of the proof of Theorem 1. Note that if $r > \min\{s, t\}$ then $R^{(r)}(s, t) = \min\{s, t\}$, and if $r = s \le t$ then $R^{(r)}(s, t) = t$.

Theorem 2 *Let* $1 < r < \min\{s, t\}$. *Then* $R^{(r)}(s, t)$ *is finite and*

$$R^{(r)}(s, t) \le R^{(r-1)}\left(R^{(r)}(s-1, t), R^{(r)}(s, t-1)\right) + 1.$$

Proof. Both assertions follow immediately if we prove the inequality under the assumption that $R^{(r-1)}(u, v)$ is finite for all u, v, and both $R^{(r)}(s-1, t)$ and $R^{(r)}(s, t-1)$ are also finite.

Let X be a set with $R^{(r-1)}(R^{(r)}(s-1, t), R^{(r)}(s, t-1)) + 1$ elements. Given any red–blue colouring c of $X^{(r)}$, pick an $x \in X$ and define a red–blue colouring c' of the $(r-1)$-sets of $Y = X - \{x\}$ by colouring $\sigma \in Y^{(r-1)}$ the colour of $\sigma \cup \{x\} \in X^{(r)}$. By the definition of the function $R^{(r-1)}(u, v)$ we may assume that Y has a red subset Z (for c') with $R^{(r)}(s-1, t)$ elements.

Now let us look at the restriction of c to $Z^{(r)}$. If it has a blue t-set, we are done, since $Z^{(r)} \subset X^{(r)}$, so a blue t-set of Z is certainly also a blue t-set of X. On the other hand, if there is no blue t-set of Z then there is a red $(s-1)$-set. The union of this red $(s-1)$-set with $\{x\}$ is then a red s-set of X, because $\{x\} \cup \sigma$ is red for every $\sigma \in Z^{(r-1)}$. $\square$

It is easily seen that Theorem 2 and the colour-grouping argument described after Theorem 1 imply the following assertion. Given r and $s_1, s_2, \ldots, s_k$, then for large enough $|X|$ every colouring of $X^{(r)}$ with k colours is such that for some i, $1 \le i \le k$, there is a set $S_i \subset X$, $|S_i| = s_i$, all of whose r-sets have colour i. The smallest value of $|X|$ for which this is true is denoted by $R_k^{(r)}(s_1, s_2, \ldots, s_k)$; thus $R^{(r)}(s, t) = R_2^{(r)}(s, t)$ and $R_k(s_1, s_2, \ldots, s_k) = R_k^{(2)}(s_1, s_2, \ldots, s_k)$. The upper bound for $R_k^{(r)}(s_1, s_2, \ldots, s_k)$ implied (via colour-grouping) by Theorem 2 is not very good. Imitating the proof of Theorem 1 one arrives at a better upper bound (cf. Exercise 8):

$$R_k^{(r)}(s_1, s_2, \ldots, s_k) \leq R_k^{(r-1)}(R_k^{(r)}(s_1 - 1, s_2, \ldots, s_k), \ldots,$$
$$R_k^{(r)}(s_1, \ldots, s_{k-1}, s_k - 1)) + 1.$$

Very few of the nontrivial Ramsey numbers are known, even in the case $r = 2$. It is easily seen that $R(3, 3) = 6$, and with some work one can show that $R(3, 4) = 9$, $R(3, 5) = 14$, $R(3, 6) = 18$, $R(3, 7) = 23$ and $R(4, 4) = 18$. Considerably more effort is needed to prove that $R(3, 8) = 28$ and $R(3, 9) = 36$. Furthermore, McKay and Radziszowski proved in 1995 that $R(4, 5) = 25$. These are the only known two-colour Ramsey numbers. For the other ones, all that is known are bounds, as shown in Table VI. 1. The proofs of many of these bounds needed a surprising amount of ingenuity, work and computing time.

At first sight, the paucity of exact Ramsey numbers may well seem surprising. However, there are many reasons why it is unlikely that a large Ramsey number, like $R(6, 6)$, will ever be determined. The two-colourings of K_n without large monochromatic complete subgraphs lack order: they look as if they had been chosen at random. This apparent disorder makes it highly unlikely that a simple induction argument will give a tight *upper bound* for $R(s, t)$. On the other hand, a head-on attack by computers is also doomed to failure, even for $R(5, 5)$. For example, if all we want to prove is that 48 is an upper bound for $R(5, 5)$, we have to examine over 2^{1000} graphs of order 48: a task well beyond the power of computers.

It is not too easy to prove general *lower bounds* for Ramsey numbers either. As the colourings without large complete monochromatic subgraphs are 'disorderly', it is not surprising that random methods can be used to give fairly good lower bounds. In Chapter VII we shall show some beautiful examples of this.

As it is very difficult to find good estimates for $R(s, t)$ as $s, t \to \infty$, it is not surprising that very few fast-growing Ramsey functions have been determined exactly. In fact, Erdős and Szekeres proved that the right-hand side of (2) is exactly 1 smaller than the value of a natural Ramsey function. In order to present this result, we introduce some terminology. Call a set $S \subset \mathbb{R}^2$ *non-degenerate* if any two points of it have different x coordinates. A *k-cup*, or a *convex k-set*, is a non-degenerate set of k points of the form $\{(x_i, h(x_i)) : i = 1, \ldots, k\}$, where h is a convex function. Writing $s(p, p') = (y - y')/(x - x')$ for the *slope* of the line through the points $p = (x, y)$ and $p' = (x', y')$, if $K = \{p_1, \ldots, p_k\}$ with $p_i = (x_i, y_i), x_1 < \cdots < x_k$, then K is a k-cup iff $s(p_1, p_2) \leq s(p_2, p_3) \leq \cdots \leq s(p_{k-1}, p_k)$. An *$\ell$-cap*, or a *concave ℓ-set*, is defined analogously.

Here is then the beautiful result of Erdős and Szekeres about k-cups and ℓ-caps. The first part was published in 1935, the second in 1960.

Theorem 3 *For $k, \ell \geq 2$, every non-degenerate set of $\binom{k+\ell-4}{k-2} + 1$ points contains a k-cup or an ℓ-cap. Also, for all $k, \ell \geq 2$, there is a non-degenerate set $S_{k,\ell}$ of $\binom{k+\ell-4}{k-2}$ points that contains neither a k-cup nor an ℓ-cap.*

Proof. Let us write $\phi(k, \ell)$ for the binomial coefficient $\binom{k+\ell-4}{k-2}$.

(i) We shall prove by induction on $k + \ell$ that every non-degenerate set of $\phi(k, \ell) + 1$ points contains a k-cup or an ℓ-cap. Since a non-degenerate set of 2 points is both a 2-cup and a 2-cap, this is clear if $\min\{k, \ell\} = 2$, since $\phi(k, 2) = \phi(2, \ell) = 1$ for all $k, \ell \geq 2$. Suppose then that $k, \ell \geq 3$ and the assertion holds for smaller values of $k+\ell$. Let S be a non-degenerate set of $\phi(k, \ell)+1$ points and suppose that, contrary to the assertion, S contains neither a k-cup nor an ℓ-cap. Let $L \subset S$ be the set of last points of $(k-1)$-cups. Then $S \backslash L$ has neither a $(k-1)$-cup nor an ℓ-cap so, by the induction hypothesis, $|S\backslash L| \leq \phi(k-1, \ell)$. Therefore $|L| \geq \phi(k, \ell) + 1 - \phi(k-1, \ell) = \phi(k, \ell-1) + 1$ so, again by the induction hypothesis, L contains an $(\ell-1)$-cap, say $\{q_1, \ldots, q_{\ell-1}\}$, with first point our set S contains q_1. Since $q_1 \in L$, a $(k-1)$-cup $\{p_1, \ldots, p_{k-1}\}$, whose last point, p_{k-1}, is precisely q_1. Now, if $s(p_{k-2}, p_{k-1}) \leq s(p_{k-1}, q_2)$ then $\{p_1, \ldots, p_{k-1}, q_2\}$ is a k-cup. Otherwise, $s(p_{k-2}, q_1) > s(q_1, q_2)$, so $\{p_{k-2}, q_1, \ldots, q_{\ell-1}\}$ is an ℓ-cup. This contradiction completes the proof of the induction step, and we are done.

(ii) We shall construct $S_{k,\ell}$ also be induction on $k+\ell$. In fact, we shall construct $S_{k,\ell}$ in the form $\{(i, y_i) : 1 \leq i \leq \phi(k, \ell)\}$.

If $\min\{k, \ell\} = 2$ then $\phi(k, \ell) = 1$ and we may take $S_{k,\ell} = \{(1, 0)\}$. Suppose then that $k, \ell \geq 3$ and we have constructed $S_{k,\ell}$ for smaller values of $k + \ell$. Set $Y = S_{k-1,\ell}$, $Z = S_{k,\ell-1}$, $m = \phi(k-1, \ell)$ and $n = \phi(k, \ell-1)$, so that $Y = \{(i, y_i) : 1 \leq i \leq m\}$ contains neither a $(k-1)$-cup nor an ℓ-cap, and $Z = \{(i, z_i) : 1 \leq i \leq n\}$ contains neither a k-cup nor an $(\ell-1)$-cup.

For $\varepsilon > 0$, set $Y^{(\varepsilon)} = \{(i, \varepsilon y_i) : 1 \leq i \leq m\}$ and $Z^{(\varepsilon)} = \{(m+i, m+\varepsilon z_i)$: $1 \leq i \leq n\}$. Now, if $\varepsilon > 0$ is small enough then every line through two points of $Y^{(\varepsilon)}$ goes below the entire set $Z^{(\varepsilon)}$, and every line through two points of $Z^{(\varepsilon)}$ goes above the entire set $Y^{(\varepsilon)}$. Hence, in this case, every cup meeting $Z^{(\varepsilon)}$ in at least two points is entirely in $Z^{(\varepsilon)}$, and every cup meeting $Y^{(\varepsilon)}$ in at least two points is entirely in $Y^{(\varepsilon)}$. But then $Y^{(\varepsilon)} \cup Z^{(\varepsilon)}$ will do for $S_{k,\ell}$ since it continues neither a k-cup nor an ℓ-cup. □

As an easy consequence of Theorem 3, we see that every set of $\binom{2k-4}{k-2} + 1$ points in the plane in general position contains the vertices of some convex k-gon. In 1935, Erdős and Szekeres conjectured that, in fact, every set of $2^{k-2} + 1$ points in general position contains a convex k-gon. It does not seem likely that the conjecture will be proved in the near future, but it is known that, if true, the conjecture is best possible (see Exercise 23).

After this brief diversion, let us return to hypergraphs. Theorem 2 implies that every red–blue colouring of the r-tuples of the natural numbers contains arbitrarily large monochromatic subsets; a subset is *monochromatic* if its r-tuples have the same colour. Ramsey proved that, in fact, we can find an *infinite monochromatic* set.

Theorem 4 *Let $1 \leq r < \infty$ and let $c : A^{(r)} \to [k] = \{1, 2, \ldots, k\}$ be a k-colouring of the r-tuples of an infinite set A. Then A contains a monochromatic infinite set.*

l k	3	4	5	6	7	8	9	10	11	12	13	14	15
3	6	9	14	18	23	28	36	40 43	46 51	51 60	59 69	66 78	73 89
4		18	25	35 41	49 61	53 84	69 115	80 149	96 191	106 238	118 291	129 349	134 417
5			43 49	58 87	80 143	95 216	114 316	442					
6				102 165	298	495	780	1171					
7					205 540	1031	1713	2826					
8						1870	282 3583	6090					
9							6625	565 12715					
10								23854	798				

TABLE VI.1. Some values and bounds for two colour Ramsey numbers.

Proof. We apply induction on r. Note that the result is trivial for $r = 1$, so we may assume that $r > 1$ and the theorem holds for smaller values of r.

Put $A_0 = A$ and pick an element $x_1 \in A_0$. As in the proof of Theorem 2, define a a colouring $c_1 : B_1^{(r-1)} \to [k]$ of the $(r-1)$-tuples of $B_1 = A_0 - \{x_1\}$ by putting $c_1(\tau) = c(\tau \cup \{x_1\})$, $\tau \in B_1^{(r-1)}$. By the induction hypothesis B_1 contains an infinite set A_1 all of whose $(r-1)$-tuples have the same colour, say d_1, where $d_1 \in \{1, \ldots, k\}$. Let now $x_2 \in A_1$, $B_2 = A_1 - \{x_2\}$ and define a k-colouring $c_2 : B_2^{(r-1)} \to [k]$ by putting $c_2(\tau) = c(\tau \cup \{x_2\})$, $\tau \in B_2^{(r-1)}$. Then B_2 has an infinite set A_2 all of whose $(r-1)$-tuples have the same colour, say d_2. Continuing in this way we obtain an infinite sequence of elements: $x_1, x_2, \ldots$, an infinite sequence of colours: $d_1, d_2, \ldots$, and an infinite nested sequence of sets: $A_0 \supset A_1 \supset A_2 \supset \cdots$, such that $x_i \in A_{i-1}$, and for $i = 0, 1, \ldots$, all r-tuples whose only element outside A_i is x_i have the same colour d_i. The infinite sequence $(d_n)_1^\infty$ must take at least one of the k values $1, 2, \ldots, k$ infinitely often, say $d = d_{n_1} = d_{n_2} = \ldots$. Then, by the construction, each r-tuple of the infinite set $\{x_{n_1}, x_{n_2}, \ldots\}$ has colour d. □

In some cases it is more convenient to apply the following version of Theorem 4. As usual, the set of natural numbers is denoted by $\mathbb{N}$.

Theorem 5 *For each $r \in \mathbb{N}$, colour the set $\mathbb{N}^{(r)}$ of r-tuples of $\mathbb{N}$ with k_r colours, where $k_r \in \mathbb{N}$. Then there is an infinite set $M \subset \mathbb{N}$ such that for every r any two r-tuples of M have the same colour, provided their minimal elements are not less than the r^{th} element of M.*

Proof. Put $M_0 = \mathbb{N}$. Having chosen infinite sets $M_0 \supset \cdots \supset M_{r-1}$, let M_r be an infinite subset of M_{r-1} such that all the r-tuples of M_r have the same colour. This way we obtain an infinite nested sequence of infinite sets: $M_0 \supset M_1 \supset \cdots$. Pick $a_1 \in M_1$, $a_2 \in M_2 - \{1, \ldots, a_1\}$, $a_3 \in M_3 - \{1, \ldots, a_2\}$, etc. Clearly, $M = \{a_1, a_2, \ldots\}$ has the required properties. □

It is interesting to note that Ramsey's theorem for infinite sets, Theorem 3, easily implies the corresponding result for finite sets, although it fails to give bounds on the numbers $R^{(r)}(s_1, s_2, \ldots, s_k)$. To see this, all one needs is a simple compactness argument, a special case of Tychonov's theorem that the product of compact spaces is compact.

We have already formulated this (see Exercise III.30) but here we spell it out again in a convenient form.

Theorem 6 *Let r and k be natural numbers, and for every $n \geq 1$, let C_n be a non-empty set of k-colourings of $[n]^{(r)}$ such that if $n < m$ and $c_m \in C_m$ then the restriction $c_m^{(n)}$ of c_m to $[n]^{(r)}$ belongs to C_n. Then there is a colouring $c : \mathbb{N}^{(r)} \to [k]$ such that, for every n, the restriction $c^{(n)}$ of c to $[n]^{(r)}$ belongs to C_n.*

Proof. For $m > n$, write $C_{n,m}$ for the set of colourings $[n]^{(r)} \to [k]$ that are restrictions of colourings in C_m. Then $C_{n,m+1} \subset C_{n,m} \subset C_n$ and so $\tilde{C}_n = \bigcap_{m=n+1}^{\infty} C_{n,m} \neq \emptyset$ for every n, since each $C_{n,m}$ is finite. Let $c_r \in \tilde{C}_r$, and pick $c_{r+1} \in \tilde{C}_{r+1}$, $c_{r+2} \in \tilde{C}_{r+2}$, and so on, such that each is in the preimage of the previous one: $c_n = c_{n+1}^{(n)}$. Finally, define $c : \mathbb{N}^{(r)} \to [k]$ by setting, for $\rho \in \mathbb{N}^{(r)}$,

$$c(\rho) = c_n(\rho) = c_{n+1}(\rho) = \cdots,$$

where $n = \max \rho$. This colouring c is as required. □

Let us see then that Theorem 5 implies that $R^{(r)}(s_1, s_2, \ldots, s_k)$ exists. Indeed, otherwise for every n there is a colouring $[n]^{(r)} \to [k]$ such that, for each i, there is no s_i-set all of whose r-sets have colour i. Writing C_n for the set of all such colourings, we see that $C_n \neq \emptyset$ and $C_{n,m} \subset C_n$ for all $n < m$, where $C_{n,m}$ is as in the proof of Theorem 5. But then there is a colouring $c : \mathbb{N}^{(r)} \to [k]$ such that every monochromatic set has fewer than $s = \max s_i$ elements, contradicting Theorem 4.

To conclude this section, we point out a fascinating phenomenon. First, let us see an extension of the fact that $R_k^{(r)}(s_1, \ldots, s_k)$ exists.

Theorem 7 *Let r, k and $s \geq 2$. If n is sufficiently large then for every k-colouring of $[n]^{(r)}$ there is a monochromatic set $S \subset [n]$ such that*

$$|S| \geq \max\{s, \min S\}.$$

Proof. Suppose that there is no such n, that is, for every n there is a colouring $[n]^{(r)} \to [k]$ without an appropriate monochromatic set. Let C_n be the set of all such colourings. Then $C_n \neq \emptyset$ and, in the earlier notation, $C_{n,m} \subset C_n$ for all $n < m$. But then there is a colouring $c : \mathbb{N}^{(r)} \to [k]$ such that its restriction $c^{(n)}$ to $[n]^{(r)}$ belongs to C_n. Now, by Theorem 4, there is an infinite monochromatic set $M \subset \mathbb{N}$. Set $m = \min M$, $t = \max\{m, s\}$, and let S consist of the first t elements of M. Then, with $n = \max S$, the colouring $c^{(n)}$ does have an appropriate monochromatic set, namely S, contradicting $c^{(n)} \in C_n$. □

This is a beautiful result but it is not too unexpected. What *is* surprising and deep is that, as proved by Paris and Harrington in 1977, although Theorem 7 is a (fairly simple) assertion concerning finite sets, it cannot be deduced from the Peano axioms, that is, it cannot be proved within the theory of finite sets. In other words, we actually *need* the notion of a finite set to prove Theorem 7. This theorem of Paris and Harrington became the starting point of an active area connecting combinatorics and logic.

As this is a book on graph theory, we cannot digress too far into logic, so let us return to graphs. Let $R^*(s)$ be the minimal integer n such that for every two-colouring of $[n]^{(2)}$ there is a monochromatic set $S \subset [n]$ with $|S| \geq \max\{s, |S|\}$. Thus $R^*(s)$ is the minimal value of n such that for every graph G with vertex set $[n]$ there is a set $S \subset [n]$ with $|S| \geq \max\{s, |S|\}$ such that $G[S]$ is trivial, that is, either complete or empty. We know from Theorem 7 that $R^*(s)$ exists. Clearly, $R^*(s) \geq R(s)$ but, not surprisingly, $R^*(s)$ is of a greater order of magnitude than $R(s)$: it turns out that there are positive constants c and d such that $2^{2^{cs}} < R^*(s) < 2^{2^{ds}}$.

VI.2 Canonical Ramsey Theorems

Can anything significant be said about colourings of $\mathbb{N}^{(r)}$ with infinitely many colours? Can we guarantee that there is an infinite set $M \subset \mathbb{N}$ such that on $M^{(r)}$ our colouring is particularly 'nice'? In 1950, Erdős and Rado proved that, unexpectedly, this is precisely the case.

In what follows, M, N, M_1, N_1, ... denote countable infinite sets, and r, s, ... are natural numbers.

We call two colourings $c_1 : N_1^{(r)} \to C_1$ and $c_2 : N_2^{(r)} \to C_2$ *equivalent* if there is a 1-to-1 map ϕ of N_1 onto N_2 such that for $\rho, \rho' \in N_1^{(r)}$ we have $c_1(\rho) = c_1(\rho')$ if and only if $c_2(\phi(\rho)) = c_2(\phi(\rho'))$.

In an ideal world, for every colouring of $N^{(r)}$ (with any number of colours) there would be an infinite set $M \subset N$ on which the colouring is equivalent to one of finitely many colourings. Surprisingly, even more is true.

Call a colouring $c : N^{(r)} \to C$ *irreducible* if for every infinite subset N_1 of N, the restriction of c to $N_1^{(r)}$ is equivalent to c. Also, call a set $\mathcal{C}$ of colourings $\mathbb{N}^{(r)} \to \mathbb{N}$ *unavoidable* if for every colouring c of $\mathbb{N}^{(r)}$ there is an infinite set $M \subset \mathbb{N}$ such that the restriction of c to $M^{(r)}$ is equivalent to a member of $\mathcal{C}$. Erdős and Rado proved that for every r there is a finite unavoidable family of irreducible colourings.

What are examples of irreducible colourings of $\mathbb{N}^{(r)}$? Two constructions spring to mind: a monochromatic colouring, in which all r-sets get the same colour, and an all-distinct colouring, in which no two sets get the same colour. After a moment's thought, we can construct more irreducible colourings. Given $N \subset \mathbb{N}$, $\alpha = \{a_1, \cdots, a_r\} \in N^{(r)}$, $a_1 < \ldots < a_r$, and $S \subset [r] = \{1, \ldots, r\}$, $|S| = s$, set $\alpha_S = \{a_i : i \in S\}$. Define the *S-canonical* colouring $c_S : N^{(r)} \to N^{(s)}$, by setting

$c_S(\alpha) = \alpha_S$. Thus we are colouring the elements of $\mathbb{N}^{(r)}$ with s-sets, and two r-sets get the same colour iff their i th elements coincide for $i \in S$ and are different for $i \notin S$. It is easily seen (see Exercise 29) that c_S is an irreducible colouring for every $S \subset [r]$; also, these colourings include the two irreducible colourings mentioned above: $c_\emptyset$ is a monochromatic colouring and $c_{[r]}$ is an all-distinct colouring.

Clearly, for $S \neq S'$ the colourings c_S and $c_{S'}$ of $N^{(r)}$ are not equivalent, so $N^{(r)}$ has at least 2^r irreducible colourings, namely the 2^r canonical colourings. As we shall see, there are no other irreducible colourings. At first sight this might be rather surprising since a canonical colouring of $N^{(r)}$ depends on the order of elements of N. To resolve this 'paradox', note that if $\{a_1, a_2, \ldots\}$ and $\{b_1, b_2, \ldots\}$ are two enumerations of N then there are subsequences $a_{k_1}, a_{k_2}, \ldots$ and $b_{l_1}, b_{l_2}, \ldots$, $k_1 < k_2 < \cdots, l_1 < l_2 < \cdots$, such that $a_{k_i} = b_{l_i}$ for every i.

Before we turn to the results, let us introduce a concept similar to the equivalence of colourings, but taking into account the order on the underlying set. Let $c : \mathbb{N}^{(r)} \to C$ and $T, U \in \mathbb{N}^{(t)}$ for some $t \geq r$. Also, let $\phi : T \to U$ be the unique order-preserving map from T onto U. The sets T and U are said to have the same *pattern* (with respect to c) if for $\rho, \rho' \in T^{(r)}$ we have $c(\rho) = c(\rho')$ if, and only if, $c(\phi(\rho)) = c(\phi(\rho'))$. Note that the number of patterns of t-sets is precisely the number of partitions of $\binom{t}{r}$ distinguishable objects; clearly, $\binom{t}{r}^{\binom{t}{r}}$ is a crude upper bound for this number.

After all this preparation, let us prove the Erdős–Rado canonical theorem for graphs, that is, for $r = 2$. Note that for every infinite set $N \subset \mathbb{N}$ there are four canonical colourings of $N^{(2)}$. In the $\emptyset$-canonical colouring of $N^{(2)}$, all edges have the same colour, in the $\{1, 2\}$-canonical colouring all edges have distinct colours, in the $\{1\}$-canonical colouring two edges have the same colour iff their first vertices coincide, and in the $\{2\}$-canonical colouring two edges have the same colour iff their second vertices coincide.

Theorem 8 *For every colouring $c : \mathbb{N}^{(2)} \to \mathbb{N}$ there is an infinite subset M of $\mathbb{N}$ such that the restriction of c to $M^{(2)}$ is canonical.*

Proof. As there are only finitely many patterns for the colourings of $[4]^{(2)}$, we may apply Ramsey's theorem for infinite sets, obtaining an infinite set $M \subset \mathbb{N}$ such that all 4-sets of M have the same pattern π. We claim that this set M will do.

Let $M = \{m_1, m_2, \cdots\}$, where $m_1 < m_2 < \ldots$. Since all 4-sets have the same pattern, for any two edges $m_i m_j$ and $m_k m_l$, the colours $c(m_i m_j)$ and $c(m_k m_l)$ do or do not coincide, according to the relative position of the pairs ij and kl in the set $\{i, j, k, l\}$. For example, 25 and 57 have the same relative position as 36 and 67; similarly, 38 and 46 have the same position as 29 and 78.

After these observations, let us prove that the restriction of c to $M^{(2)}$ is canonical. With a slight abuse of notation, from now on write c for the restriction of c to $M^{(2)}$. We may assume that $c \neq c_{\{1,2\}}$, that is, $M^{(2)}$ has two edges of the same colour: say $c(m_i m_j) = c(m_k m_l)$, where $m_i \notin \{m_j, m_k\}$. Note that we do not (and can not) assume that $i < j$ or $i > j$. But then $c(m_{2i} m_{2j}) = c(m_{2k} m_{2l}) = c(m_{2i+1} m_{2j})$

so $M^{(2)}$ has two *adjacent* edges of the same colour. Let us distinguish three cases according to the positions of these adjacent edges, and see what we can deduce.

(i) Suppose first that $c(m_i m_j) = c(m_i m_k)$ for some $i < j < k$. Then, by considering the 4-set $\{m_i, m_j, m_k, m_{k+1}\}$, we see that in the pattern π the edges 12 and 13 get the same colour. But then any two edges sharing their first vertices have the same colour, since if $r < s < t$ then the restriction of c to $\{m_r, m_s, m_t, m_{t+1}\}^{(2)}$ shows that $c(m_r m_s) = c(m_r m_t)$. This means that there is a colouring $d : M \to \mathbb{N}$ such that for $r < s$ we have $c(m_r m_s) = d(m_r)$.

(ii) Suppose next that $c(m_i m_k) = c(m_j m_k)$ for some $i < j < k$. Then, similarly, there is a map $e : M \to \mathbb{N}$ such that $c(m_r m_s) = e(m_s)$ if $r < s$.

(iii) Finally suppose that $c(m_i m_j) = c(m_j m_k)$ for some $i < j < k$. Then $c(m_r m_s) = c(m_s m_t)$ for all $r < s < t$. Hence $c(m_1 m_3) = c(m_3 m_5) = c(m_2 m_3) = c(m_3 m_4)$, say, so there are edges of the same colour sharing their second vertices. Therefore, there are maps $d : M \to \mathbb{N}$ and $e : M \to \mathbb{N}$ such that if $i < j$ then $c(m_i m_j) = d(m_i) = e(m_j)$. But then any two edges of $M^{(2)}$ have the same colour, so $c = c_\emptyset$.

What we have seen so far is that if (iii) holds then we are done. In fact, it is very easy to complete the proof in the case when (iii) does not hold. Indeed, if (i) holds but (iii) does not then $d(m_i) \neq d(m_j)$ for all $i \neq j$, so $c = c_{\{1\}}$, and if (ii) holds but (iii) does not then $e(m_i) \neq e(m_j)$ for $i \neq j$, so $c = c_{\{2\}}$. □

As it happens, the proof of the full Erdős–Rado canonical theorem is hardly more complicated than the proof above.

Theorem 9 *Let r be a positive integer and $c : \mathbb{N}^{(r)} \to \mathbb{N}$ a colouring. Then there is an infinite subset M of $\mathbb{N}$ such that the restriction of c to $M^{(r)}$ is canonical.*

Proof. Let us apply induction on r. For $r = 1$ there is nothing to prove, so suppose that $r \geq 2$ and the theorem holds for smaller values of r. Given $c : \mathbb{N}^{(r)} \to \mathbb{N}$, colour each $T \in \mathbb{N}^{(2r)}$ with the pattern of the restriction of c to $T^{(r)}$. As there are only finitely many patterns, there is an infinite set $N \subset \mathbb{N}$ such that all $2r$-subsets of N have the same pattern π. In order to simplify the notation, we assume that $N = \mathbb{N}$: all this amounts to is an appropriate relabelling.

If no two r-subsets of N have the same colour then we are done: $c = c_{[r]}$. Therefore we may assume that $c(\rho) = c(\sigma)$ for some $\rho, \sigma \in N^{(r)}$, $\rho \neq \sigma$; say $\rho = \{a_1, \ldots, a_r\}$ and $\sigma = \{b_1, \ldots, b_r\}$, where $a_1 < \cdots < a_r$ and $b_1 < \cdots < b_r$. As $\rho \neq \sigma$, there is an element $b_i \in \sigma \backslash \rho$. Note that all the sets $\rho_0 = \{2a_1, 2a_2, \ldots, 2a_r\}$, $\sigma_1 = \{2b_1, 2b_2, \ldots, 2b_r\}$ and $\sigma_2 = \{2b_1, 2b_2, \ldots, 2b_{i-1}, 2b_i - 1, 2b_{i+1}, \ldots, 2b_r\}$ get the same colour. Indeed, $|\rho_0 \cup \sigma_1| = |\rho_0 \cup \sigma_2| = u$, say, so there are sets $T_1, T_2 \in N^{(2r)}$ such that $\rho_0 \cup \sigma_1$ is the set of the first u elements of T_1, and $\rho_0 \cup \sigma_2$ is the set of the first u elements of T_2. As $T_1^{(r)}$ has pattern π, we have $c(\rho_0) = c(\sigma_1)$, and as $T_2^{(r)}$ has pattern π, we have $c(\rho_0) = c(\sigma_2)$.

Now, since σ_1 and σ_2 get the same colour, any two r-subsets of N differing only in the i th place also get the same colour: if $\tau, \tau \in N^{(r)}$ and $\tau_{[r]-\{i\}} = \tau'_{[r]-\{i\}}$ then

$c(\tau) = c(\tau')$, that is,the colour of $\tau \in N^{(r)}$ depends only on $\tau_{[r]-\{i\}}$. This enables us to define a colouring $c' : N^{(r-1)} \to \mathbb{N} \cup \{\infty\}$ as follows: for $\nu \in N^{(r-1)}$ set $c'(\nu) = c(\tau)$ if $\nu = \tau_{[r]-\{i\}}$ for some $\tau \in N^{(r)}$, and $c'(\nu) = \infty$ otherwise.

By the induction hypothesis, there is an infinite set $M \subset N$ such that c' is canonical on $M^{(r-1)}$. Then $c'(\nu) \neq \infty$ for $\nu \in M^{(r-1)}$, and c is a canonical colouring of $M^{(r)}$. □

As an amusing point, note that Theorem 9 is clearly stronger than Theorem 4 since for an infinite set $M \subset \mathbb{N}$ the only canonical colouring of $M^{(r)}$ that uses finitely many colours is $c_\emptyset$, the canonical colouring using only one colour.

VI.3 Ramsey Theory For Graphs

Let H_1 and H_2 be arbitrary graphs. Given n, is it true that every red–blue colouring of the edges of K_n contains a red H_1 or a blue H_2? Since H_i is a subgraph of K_{s_1}, where $s_i = |H_i|$, the answer is clearly "yes" if $n \geq R(s_1, s_2)$. Let $r(H_1, H_2)$ be the smallest value of n that will ensure an affirmative answer, and define $r(H_1, \ldots, H_k)$ analogously for k colours. Note that this notation is similar to the one introduced earlier: $R(s_1, s_2) = r(K_{s_1}, K_{s_2})$. Instead of a red–blue colouring, one frequently works with a graph and its complement: clearly, $r(H_1, H_2) - 1$ is the maximal value of n for which there is a graph G of order n such that $H_1 \not\subset G$ and $H_2 \not\subset \overline{G}$.

The numbers $r(H_1, \ldots, H_k)$, called *generalized Ramsey numbers* or *graphical Ramsey numbers*, have been the subject of much study, and by now there is a large body of results about them. Nevertheless, there is a long way to go, which is not surprising, since the generalized Ramsey numbers include the classical Ramsey numbers $R(s, t)$. Here we shall present some of the basic results about generalized Ramsey numbers.

In order to avoid trivialities, throughout this discussion we shall assume that $H_1, H_2, \ldots$ do not have isolated vertices. Let us start with the observation that if H_1 is very sparse, say it consists of ℓ independent edges, then $r(H_1, H_2)$ is rather small. In fact, if H_1 consists of ℓ independent edges, and H_2 is a complete graph, then we can determine $r(H_1, H_2)$ exactly.

Theorem 10 *For $\ell \geq 1$ and $p \geq 2$ we have*

$$r(\ell K_2, K_p) = 2\ell + p - 2.$$

Proof. The graph $K_{2\ell-1} \cup E_{p-2}$ does not contain ℓ independent edges, and its complement, $E_{2\ell-1} + K_{p-2}$, does not contain a complete graph of order p. Hence $r(\ell K_2, K_p) \geq 2\ell + p - 2$.

On the other hand, let G be a graph of order $n = 2\ell + p - 2$, containing a maximal set of $s \leq \ell - 1$ independent edges. Then the set of $n - 2s \geq 2\ell + p - 2 - 2(\ell - 1) = p$ vertices not on these edges spans a complete graph of order at least p. Therefore $r(\ell K_2, K_p) \leq 2\ell + p - 2$. □

Note that if H is any graph of order h then, by Theorem 10, $r(\ell K_2, H) \leq r(\ell K_2, K_h) \leq 2\ell + h - 2$.

The next observation is a lower bound for $r(H_1, H_2)$, valid for all pairs (H_1, H_2). For a graph G, denote by $c(G)$ the maximal order of a component of G, and by $u(G)$ the *chromatic surplus* of G: the minimal number of vertices in a colour class, taken over all proper $\chi(G)$-colourings of G. Thus $u(G) = \min\{U \subset V(G) : \chi(G - U) < \chi(G)\}$. For example, $u(C_2k) = k$ and $u(C_{2k+1}) = 1$.

Theorem 11 *For all nonempty graphs H_1 and H_2 we have*

$$r(H_1, H_2) \geq (\chi(H_1) - 1)(c(H_2) - 1) + u(H_1).$$

In particular, if H_2 is connected then

$$r(H_1, H_2) \geq (\chi(H_1) - 1)(|H_2| - 1) + 1.$$

Proof. Set $k = \chi(H_1)$, $u = u(H_1)$ and $c = c(H_2)$. Trivially, $r(H_1, H_2) \geq r(H_1, K_2) = |H_1| \geq \chi(H_1)u(H_1) = ku$. Hence, if $c \leq u$ then $r(H_1, H_2) \geq ku \geq (k - 1)c + u$. On the other hand, if $c > u$ then the graph $G = (k - 1)K_{c-1} \cup K_{u-1}$ does not contain H_2, and its complement does not contain H_1. Therefore, $r(H_1, H_2) \geq |G| + 1 = (k - 1)(c - 1) + u$. □

Although the inequalities in Theorem 9 are very simple, in some cases they are best possible. Let us see two examples of this: the first is a beautiful result of Chvátal.

Theorem 12 *Let $s, t \geq 2$. then for every tree T of order t we have $r(K_s, T) = (s - 1)(t - 1) + 1$.*

Proof. From Theorem 10 we know that $r(K_s, T) \geq (s - 1)(t - 1) + 1$. To prove the reverse inequality, let G be a graph of order $n = (s - 1)(t - 1) + 1$ whose complement does not contain K_s. Then $\chi(G) \geq \lceil n/(s - 1) \rceil = t$ so it contains a critical subgraph H of minimal degree at least $t - 1$ (see Theorem V.1). It is easily seen that H contains (a copy of) T. Indeed, we may assume that $T_1 \subset H$, where $T_1 = T - x$ and x is an endvertex of T, adjacent to a vertex y of T_1 (and of H). Since y has at least $t - 1$ neighbours in H, at least one of its neighbours, say z, does not belong to T_1. Then the subgraph of H spanned by T_1 and z clearly contains (a copy of) T. □

The second example of equality in Theorem 8 concerns fans. For $l \geq 1$, the graph $H_\ell = K_1 + \ell K_2$ is called a *fan with ℓ blades.* Thus $F_1 = K_3$, and F_ℓ is made up of ℓ triangles with a vertex in common. In 1996, Li and Rousseau demonstrated the following result.

Theorem 13 *For $\ell \geq 2$ we have $r(F_1, F_\ell) = r(K_3, F_\ell) = 4\ell + 1$.*

Proof. We know from Theorem 11 that $r(K_3, F_\ell) \geq 2(|F_\ell| - 1) + 1 = 4\ell + 1$.

To prove the reverse inequality, suppose that the inequality is false; that is, there is a triangle-free graph G of order $n = 4\ell + 1$ whose complement does not contain F_ℓ.

For $x \in G$, let $U = \Gamma_G(x)$. Then U is a set of independent vertices and since $\overline{G}$ does not contains F_ℓ, we see that $d_G(x) = |U| \leq 2\ell$.

On the other hand, how large can the degree of x be in $\overline{G}$? Set $W = \Gamma_{\overline{G}}(x) = V(G) - (U \cup \{x\})$. Then $\overline{G}[W]$ does not contain ℓ independent edges, and its complement, $G[W]$, has no triangle. Hence, by Theorem 10, $d_{\overline{G}}(x) = |W| \leq 2\ell$.

This shows that $d_G(x) = d_{\overline{G}}(x) = 2\ell$, that is, G is a triangle-free 2ℓ-regular graph of order $4\ell + 1$. But from the result in Exercise IV.48 we know that this is impossible. □

If we define a graph H_2 to be H_1-*good* if equality holds in Theorem 11, then the previous two results claim that every tree is K_s-good for $s \geq 2$, and the fan F_ℓ is K_3-good for $\ell \geq 2$. In fact, Li and Rousseau proved also that, for every fixed $s \geq 2$, if ℓ is large enough then F_ℓ is K_s-good. Even more, if H_1 and H_2 are fixed graphs and ℓ is large enough then $K_1 + \ell H_2$ is $(K_2 + H_1)$-good.

As we know very little about $r(K_s, K_t)$, it is only to be expected that $r(G_1, G_2)$ has been determined mostly in the cases when at least one of G_1 and G_2 is sparse, as in Theorems 12 and 13. As we shall see now, there are particularly pleasing results for $r(sH_1, tH_2)$ when H_1 and H_2 are fixed and s and t are large. The following simple lemma shows that for fixed H_1 and H_2 the function $r(sH_1, tH_2)$ is at most $s|H_1| + t|H_2| + c$, where c depends only on H_1 and H_2, and not on s and t.

Lemma 14 *For all graphs G, H_1 and H_2 we have $r(G, H_1 \cup H_2) \leq \max\{r(G, H_1) + |H_2|, r(G, H_2)\}$. In particular, $r(sH_1, H_2) \leq r(H_1, H_2) + (s-1)|H_1|$.*

Proof. Let $n = \max\{r(G, H_1) + |H_2|, r(G, H_2)\}$, and suppose that we are given a red–blue colouring of K_n without a red G. Then $n \geq r(G, H_2)$ implies that there is a blue H_2. Remove it. Since $n - |H_2| \geq r(G, H_1)$, the remainder contains a blue H_1. Hence K_n contains a blue $H_1 \cup H_2$. □

This simple lemma can be used to determine $r(sH, tH)$ when H is K_2 or K_3.

Theorem 15 *If $s \geq t \geq 1$ then*

$$r(sK_2, tK_2) = 2s + t - 1.$$

Proof. The graph $G = K_{2s-1} \cup E_{t-1}$ does not contain s independent edges and $\overline{G} = E_{2s-1} + K_{t-1}$ does not contain t independent edges. Hence $r(sK_2, tK_2) \geq 2s + t - 1$.

Trivially (or, by Theorem 10), $r(sK_2, K_2) = 2s$, so to complete the proof it suffices to show that

$$r((s+1)K_2, (t+1)K_2) \leq r(sK_2, tK_2) + 3.$$

To see this, let G be a graph of order $n = r(sK_2, tK_2) + 3 \geq 2s + t + 2$. If $G = K_n$ then $G \supset (s+1)K_2$, and if $G = E_n$ then $\overline{G} \supset (t+1)K_2$. Otherwise, there are three vertices, say x, y and z, such that $xy \in G$, $xz \notin G$. Now, either $G - \{x, y, z\}$ contains s independent edges of G and then xy can be added to them to form $s + 1$ independent edges of G, or else $\overline{G} - \{x, y, z\}$ contains t independent edges and then xz can be added to them to form $t + 1$ independent edges of $\overline{G}$. □

Theorem 16 *If $s \geq t \geq 1$ and $s \geq 2$ then $r(sK_3, tK_3) = 3s + 2t$.*

Proof. Let $G = K_{3s-1} \cup (K_1 + E_{2t-1})$. Then G does not contain s independent triangles and $\overline{G} = E_{3s-1} + (K_1 \cup K_{2t-1})$ does not contain t independent triangles. Hence $r(sK_3, tK_3)$ is at least as large as claimed.

It is not difficult to show that $r(2K_3, K_3) = 8$ and $r(2K_3, 2K_3) = 10$ (Exercise 15). Hence repeated applications of Lemma 14 give

$$r(sK_3, K_3) \leq 3s + 2,$$

and to complete the proof it suffices to show that for $s \geq 1, t \geq 1$ we have

$$r((s+1)K_3, (t+1)K_3) \leq r(sK_3, tK_3) + 5.$$

To see this, let $n = r(sK_3, tK_3) + 5$ and consider a red–blue colouring of K_n. Select a monochromatic (say red) triangle R_3 in K_n. If $K_n - R_3$ contains a red sK_3 then we are home. Otherwise, $K_n - R_3$ contains a blue triangle B_3 (it even contains a blue tK_3). We may assume that at least five of the nine $R_3 - B_3$ edges are red. At least two of these edges are incident with a vertex of B_3, and together with an edge of R_3 they form a red triangle R_3^* meeting B_3. Since $K_n - R_3^* - B_3$ has $r(sK_3, tK_3)$ vertices, it contains either a red sK_3 or a blue tK_3. These are disjoint from both R_3^* and B_3, so K_n contains either a red $(s+1)K_3$ or a blue $(t+1)K_3$. □

By elaborating the idea used in the proofs of the previous two theorems we can obtain good bounds on $r(sK_p, tK_q)$, provided that $\max(s, t)$ is large compared to $\max(p, q)$. Let $p, q \geq 2$ be fixed and choose t_0 such that

$$t_0 \min\{p, q\} \geq 2r(K_p, K_q).$$

Put $C = r(t_0K_p, t_0K_q)$.

Theorem 17 *If $s \geq t \geq 1$ then*

$$ps + (q-1)t - 1 \leq r(sK_p, tK_q) \leq ps + (q-1)t + C.$$

Proof. The graph $K_{ps-1} \cup E_{(q-1)t-1}$ shows the first inequality. As in the proofs of the previous theorems, we fix $s - t$ and apply induction on t. By Lemma 14 we have

$$r(sK_p, tK_q) \leq (s-t)p + r(tK_p, tK_q) \leq ps + C,$$

provided that $t \leq t_0$. Assume now that $t \geq t_0$ and the second inequality of the theorem holds for s, t.

Let G be a graph of order $n = p(s+1) + (q-1)(t+1) + C$ such that $G \not\supset (s+1)K_p$ and $\overline{G} \not\supset (t+1)K_q$. We claim that some K_p of G and K_q of $\overline{G}$ share a vertex. Indeed, suppose that this is not the case. By altering G, if necessary, we may assume that $G \supset sK_p$ and $\overline{G} \supset tK_q$. Denote by V_p the set of vertices of G that are in K_p subgraphs and put $V_q = V \backslash V_p$, $n_p = |V_p|$, and $n_q = |V_q|$.

By our assumption, $n_p \geq sp$ and $n_q \geq tq$. In the graph G, a vertex $x \in V_q$ is joined to at most $r(K_{p-1}, K_q) - 1$ vertices of V_p, since otherwise there is a K_p of G containing x or else a K_q of $\overline{G}$ consisting of vertices of V_p. Similarly, in the graph $\overline{G}$ every vertex $y \in V_q$ is joined to at most $r(K_p, K_{q-1}) - 1$ vertices of V_p. Hence, counting the $V_p - V_q$ edges in G and $\overline{G}$, we find that

$$n_q r(K_{p-1}, K_q) + n_p r(K_p, K_{q-1}) > n_p n_q.$$

However, this is impossible, since $n_p \geq sp \geq t_0 p$ and $n_q \geq tq \geq t_0 q$, so $n_p \geq 2r(K_{p-1}, K_q)$ and $n_q \geq 2r(K_p, K_{q-1})$. Therefore, we *can* find a K_p of G and a K_q of $\overline{G}$ with a vertex in common.

When we omit the $p + q - 1$ vertices of these two subgraphs, we find that the remainder H is such that $H \not\supset sK_p$ and $\overline{H} \not\supset tK_q$. However, $|H| = ps + (q-1)t + C$, so this is impossible. □

In all the results above, we have $r(H_1, H_2) \leq C(|H_1| + |H_2|)$, where C depends only on the maximal degrees of H_1 and H_2. That this is not by chance is a beautiful and deep theorem, proved by Chvátal, Rödl, Szemerédi and Trotter in 1983.

Theorem 18 *For every $d \geq 1$ there is a constant $c = c(d)$ such that if $\Delta(H) \leq d$ then $r(H, H) \leq c|H|$.* □

In fact, it is likely that much more is true. Burr and Erdős conjectured in 1975 that the maximal degree can be replaced by the maximum of the minimal degrees of subgraphs, as in Theorem V.1: for every d there is a constant $c = c(d)$ such that if every subgraph of H has a vertex of degree at most d then $r(H) \leq c|H|$.

Additional evidence for the truth of this conjecture was provided by Chen and Schelp: they proved that $r(H) \leq c|H|$ for some absolute constant c and every planar graph H. Extending this result, Rödl and Thomas proved in 1995 that for every k there is a constant $c = c(k)$ such that if H has no subcontraction to K_k then $r(H) \leq c|H|$.

It would not be unreasonable to think that the various Ramsey theorems hold for finite graphs, because the graph whose edges we colour is K_n and not some sparse graph with few edges. For example, one might guess that, if G is a graph such that whenever the edges of G are k-coloured there is a monochromatic K_s, then G has to be rather dense. In fact, this is not the case at all. For every graph H with clique number $r = \omega(H)$ and every $k \geq 1$ there is a graph G with clique number also r such that every k-colouring of G contains a monochromatic copy of H. This beautiful result was proved by Nešetřil and Rödl in the following stronger form, extending earlier results of Graham and Folkman.

Theorem 19 *For every graph H and integer $k \geq 1$, there is a graph G with $\omega(G) = \omega(H)$ such that every k-colouring of the edges of G contains a monochromatic induced subgraph isomorphic to H.* □

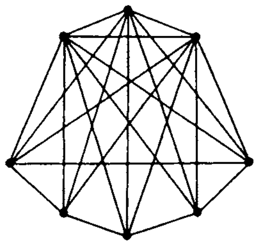


FIGURE VI.1. The graph $C_3 + C_5$.

To conclude this section, let us note that occasionally mainstream problems of extremal graph theory masquerade as problems of Ramsey theory, as the problems have very little to do with *partitioning* the edges. For example, what is $r(H, K_{1,\ell})$? It is the smallest value of n such that every graph G of order n and minimal degree at least $n - \ell$ contains a copy of H. As another example, if

$$k\,\mathrm{ex}(n; H) < \binom{n}{2},$$

then in every k-colouring of the edges of K_n there is a colour class with more than $\mathrm{ex}(n; H)$ edges, so that colour class automatically contains a copy of H.

For example, by Theorem IV.12,

$$\mathrm{ex}(n; C_4) \le \frac{n}{4}(1 + \sqrt{4n-3}),$$

so for $n = k^2 + k + 2$ we have

$$\begin{aligned} k\,\mathrm{ex}(n; C_4) &\le k\frac{k^2+k+2}{4}(1 + (2k+1)) \\ &= \frac{(k^2+k)(k^2+k+2)}{2} < \binom{n}{2}. \end{aligned}$$

Therefore, $r_k(C_4) \le k^2 + k + 2$. Chung and Graham showed that this bound is close to being best possible: $r_k(C_4) \ge k^2 - k + 2$ if $k - 2$ is a prime power (see Exercise 17).

VI.4 Ramsey Theory for Integers

It may sound strange that the first results concerning monochromatic substructures arose in connection with the integers, rather than graphs; however, as graph theory is very young indeed, this is not too surprising. In this section we shall present three classical results, together with some substantial recent developments.

Perhaps the first result of Ramsey theory is a theorem of Hilbert concerning 'cubes' in the set of natural numbers. Although the result is simple, its proof is clearly more than a straightforward application of the pigeon-hole principle.

Let us call a set $C \subset \mathbb{N}$ an *ℓ-cube in* $\mathbb{N}$ if there are natural numbers $s_0, s_1, \ldots, s_\ell$, with $s_1 + \ldots + s_i < s_{i+1}$ for $1 \le i < \ell$, such that

$$C = C(s_0; s_1, \ldots, s_\ell) = \{s_0 + \sum_{i=1}^{\ell} \varepsilon_i s_1 : \varepsilon_i = 0 \text{ or } 1\}.$$

Thus an ℓ-cube in $\mathbb{N}$ is the affine image of the unit cube $\{0, 1\}^\ell \subset \mathbb{R}$, and this affine image has 2^ℓ vertices.

In 1892 Hilbert proved the following result.

Theorem 20 *If $\mathbb{N}$ is coloured with finitely many colours then, for every $\ell \ge 1$, one of the colour classes contains infinitely many translates of the same ℓ-cube.*

Proof. It clearly suffices to prove the following finite version of this result.

There is a function $H : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that if $N \ge H(k, l)$ then every k-colouring of $[N]$ contains a monochromatic ℓ-cube.

Since a 1-cube in $\mathbb{N}$ is just a pair of integers, $H(k, 1) = k + 1$ will do in this assertion. Therefore, it suffices to show that if we can have $H(k, \ell) \le n$ then $H(k, \ell + 1) = N = kn^{\ell+1}$ will do.

To see this, let $c : [N] \to [k]$ be a k-colouring, and partition $[N]$ into $N/n = kn^\ell$ intervals, each of length n:

$$[N] = \bigcup_{j=1}^{N/n} I_j,$$

where $I_j = [(j - 1)n + 1, jn], j = 1, \ldots, N/n$. Then each I_j contains a monochromatic ℓ-cube. But, up to translation, there are at most $(n - 1)^\ell < n^\ell$ cubes in these intervals, and each monochromatic cube can get one of k colours. Since there are kn^ℓ intervals, some two of these intervals, say I_j and I_h, contain translations of the *same* ℓ-cube C_ℓ in the *same* colour. The union of these two translations is a monochromatic $(\ell + 1)$-cube. □

The result above had essentially no influence on the development of Ramsey theory, but the following theorem, proved by Schur in 1916, became the starting point of an area that is still very active today.

Theorem 21 *For every $k \ge 1$ there is an integer m such that every k-colouring of $[m]$ contains integers x, y, z of the same colour such that*

$$x + y = z.$$

Proof. We claim that $m = R_k(3) - 1$ will do, where $R_k(3) = R_k(3, \ldots, 3)$ is the graphical Ramsey number for k colours and triangles, i.e., the minimal integer n such that every k-colouring of the edges of K_n contains a monochromatic triangle.

Let then $n = R_k(3)$ and let $c : [m] = [n - 1] \to [k]$ be a k-colouring. Induce a k-colouring of $[n]^{(2)}$, the edge set of the complete graph with vertex set $[n]$, as follows: for $ij \in E(K_n) = [n]^{(2)}$ set $c'(ij) = c(|i - j|)$. By the definition of $n = R_k(3)$, there is a monochromatic triangle, say with vertex set $\{h, i, j\}$, so that $1 \le h < i < j \le n$ and $c'(hi) = c'(ij) = c'(hj) = \ell$ for some ℓ. But then

$x = i - h$, $y = j - i$ and $z = j - h$ are such that $c(x) = c(y) = c(z) = \ell$ and $x + y = z$. □

Writing $S(k)$ for the minimal integer m that will do in Theorem 20, we see that $S(k) \leq R_k(3) - 1$. As it is easily shown that $R_k(3) \leq \lfloor ek! \rfloor + 1$ (see Exercise 25), we find that $S(k) \leq ek!$.

The third and most important classical result, predating Ramsey's theorem, was proved by van der Waerden in 1927. The *length* of a sequence is the number of its terms.

Theorem 22 *Given p and k, if n is large enough, then every k-colouring of $[n]$ contains a monochromatic arithmetic progression of length p.*

In view of Theorem 22, we can define the *van der Waerden functions* $W(p)$ and $W(p, k)$. Here $W(p) = W(p, 2)$, and $W(p, k)$ is the minimal value of n that will do in Theorem 22; thus $W(p, k)$ is the minimal integer n such that if $[n] = \bigcup_{i=1}^{k} N_i$ then there are $a, d \geq 1$, and $1 \leq i \leq k$ such that $a, a+d, a+2d, \ldots, a+(p-1)d \in N_i$. Not surprisingly, the two-colour function function $W(p)$ has been studied most, and by now it is known that $W(2) = 3$, $W(3) = 9$, $W(4) = 35$ and $W(5) = 178$. However, very little is known about the growth of the functions $W(p)$ and $W(p, k)$.

Rather than proving van der Waerden's theorem directly, we shall deduce it from a remarkable extension of the theorem proved by Hales and Jewett in 1963. In order to state it, we need some definitions.

For a finite set A and integer n the *cube* of dimension $n \geq 1$ over the *alphabet* A is the set $A^n = A^{[n]} = \{(a_1, \ldots, a_n) : a_i \in A \text{ for every } i\}$. A *combinatorial line,* or simply a *line,* in A^n is a set L of the form

$$L = \{(a_1, \ldots, a_n) \in A^n : a_i = a_j \text{ for } i, j \in I \text{ and } a_i = a_i^0 \text{ for } i \notin I\},$$

where I is a non-empty subset of $[n]$ and a_i^0 is a fixed element of A for $i \in [n] - I$. Note that every line in A^n has precisely $|A|$ elements. Taking $A = [p]$, as we often do, the 'points' $a^1, a^2, \ldots, a^p$ of a line can be renumbered in such a way that $a^j = (a_1^j, a_2^j, \ldots, a_n^j)$ satisfies

$$a_i^j = \begin{cases} j & \text{if } i \in I, \\ a_i^0 & \text{if } i \notin I. \end{cases}$$

Clearly, every line has p elements, and there are

$$\sum_{I \subset [n], I \neq \emptyset} p^{n-|I|} = \sum_{\mathcal{J} \subset [n], \mathcal{J} \neq [n]} p^{|\mathcal{J}|} = (p+1)^n - p^n$$

lines in A^n. For example, the cube $[p]^2$ of dimension 2 has $2p+1$ lines: p 'vertical' lines, p 'horizontal' lines, and one 'diagonal' line, namely $\{(a, a) : a \in A\}$.

Theorem 23 *For every p and k, there is an integer n such that if A is an alphabet with p letters then every k-colouring $c : A^n \to [k]$ contains a monochromatic line.*

The *Hales–Jewett function* $HJ(p,k)$ is defined much like the corresponding van der Waerden function: $HJ(p,k)$ is the minimal value of n that will do in Theorem 23.

To see that the Hales–Jewett theorem implies van der Waerden's theorem, all we need is a map $\theta : [p]^n \to [p^n]$ sending every combinatorial line onto an arithmetic progression of length p. For example, we could take $\theta : [p]^n \to [p^n]$ given by $(a_i)_1^n \mapsto a_1 + \cdots + a_n$ or $(a_i)_1^n \mapsto 1 + \sum_{i=1}^n (a_i - 1)p^{i-1}$. Then, every k-colouring c of $[p^n]$ induces a k-colouring $\tilde{c}$ of the cube $[p]^n$ by setting $\tilde{c}((a_i)_1^n) = c(\theta((a_i)_1^n))$. Now, a monochromatic line in the colouring $\tilde{c}$ is mapped onto a monochromatic arithmetic progression of length p in the colouring c of $[p^n]$. Hence, $W(p,k) \le p^{HJ(p,k)}$.

The original proofs of Theorems 21 and 22 used double induction (on p and k) and not even for $W(p) = W(p,2)$ did they provide a primitive recursive upper bound, so the bound grew remarkably fast. In fact, the upper bound for $W(p)$ grew like *Ackerman's function* $A(p)$. To define $A(p)$, first define $f_1, f_2, \ldots : \mathbb{N} \to \mathbb{N}$ by setting $f_1(m) = 2m$ and

$$f_{n+1}(m) = \underbrace{f_n \circ f_n \circ \cdots \circ f_n}_{m}(1).$$

In particular, $f_2(1) = f_1(1) = 2$, $f_2(2) = f_1 \circ f_1(1) = f_1(2) = 2^2$, $f_2(3) = f_1 \circ f_1 \circ f_1(1) = f_1(2^2) = 2^3$, and so on, so that $f_2(m) = 2^m$; then $f_3(1) = 2$, $f_3(2) = f_2 \circ f_2(2) = f_2(2^2) = 2^{2^2}$, $f_3(3) = f_2(2^{2^2}) = 2^{2^{2^2}}$, and so on. Then Ackerman's function is $A(p) = f_p(p)$. For obvious reasons, f_3 is known as a 'tower' function. At the end of Section IV. 5 we already encountered this tower function in connection with Szemerédi's regularity lemma.

The breakthrough in the upper bounds for van der Waerden's function came over 60 years after van der Waerden proved his theorem, when Shelah, in a remarkable tour de force, gave a primitive recursive upper bound for the Hales–Jewett function $HJ(p,k)$ and so for the van der Waerden function $W(p,k)$. The main aim of this section is to present this beautiful theorem of Shelah.

Let us start with a technical lemma, known as Shelah's pigeon-hole principle.

Lemma 24 *Given integers n and k, if m is large enough, the following assertion holds. Let $c_j : [m]^{[2n-1]} \to [k]$, $j = 1, \ldots, n$, be k-colourings. Then there are integers $1 \le a_j < b_j \le m$ such that for every j, $1 \le j \le n$, we have*

$$\begin{aligned} &c_j(a_1, b_1, \ldots, a_{j-1}, b_{j-1}, a_j, a_{j+1}, b_{j+1}, \ldots, a_n, b_n) \\ &\quad = c_j(a_1, b_1, \ldots, a_{j-1}, b_{j-1}, b_j, a_{j+1}, b_{j+1}, \ldots, a_n, b_n). \end{aligned}$$

Proof. Let us apply induction on n. For $n = 1$ we may take any $m \ge k+1$. Suppose now that m_0 will do for n and k, and let us prove that any $m \ge k^{m_0^{2n}} + 1$ will do for $n+1$ and k.

Given colourings $c_j : [m_0]^{[2n+1]} \to [k]$, $j = 1, \ldots, n+1$, induce a colouring $c : [m] \to [k]^{[m_0]^{[2n]}}$ by setting, for every $a \in [m]$ and $(a_1, b_1, \ldots, a_n, b_n) \in$

$[m_0]^{[2n]}$,

$$c_a(a_1, b_1, \ldots, a_n, b_n) = c_{n+1}(a_1, b_1, \ldots, a_n, b_n, a).$$

By our choice of m, there are $1 \le a_{n+1} < b_{n+1} \le m$ such that $c_{a_{n+1}} = c_{b_{n+1}}$; that is,

$$c_{n+1}(a_1, b_1, \ldots, a_n, b_n, a_{n+1}) = c_{n+1}(a_1, b_1, \ldots, a_n, b_n, b_{n+1})$$

for all $(a_1, b_1, \ldots, a_n, b_n) \in [m_0]^{[2n]}$. Now, for $j = 1, \ldots, n$, define c'_j : $[m_0]^{[2n-1]}$,

$$c'_j(x_1, \ldots, x_{2n-1}) = c_j(x_1, \ldots, x_{2n-1}, a_{n+1}, b_{n+1}).$$

By the induction hypothesis, there are $1 \le a_i < b_i \le m_0$ for $i = 1, \ldots, n$ such that

$$\begin{aligned} &c'_j(a_1, b_1, \ldots, a_{j-1}, b_{j-1}, a_j, a_{j+1}, b_{j+1}, \ldots, a_n, b_n) \\ &\quad = c'_j(a_1, b_1, \ldots, a_{j-1}, b_{j-1}, b_j, a_{j+1}, b_{j+1}, \ldots, a_n, b_n) \end{aligned}$$

for every j. But then the numbers $1 \le a_i < b_i \le m$, $i = 1, \ldots, n+1$, have the required property. □

Writing $S(n, k)$ for the smallest value of m that will do in Lemma 24, the proof above shows that $S(1, k) = k + 1$ and

$$S(n+1, k) \le k^{S(n,k)^{2n}} + 1$$

for $n \ge 1$. We call $S(n, k)$ *Shelah's function.*

Before we turn to Shelah's theorem, let us reformulate Lemma 24 in a more convenient form. Given positive integers n and m, define

$$S = \{(a_1, b_1, \ldots, a_n, b_n) \in [m]^{(2n)}\}$$

and

$$S_0 = \{(a_1, b_1, \ldots, a_n, b_n) \in [m]^{[2n]} : a_i \le b_i \text{ for every } i\}.$$

We call S and S_0 the *Shelah subsets* of $[m]^{[2n]}$. Also, for $s = (a_1, b_1, \ldots, a_n, b_n) \in S$ and $1 \le j \le n$, set

$$s^{j,1} = (a_1, b_1, \ldots, a_{j-1}, b_{j-1}, a_j, a_j, a_{j+1}, b_{j+1}, \ldots, a_n, b_n)$$

and

$$s^{j,2} = (a_1, b_1, \ldots, a_{j-1}, b_{j-1}, b_j, b_j, a_{j+1}, b_{j+1}, \ldots, a_n, b_n).$$

With this notation, the proof of Lemma 24 gives the following assertion.

Lemma 25 *Given n and k, if $m \ge S(n, k)$ then the following assertions hold. Let S and S_0 be the Shelah subsets of $[m]^{[2n]}$ and let $c : S \to [k]$. Then there is a point $s \in S_0$ such that $c(s^{j,1}) = c(s^{j,2})$ for every j, $1 \le j \le n$.* □

After all this preparation, we are ready for Shelah's theorem, greatly strengthening the Hales–Jewett theorem.

Theorem 26 *Given integers p and k, there is a minimal integer* $n = HJ(p, k)$ *such that every k-colouring* $[p]^{[n]} \to [k]$ *contains a monochromatic line. Furthermore, if* $n = HJ(p, k)$ *then* $HJ(p+1, k) \le nS(n, k^{(p+1)^n})$.

Proof. Clearly, $HJ(1, k) = 1$. Let us assume that $n = HJ(p, k)$ exists and $m = S(n, k^{(p+1)^n})$. Let

$$c : [p+1]^{[nm]} \to [k]$$

be a k-colouring. We have to show that this k-colouring contains a monochromatic line.

Partition $[nm]$ into n intervals, each of length m:

$$[nm] = \bigcup_{j=1}^{n} I_j,$$

$I_j = [(j-1)m+1, jm]$. Also, let S and S_0 be the Shelah subsets of $[m]^{[2n]}$. For $s = (a_1, b_1 \dots, a_n, b_n) \in S_0, x = (x_1, \dots, x_n) \in [p+1]^{[n]}$ and $\ell = (j-1)m+i \in I_j$, set

$$e_\ell = \begin{cases} p+1 & \text{if } i \le a_j, \\ x_j & \text{if } a_j < i \le b_j, \\ p & \text{if } i > b_j. \end{cases}$$

We call $e_s(x) = e(s, x) = (e_1, \dots, e_{nm}) \in [p+1]^{[nm]}$ the *s-extension* of x. Clearly, for every $s \in S_0$, the map $e_s : [p+1]^{[n]} \to [p+1]^{[nm]}$ is an injection, mapping a line of $[p+1]^{[n]}$ onto a line of $[p+1]^{[nm]}$.

Let us use c to induce a colouring $\tilde{c} : S \to [k]^{[p+1]^{[n]}}$, $s \mapsto \tilde{c}_s$, where $\tilde{c}_s(x) = c(e(s, x))$ for $x \in [p+1]^{[n]}$. Since $m = S(n, k^{(p+1)^n})$, by Lemma 25 there is an $s \in S_0$ such that

$$c(e(s^{j,1}, x)) = c(e(s^{j,2}, x))$$

for all $x \in [p+1]^{[n]}$. Let us fix this point $s = (a_1, b_1, \dots, a_n, b_n) \in S_0$, and consider the k-colouring $[p]^{[n]} \to [k]$ given by $x \to c(e(s, x))$.

Since $n = HJ(p, k)$, the cube $[p]^{[n]}$ contains a monochromatic line. This means that, by renumbering the points, if necessary, there are points $x^1, \dots, x^p \in [p]^{[n]}$ and an interval $I = [h]$, $h \ge 1$, such that with $x^j = (x_1^j, \dots, x_n^j)$ we have

$$x_i^j = \begin{cases} j & \text{if } 1 \le i \le h, \\ x_i^0 & \text{if } h < i \le n, \end{cases}$$

where $x_{h+1}^0, \dots, x_n^0 \in [p]$.

Now define $x^{p+1} = (x_1^{p+1}, \dots, x_n^{p+1}) \in [p+1]^{[n]}$ to continue this sequence:

$$x_i^{p+1} = \begin{cases} p+1 & \text{if } 1 \le i \le h, \\ x_i^0 & \text{if } h < i \le n. \end{cases}$$

Then $\{x^1, \ldots, x^{p+1}\}$ is a line in $[p+1]^{[n]}$, and so $\{e_s(x^1), \ldots, e_s(x^{p+1})\}$ is a line in $[p+1]^{[nm]}$. To complete the proof, all we have to check is that this line is monochromatic, that is,

$$c(e(s, x^{p+1})) = c(e(s, x^p)). \tag{5}$$

We prove (5) by a telescoping argument. For $0 \le j \le h$, define $y^j = (y_1^j, \ldots, y_n^j) \in [p+1]^{[n]}$ by

$$y_i^j = \begin{cases} p+1 & \text{if } 1 \le i \le j, \\ p & \text{if } j < i \le h, \\ x_i^0 & \text{if } h < i \le n, \end{cases}$$

so that $y^0 = x^p$ and $y^h = x^{p+1}$.

Note that, for every j,

$$e(s, y^j) = e(s^{j+1,1}, y^{j+1}) = e(s^{j,2}, y^j)$$

and

$$c(e(s^{j,1}, y^j)) = c(e(s^{j,2}, y^j)).$$

Hence

$$c(e(s, y^0)) = c(e(s^{1,1}, y^1)) = c(e(s^{1,2}, y^1)) = c(e(s, y^1)).$$

Similarly, $c(e(s, y^1)) = c(e(s, y^2)) = \cdots = c(e(s, y^h))$, so $c(e(s, x^p)) = c(e(s, x^{p+1}))$. Thus (5) holds, and we are done. □

Theorem 26 implies that the Hales-Jewett function does not grow anywhere near as fast as Ackerman's function: in fact, for some constant c we have $HJ(p,k) \le f_4(c(p+k))$, where f_4 is the fourth function in the Ackerman hierarchy, the next function after the tower function $2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$.

The bound on the van der Waerden function $W(k)$ implied by Theorem 26 is, in fact, the best upper bound on $W(k)$ known at the moment.

The Hales–Jewett theorem and Shelah's theorem extend van der Waerden's theorem in an abstract, combinatorial direction. Van der Waerden's theorem also has many beautiful and deep extensions in the ring of integers: the starting point of these extensions is Rado's theorem concerning systems of linear equations. Let $A = (a_{ij})$ be an n by m matrix with integer entries. Call A *partition regular* if $Ax = 0$ has a monochromatic solution in every colouring of $\mathbb{N}$ with finitely many colours. In other words, $A = (a_{ij})$ is partition regular if for every partition $\mathbb{N} = \bigcup_{\ell=1}^{k} N_\ell$, one of the classes N_ℓ contains integers $x_1, \ldots, x_n$ such that

$$\sum_{j=1}^{n} a_{ij} x_j = 0$$

for $i = 1, \ldots, m$. Note that not every matrix is partition regular: for example a matrix with some positive and no negative entries is not partition regular, since

$0 \notin \mathbb{N}$, and neither is $A = (\,1 \ \ -2\,)$. On the other hand, Schur's theorem, Theorem 21 says that $(\,1 \ \ 1 \ \ -1\,)$ is partition regular and, as we shall see, van der Waerden's theorem is slightly weaker than the assertion that a certain matrix is partition regular. In 1933 Rado gave a remarkable characterization of partition regular matrices.

Let us write $a_1, \ldots, a_n \in \mathbb{Z}^m$ for the column vectors of $A = (a_{ij})$. Thus $a_j = (a_{1j}, a_{2j}, \ldots, a_{mj})^T$. We say that A satisfies the *columns condition* if, by renumbering the column vectors if necessary, there are indices $1 < n_1 < \cdots < n_\ell = n$ such that, with $b_i = \sum_{j=1}^{n_i} a_j$, the vector b_1 is 0, and for $i > 1$ the vector b_i is a rational linear combination of the vectors $a_1, a_2, \ldots, a_{n_{i-1}}$, that is, b_i is in the linear subspace of $\mathbb{Q}^m$ spanned by the set $\{a_1, a_2, \ldots, a_{n_{i-1}}\}$. For example, if $A = (a_1 \cdots a_m)$ with $a_i > 0$ for every i, then A satisfies the columns condition iff some collection of the a_i sums to 0.

Here then is Rado's partition regularity theorem.

Theorem 27 *A matrix with integer entries is partition regular if and only if it satisfies the columns condition.* □

This beautiful theorem reduces partition regularity to a property can be checked in finite time. It is worth remarking that neither of the two implications is easy. Also, as in most Ramsey type results, by the standard compactness argument we have encountered several times, the infinite version implies the finite version. Thus if A is partition regular then, for each k, there is a natural number $R = R(A, k)$ such that $Ax = 0$ has a monochromatic solution in every k-colouring of $[R]$.

In order to deduce Schur's theorem from Rado's theorem, all we have to notice is that $(1 \ 1 \ \ -1)$ is partition regular, since it satisfies the columns condition. Furthermore, as a matrix of the type

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix}$$

satisfies the columns condition, Rado's theorem implies that for integers p and k there is an integer n such that every k-colouring of $[n]$ contains a monochromatic arithmetic progression of length p whose difference is also in the same colour class: this is a little more than van der Waerden's theorem (see Exercise 44).

Yet another immediate consequence of Rado's theorem is that, given integers k and n, there exists $N = N(k, n)$ such that if $[N]$ is k-coloured then there is a set A of n natural numbers such that $\sum_{a \in A} a \leq N$ and all the sums $\sum_{b \in B} b$, $\emptyset \neq B \subset A$ have the same colour. At the end of the next section we shall discuss a beautiful extension of this to infinite sets.

In conclusion, let us say a few words about Szemerédi's theorem. In the 1930s, Erdős and Turán conjectured the far-reaching extension of van der Waerden's theorem that the *largest* colour class will do: it suffices to know that our set is 'large' rather than a part of a partition. To be precise, they conjectured that every set of natural numbers with positive upper density contains arbitrarily long arithmetic progressions: if $A \subset \mathbb{N}$ is such that $\limsup_{A\to\infty} |A \cap [N]|/N > 0$, then A contains arbitrarily long arithmetic progressions. By the standard compactness argument, this means that if $\delta > 0$ and $p \geq 1$ then there exists an N such that every subset of $[N]$ with at least δN elements contains an arithmetic progression of length p.

The first evidence for the truth of the Erdős–Turán conjecture was provided by Roth in 1953, when he proved the conjecture in the special case $p = 3$. The full conjecture was proved by Szemerédi in 1975, by a deep and intricate combinatorial argument. It is in this proof that Szemerédi needed his regularity lemma, presented in Chapter IV. Section 5. As we saw there, this result revolutionized extremal graph theory.

In fact, Szemerédi's theorem greatly influenced ergodic–theory as well: in 1977, Fürstenberg gave an ergodic theoretic new proof of the theorem, and thereby revitalized ergodic theory. But all that is well beyond our brief.

VI.5 Subsequences

Let (f_n) be a sequence of functions on a space T. Then we can find an infinite subsequence (g_n) such that one of the following two alternatives holds:

a if (h_n) is any subsequence of (g_n), then $\sup_{t\in T} |\sum_1^N h_n(t)| \geq 1/N$ for every $N \geq 1$,

b if (h_n) is any subsequence of (g_n), then $\sup_{t\in T} |\sum_1^N h_n(t)| < 1/N$ for every $N \geq 1$.

This rather difficult assertion about sequences of functions is, in fact, an immediate consequence of a Ramsey-type result about infinite sets.

As usual, given a set M, we write 2^M for the set of subsets of M, $M^{(r)}$ for the set of r-tuples of M, and $M^{(\omega)}$ for the set of countably infinite subsets of M. In view of Theorem 4, it is natural to ask whether every red–blue colouring of $\mathbb{N}^{(\omega)}$ contains an infinite monochromatic set. It does not take long to realize that this is not the case (see Exercise 45). Motivated by this observation, call a family $\mathcal{F} \subset 2^{\mathbb{N}}$ *Ramsey* if there exists an $M \in \mathbb{N}^{(\omega)}$ such that either $M^{(\omega)} \subset \mathcal{F}$ or $M^{(\omega)} \subset 2^{\mathbb{N}} - \mathcal{F}$. In other words, if a red–blue colouring of $\mathbb{N}^{(\omega)}$ contains an infinite monochromatic set, then we say that the collection $\mathcal{F}$ of red elements of $\mathbb{N}^{(\omega)}$ is Ramsey.

Of course, $2^{\mathbb{N}}$ can be identified with the Cartesian product $\prod_{i\in\mathbb{N}} T_n$, where $T_n = \{0, 1\}$ for all n. We give T_n the discrete topology and the product $2^{\mathbb{N}}$ the product topology: in this topology $2^{\mathbb{N}}$ is a compact Hausdorff space. A weak form of a theorem due to Galvin and Prikry states that open subsets of $2^{\mathbb{N}}$ are Ramsey. (This is easily seen to imply the above assertion about sequences of functions: see Exercise 46.) To prove this result it is convenient to use the notation and terminology introduced by Galvin and Prikry. We use M, N, A and B to refer to infinite subsets of $\mathbb{N}$, and X and Y for finite subsets of $\mathbb{N}$. We write $X < a$ if $x < a$ for every $x \in X$; $X < M$ means that $X < m$ for every $m \in M$. An *M-extension* of X is a set of the form $X \cup N$, where $X < N$ and $N \subset M$. Let us fix now a family $\mathcal{F} \subset 2^{\mathbb{N}}$. We say that M *accepts* X if every M-extension of X belongs to $\mathcal{F}$; M *rejects* X if no $N \subset M$ accepts X.

Lemma 28 *If $\mathbb{N}$ rejects $\emptyset$ then there exists an $M \in \mathbb{N}^{(\omega)}$ that rejects every $X \subset M$.*

Proof. Note first that there is an M_0 such that every $X \subset M_0$ is either accepted or rejected by M_0. Indeed, put $N_0 = \mathbb{N}$, $a_0 = 1$. Suppose that we have defined $N_0 \supset N_1 \supset \cdots \supset N_k$ and $a_i \in N_i - N_{i+1}$, $0 \le i \le k-1$. Pick $a_k \in N_k$. If $N_k - \{a_k\}$ rejects $\{a_0, \ldots, a_k\}$ then put $N_{k+1} = N_k - \{a_k\}$; otherwise, let N_{k+1} be an infinite subset of $N_k - \{a_k\}$ that accepts $\{a_0, \ldots, a_k\}$. Then $M_0 = \{a_0, a_1 \ldots\}$ will do.

By assumption M_0 rejects $\emptyset$. Suppose now that we have chosen $b_0, b_1, \ldots, b_{k-1}$ such that M_0 rejects every $X \subset \{b_0, b_1, \ldots, b_{k-1}\}$. Then M_0 cannot accept infinitely many sets of the form $X \cup \{c_j\}$, $j = 1, 2, \ldots$, since otherwise $\{c_1, c_2, \ldots\}$ accepts X. Hence M_0 rejects all but finitely many sets of the form $X \cup \{c\}$. As there are only 2^k choices for X, there exists a b_k such that M_0 rejects every $X \subset \{b_0, b_1, \ldots, b_k\}$. By construction the set $M = \{b_0, b_1, \ldots\}$ has the required property. □

Armed with Lemma 28, it is easy to prove the promised theorem of Galvin and Prikry.

Theorem 29 *Every open subset of $2^{\mathbb{N}}$ is Ramsey.*

Proof. Let $\mathcal{F} \subset 2^{\mathbb{N}}$ be open and assume that $A^{(\omega)} \not\subset \mathcal{F}$ for every $A \in \mathbb{N}^{(\omega)}$, i.e., $\mathbb{N}$ rejects $\emptyset$. Let M be the set whose existence is guaranteed by Lemma 28. If $M^{(\omega)} \not\subset 2^{\mathbb{N}} - \mathcal{F}$, let $A \in M^{(\omega)} \cap \mathcal{F}$. Since $\mathcal{F}$ is open, it contains a neighbourhood of A, so there is an integer $a \in A$ such that if $B \cap \{1, 2, \ldots, a\} = A \cap \{1, 2, \ldots, a\}$ then $B \in \mathcal{F}$. But this implies that M accepts $A \cap \{1, 2, \ldots, a\}$, contrary to the choice of M. Hence $M^{(\omega)} \subset 2^{\mathbb{N}} - \mathcal{F}$, proving that $\mathcal{F}$ is Ramsey. □

Roughly speaking, Theorem 29 tells us that sets 'insensitive to small changes' are Ramsey; as Exercise 45 shows, sets 'sensitive to small changes' need not be Ramsey.

We now set out to show that Theorem 29 leads to an elegant extension of Theorem 4. Denote by $X^{(<\omega)}$ the family of *finite* subsets of X. A family $\mathcal{G} \subset \mathbb{N}^{(<\omega)}$ is *dense* if $\mathcal{G} \cap M^{(<\omega)} \neq \emptyset$ for every $M \in \mathbb{N}^{(\omega)}$, and it is *thin* if no member of $\mathcal{G}$ is an initial segment of another member (that is, if $X < Y$ implies $X \notin \mathcal{G}$ or $X \cup Y \notin \mathcal{G}$). For example, for each $r = 1, 2, \ldots$, the family $\mathbb{N}^{(r)}$ is thin.

Corollary 30 *Let $\mathcal{G} \subset \mathbb{N}^{(<\omega)}$ be dense. Then there is an $M \in N^{(\omega)}$ such that every $A \subset M$ has an initial segment belonging to $\mathcal{G}$.*

Proof. Let $\mathcal{F} = \{F \subset \mathbb{N} : F$ has an initial segment belonging to $\mathcal{G}\}$. Then $\mathcal{F}$ is open, so there is an $M \in \mathbb{N}^{(\omega)}$ such that either $M^{(\omega)} \subset \mathcal{F}$, in which case we are done, or else $M^{(\omega)} \subset 2^{\mathbb{N}} - \mathcal{F}$. The second alternative cannot hold since it would imply $M^{(<\omega)} \cap \mathcal{G} = \emptyset$. □

This corollary enables us to deduce a major extension of the original Ramsey theorem for infinite sets (Theorem 3).

Corollary 31 *Let $\mathcal{G} \subset \mathbb{N}^{(<\omega)}$ be a thin family, and let $k \in \mathbb{N}$. Then for any k-colouring of $\mathcal{G}$ there is an infinite set $A \subset \mathbb{N}$ such that all members of $\mathcal{G}$ contained in A have the same colour.*

Proof. It clearly suffices to prove the result for $k = 2$. Consider a red and blue colouring of $\mathcal{G} : \mathcal{G} = \mathcal{F}_{\text{red}} \cup \mathcal{F}_{\text{blue}}$. If F_{red} is dense then let M be the set guaranteed by Corollary 30. For every $F \in \mathcal{G} \cap 2^M$ there is an infinite set $N \subset M$ with initial segment F. Since $\mathcal{G}$ is thin, F is the unique initial segment of N that belongs to $\mathcal{G}$. Hence $F \in \mathcal{F}_{\text{red}}$, so every member of $\mathcal{G}$ contained in M is red.

On the other hand, if $\mathcal{F}_{\text{red}}$ is not dense, then $2^M \cap \mathcal{F}_{\text{red}} = \emptyset$ for some infinite set M. Hence $2^M \cap \mathcal{G} \subset \mathcal{F}_{\text{blue}}$. $\square$

Let us now turn to the result concerning monochromatic sums we promised at the end for the previous section. This beautiful result, conjectured by Graham and Rothschild and first proved by Hindman, is not very near to the other results given in this section, but the striking proof given by Glazer illustrates the rich methods that can be applied in infinite Ramsey theory.

Theorem 32 *For any k-colouring of $\mathbb{N}$ there is an infinite set $A \subset \mathbb{N}$ such that all sums $\sum_{x \in X} x$, $\emptyset \neq X \subset A$, have the same colour.*

Proof. We shall not give a detailed proof but only sketch one for those who are (at least vaguely) familiar with ultrafilters on $\mathbb{N}$ and know that the set $\beta\mathbb{N}$ of all ultrafilters is a compact topological space (the Stone–Čech compactification of the discrete space $\mathbb{N}$). The proof, which is due to Glazer, is at least as beautiful as the theorem and is considerably more surprising. Let us recall that a filter $\mathcal{F}$ on $\mathbb{N}$ is a non-empty collection of subsets of $\mathbb{N}$ such that (i) if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$, (ii) if $A \in \mathcal{F}$ and $A \subset B$ then $B \in \mathcal{F}$ and (iii) $\mathcal{F}\mathbb{N}e2^{\mathbb{N}}$, that is, $\emptyset \notin \mathcal{F}$. Zorn's lemma implies that every filter is contained in a maximal filter, called an *ultrafilter*. If $\mathcal{U}$ is an ultrafilter, then for every $A \subset \mathbb{N}$ either $A \in \mathcal{U}$ or else $\mathbb{N} - A \in \mathcal{U}$. This implies that every ultrafilter $\mathcal{U}$ defines a finitely additive 0–1 measure m on $w^{\mathbb{N}}$:

$$m(A) = \begin{cases} 1 & \text{if } A \in \mathcal{U} \\ 0 & \text{if } \mathbb{N} - A \in \mathcal{U}. \end{cases}$$

Conversely, clearly every finitely additive 0–1 measure on $2^{\mathbb{N}}$ defines an ultrafilter. If there is a *finite* set of measure 1, then one of the elements, say a, of that set also has measure 1, and so $\mathcal{U} = \{A \subset \mathbb{N} : a \in A\}$. These ultrafilters are called *principal.* Not every ultrafilter is principal: the ultrafilters containing the filter $\mathcal{F} = \{A \subset N : \mathbb{N} - A \text{ is finite}\}$ are not principal.

That ultrafilters can be useful in proofs of Ramsey theorems can be seen from the following very simple proof of the case $r = 2$ of Theorem 3. Fix a nonprincipal ultrafilter $\mathcal{U}$. Let $N^{(2)} = P_1 \cup P_2 \cup \cdots \cup P_k$. For $n \in \mathbb{N}$ let $A_1^{(n)} = \{m : (n, m) \in P_i\}$. Then exactly one of the sets $A_1^{(n)}, A_2^{(n)}, \ldots, A_k^{(n)}$ belongs to $\mathcal{U}$, say the set $A_{c(n)}^{(n)}$.

Now, with $B_i = \{n : c(n) = i\}$ we have $\mathbb{N} = B_1 \cup \cdots \cup B_k$, so again exactly one of these sets, say B_j, belongs to $\mathcal{U}$. Finally, pick $a_1 \in B_j$, $a_2 \in B_j \cap A_j^{a_1}$, $a_3 \in B_j \cap A_j^{a_1} \cap A_j^{a_2}$, etc. With $A = \{a_1, a_2, \ldots\}$ we have $A^{(2)} \subset P_j$.

Let us turn at last to Glazer's proof of Theorem 32. Let us define an addition on $\beta\mathbb{N}$ by

$$\mathcal{U} + \mathcal{V} = \{A \subset \mathbb{N} : \{n \in \mathbb{N} : A - n \in \mathcal{U}\} \in \mathcal{V}\}$$

where $\mathcal{U}, \mathcal{V} \in \beta\mathbb{N}$ and $A - n = \{a - n : a \in A, a > n\}$.

With some effort one can check that $\mathcal{U} + \mathcal{V}$ is indeed an ultrafilter and that with this addition $\beta\mathbb{N}$ becomes a semigroup. Furthermore, the semigroup operation is right-continuous, i.e., for a fixed $\mathcal{V} \in \beta\mathbb{N}$ the map $\beta\mathbb{N} \to \beta\mathbb{N}$, given by $\mathcal{U} \to \mathcal{V}+\mathcal{U}$, is continuous. By applying a short and standard topological argument we see that the properties above imply that $\beta\mathbb{N}$ has an idempotent element, that is,an element $\mathcal{P}$ with $\mathcal{P} + \mathcal{P} = \mathcal{P}$. This $\mathcal{P}$ is nonprincipal, since if $\{p\} \in \mathcal{P}$ then $\{2p\} \in \mathcal{P} + \mathcal{P}$ so $\{p\} \notin \mathcal{P} + \mathcal{P}$.

Let now $A \in \mathcal{P}$. Then, by the definition of addition, the set

$$A^* = \{n \in \mathbb{N} : A - n \in \mathcal{P}\}$$

belongs to $\mathcal{P}$. Thus if $a \in A \cap A^*$ then $B = (A - a) \cap (A \backslash \{a\}) \in \mathcal{P}$. (We could replace A by $A \backslash \{a\}$ since $\mathcal{P}$ is not principal.) Hence for every $A \in \mathcal{P}$ there exists $a \in A$ and $B \subset A \backslash \{a\}$ such that $B \in \mathcal{P}$ and $a + B \subset A$.

Of course, this ultrafilter $\mathcal{P}$ has nothing to do with any colouring of $\mathbb{N}$. However, just as any nonprincipal ultrafilter enabled us to find a monochromatic infinite set in a direct way, this idempotent $\mathcal{P}$ enables us to find an appropriate infinite set. Let $\mathbb{N} = C_1 \cup \cdots \cup C_k$ be the decomposition of $\mathbb{N}$ into colour classes. Exactly one of these colour classes, say C_i, belongs to $\mathcal{P}$. Put $A_1 = C_i$. Select $A_1 \in A_1$ and $A_2 \in \mathcal{P}$, $A_2 \subset A_1 - \{a_1\}$ such that $a_1 + A_2 \subset A_1$. Then select $A_2 \in A_2$ and $A_3 \in \mathcal{P}$, $A_3 \subset A_2 - \{a_2\}$ such that $a_2 + A_3 \subset A_2$, etc. The set $A = \{a_1, a_2, \ldots\}$ clearly has the required property: every infinite sum $\sum_{x \in X} x$, $X \subset A$, has colour i. □

Finally, it should be emphasized that the infinite Ramsey results presented in this section form only the tip of an iceberg: the Ramsey theory of infinite sets, called *partition calculus*, is an essential and very cultivated branch of set theory, and it has a huge literature.

VI.6 Exercises

1.⁻ Prove that every 2-colouring of the edges of K_n contains a monochromatic spanning tree (cf. Exercise I.1).

2. Prove that $R(3, 4) = 9$ (see Fig. VI.2).

3. Extending the construction in Fig. VI.2, find for each $t \geq 2$ a t-regular graph that shows that $R(3, t + 1) > 3t - 1$.

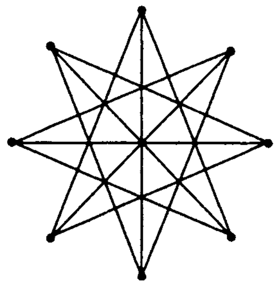


FIGURE VI.2. A graph showing that $R(3, 4) > 8$.

4. By considering the graph with vertex set $\mathbb{Z}_{17}$ (the integers modulo 17) in which i is joined to j iff $i - j = \pm 1, \pm 2, \pm 4$ or ± 8, show that $R(4, 4) = 18$.

5. Prove that $R(3, 5) = 14$.

6.$^+$ Let e be an edge of K_4. Show that $r(K_4 - e, K_4) = 11$.

7. By considering the 3-colouring of K_{16} with vertex set $GF(16)$, the field of order 16, in which the colour of an edge ij depends on the coset of the group of the cubic residues to which $i - j$ belongs, show that $R_3(3, 3, 3) = 17$. (Remember to check that the graph is well defined.)

8. Establish the upper bound for $R_k^{(r)}(s_1, \ldots, s_k)$ given after Theorem 2.

9. Give a direct proof of the result of Erdős and Szekeres in Exercise II.54 that, for all $k, \ell \geq 1$, every sequence of $k\ell + 1$ real numbers contains an increasing subsequence of $k + 1$ terms or a decreasing subsequence of $\ell + 1$ terms. [Hint. Imitate the proof of Theorem 3: consider the last elements of increasing subsequences with k terms.]

10. Show that if $x_1 < x_2 < \cdots < x_n$ are $n = \binom{k+\ell-4}{k-2} + 1$ real numbers, then we can find either $1 \leq n_1 < \cdots < n_k \leq n$ with $x_{n_2} - x_{n_1} \leq x_{n_3} - x_{n_2} \leq \cdots \leq x_{n_k} - n_{n_{k-1}}$ or else $1 \leq m_1 < \cdots < m_\ell \leq n$ with $x_{m_2} - x_{m_1} \geq x_{m_3} - x_{m_2} \geq \cdots \geq x_{m_\ell} - x_{m_{\ell-1}}$. Give two proofs: in the first, consider the numbers in the two intervals $[x_1, (x_1 + x_n)/2]$ and $[(x_1 + x_n)/2, x_n]$, and in the second imitate the proof of Theorem 3.

11. Note that the proof of Theorem 3 gives the following common extension of the upper bound in Theorem 3 and the assertion in the previous exercise. Let $k, \ell \geq 2$, $n = \binom{k+\ell-4}{k-2} + 1$, and $w : [n]^{(2)} \to \mathbb{R}$. Then either $w(n_1 n_2) \leq w(n_2 n_3) \leq \cdots \leq w(n_{k-1} k_k)$ for some $1 \leq n_1 < n_2 < \cdots < n_k \leq n$, or else $w(m_1 m_2) \geq w(m_2 m_3) \geq \cdots \geq w(m_{\ell-1} m_\ell)$ for some $1 \leq m_1 < m_2 < \cdots < m_\ell \leq n$. Note also that this need not hold for smaller values of n.

12.$^-$ Given $2 \leq k \leq n$, denote by $c_k(n)$ the maximal integer that is such that in every k-colouring of the edges of K_n we can find a connected monochromatic subgraph of order $c_k(n)$. Show that $c_2(n) = n$. (See Exercise 1.)

13.$^-$ (Exercise 12$^-$ contd.) Prove that $c_{n-1}(n) = 2$ if $n \geq 2$ is even and $c_{n-1}(n) = 3$ if $n \geq 3$ is odd. [*Hint.* Use Theorem I.9.]

14.$^+$ (Exercises 12$^-$ and 13$^-$ contd.) Prove that

$$c_3(n) = \begin{cases} \dfrac{n}{2} + 1 & \text{if } n \equiv 2 \pmod 4, \\ \left\lfloor \dfrac{n}{2} \right\rfloor & \text{otherwise.} \end{cases}$$

15. Check that $r(2K_3, K_3) = 8$ and $r(2K_3, 2K_3) = 10$.

16. Show that $r(C_4, C_4) = 6$.

17.$^+$ Using two-dimensional vector spaces over finite fields (cf. Theorem IV.12), show that

$$r_k(C_4, C_4, \ldots, C_4) = k^2 + O(k).$$

18. By considering $H_1 = P_5$ and $H_2 = K_{1,3}$, show that

$$r(H_1, H_2) \geq \min_{i=1,2} r(H_i, H_i)$$

need not hold.

19. Let H_p, H_q be graphs of order p and q, respectively, and let $\alpha(H_p) = i$, $\alpha(H_q) = j$. Then there is a constant C depending only on p and q such that, with $m(s, t) = \min\{si, tj\}$, we have

$$ps + qt - m(s, t) - 2 \leq r(sH_p, tH_q) \leq ps + qt - m(s, t) + C.$$

[*Hint.* Find a red $K_{j(p-i)}$, say R, a blue $K_{i(q-j)}$, say B, and a set N of ij other vertices such that the $R - N$ edges are red and the $B - N$ edges are blue. [Cf. the proof of Theorem 17.]

In the next four exercises, $f(n)$ is the minimal integer N such that whenever X is a set of N points in a plane, no three of which are collinear, X contains n points forming a convex n-gon.

20. Show that $f(3) = 3$ and $f(4) = 5$. Deduce that $f(n) \leq R^{(4)}(5, n)$ for every $n \geq n$.

21.$^+$ Prove that $f(5) = 9$.

22. Deduce from Theorem 3 that $f(n) \leq \binom{2n-4}{n-2} + 1$.

23.$^+$ Prove that $f(n) \geq 2^{n-2} + 1$; i.e., for every $n \geq 3$ there are 2^{n-2} points in the plane in general position, such that no n of them form a convex polygon. [*Hint.* Let $S_{k,\ell}$ be as in the proof of Theorem 2. For $i = 0, 1, \ldots, n-2$, let S_i be obtained from $S_{n+i,2-i}$ by flattening it, shrinking it, and finally translating it so that $S_0, S_1, \ldots, S_{n-2}$ are on an increasing circular arc, as in Fig. VI.3.

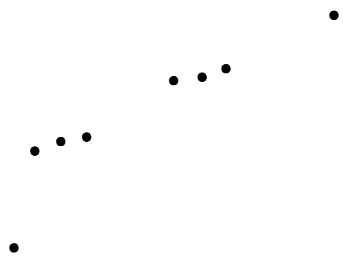


FIGURE VI.3. The case $n = 5$: a set of eight points without a convex pentagon.

Show that if the S_i are far enough apart and small enough then $\bigcup_{i=0}^{n-2} S_i$ does not contain a convex n-gon.]

24. Let S be an infinite set of points in the plane. Show that there is an infinite set $A \subset S$ such that either A is contained in a line or no three points of A are collinear.

25. Show that $R_k(3, 3, \ldots, 3) \leq \lfloor ek! \rfloor + 1$. [*Hint.* Note that $\lfloor ek! \rfloor = 1 + k\lfloor e(k-1)! \rfloor$.]

26. Show that there is an infinite set of natural numbers such that the sum of any two elements has an even number of prime factors, counted without multiplicity, say.

27. Show that there is a sequence $n_1 < n_2 < \cdots$ of natural numbers such that if $r \leq i_1 < i_2 < \cdots < i_r$, then $\sum_{j=1}^{r} n_{i_j}$ has an even number of prime factors iff r has an odd number of prime factors.

28. Define a graph with vertex set $[N]^{(2)}$ by joining $a < b$ to $b < c$. Show that this graph does not contain a triangle and its chromatic number tends to infinity with N. (Cf. Exercise V.12.)

29. Check that every S-canonical colouring $c_S : \mathbb{N}^{(r)} \to \mathbb{N}^{(s)}$, defined before Theorem 8, is irreducible.

30. Let $g_1(x), g_2(x), \ldots, g_n(x)$ be bounded real functions and let $f(x)$ be another real function. Let ε and δ be positive constants. Suppose that $\max_i(g_i(x) - g_i(y)) > \delta$ whenever $f(x) - f(y) > \varepsilon$. Prove that f is bounded.

31. Prove that every 2-colouring of the edges of K_{3n-1} contains n independent edges of the same colour. Show also that the result is best possible: there is a 2-colouring of the edges of K_{3n-2} in which no set of n independent edges is monochromatic.

32. Show that for every ℓ there is a natural number n such that if $[n]$ is partitioned into two classes then $x_1 + \cdots + x_\ell = x_{\ell+1}$ is solvable in one class. Writing $n(\ell)$ for the smallest natural number that will do here, show that $n(2) = 5$ and $n(\ell) \geq \ell^2 + \ell - 1$. Determine $n(3)$.

33. Show that if $[9] = \{1, 2, \ldots, 9\}$ is partitioned into two classes then $x_1 + x_2 + 1 = x_3$ is solvable in one class.

34.$^+$ As usual, let $\mathcal{P}(n)$ be the collection of all 2^n subsets of $[n] = \{1, \ldots, n\}$. Show that if $\mathcal{P}(n)$ is coloured with n colours then there are sets $A, B \subset [n]$, $A \neq B$, such that $A, B, A \cup B$ and $A \cap B$ all have the same colour. Show that if we use $n + 1$ colours then the sets A, B need not exist.

35.$^+$ Show that every red–blue colouring of the edges of K_{6n} contains n vertex disjoint triangles with all $3n$ edges of the same colour.

36. Let $n = 2^p$. Show that K_{n+1} is not the union of p bipartite graphs but K_n is. Deduce that if there are $2^p + 1$ points in the plane, then some three of them determine an angle of size greater than $\pi(1 - (1/p))$.

37. Let $n = 2^p$, and let K_n be the union of the bipartite graphs $G_1, \ldots, G_p$. Show that $2^k - 1 \leq \sum_{i=1}^{k} d_{G_i}(x) \leq 2^p - 2^{p-k}$ for every k, $k = 1, \ldots, p$. [*Hint.* Note that each G_i has 2^{p-1} vertices in each of its classes.]

38. For $n \geq 2$, let $g(n)$ be the minimal value of N such that any N points in $\mathbb{R}^n$ contain three that determine an angle strictly greater than $\pi/2$. Prove that
   (i) $g(2) = 5$,
   (ii)$^{++}$ $g(3) = 9$,
   (iii) $g(n) \geq 2^n + 1$,
   (iv)$^+$ $g(n)$ is finite for every n.

39.$^{++}$ Prove that for every $n \geq 2$ and $\varepsilon > 0$ there is an N such that any N points in $\mathbb{R}^n$ contain three that determine an angle greater than $\pi - \varepsilon$.

40. Prove that for every $k \geq 1$ there is a natural number $n = n(k)$ such that if the subsets of an n-element set are k-coloured, then there are disjoint sets A, B such that A, B and $A \cup B$ all have the same colour and $|A| \neq |B|$.

41.$^+$ Let (x_n) be a sequence of unit vectors in a normed space. Show that (x_n) has a subsequence (y_n) such that if $(\lambda_i)_1^k \in \mathbb{R}^k$, with $\sum_1^k |\lambda_i| = 1$, then

$$\left| \| \sum_1^k \lambda_i y_{m_i} \| - \| \sum_1^k \lambda_i y_{n_i} \| \right| < 1/k$$

whenever $k < m_1 < \cdots < m_k$ and $k < n_1 < \cdots < n_k$.

42. (i) Show that every red–blue colouring of the edges of a $K_{3,3}$ contains a monochromatic path of length 3.
(ii) Show that every red–blue colouring of the edges of a $K_{2k-1,2k-1}$ contains a monochromatic tree of order $2k - 1$, with two vertices of degree k.

43. Show that the matrix $(\ 1 \quad -2\ )$ is not partition regular by constructing a red–blue-colouring of $\mathbb{N}$ in which, for every x, the numbers x and $2x$ have different colours.

44. Deduce from van der Waerden's theorem that for all p and k there is an integer n such that every k-colouring of $[n]$ contains a monochromatic arithmetic progression of length p whose difference belongs to the same colour class.
45. Deduce from Zorn's lemma that there is a minimal set $\mathcal{M} \subset \mathbb{N}^{(\omega)}$ such that for every $N \in \mathbb{N}^{(\omega)}$ there is a unique $M \in \mathcal{M}$ with the symmetric difference $M \Delta N$ finite. Colour N *red* if $M \Delta N$ has an even number of elements, and *blue* otherwise. Show that this red–blue colouring of $\mathbb{N}^{(\omega)}$ does not contain an infinite monochromatic set.
46. Deduce from Theorem 29 the assertion at the beginning of Section 5, concerning sequences of functions.

VI.7 Notes

There are two excellent major surveys of the field, giving a great many references: the book by R.L. Graham, B.L. Rothschild and J.H. Spencer, *Ramsey Theory*, 2nd ed., Wiley and Sons, New York, 1990, xi+196 pp., and the survey article by J. Nešetřil, Ramsey Theory, in *Handbook of Combinatorics*, vol.II (R.L. Graham, M. Grötschel and L. Lovász, eds), North-Holland, Amsterdam, 1995, pp. 1331–1403.

The fundamental Ramsey theorems are in F.P. Ramsey, On a problem of formal logic, *Proc. Lond. Math. Soc.* (2) **30** (1930) 264–286.

Theorem 1 and the first part of Theorem 3 are from P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compositio Math.* **2** (1935) 464–470, and the second part of Theorem 3 is from P. Erdős and G. Szekeres, On some extremum problems in elementary geometry, *Ann. Univ. Sci. Budapestin.* **3** (1960) 53–62. A.G. Thomason's improved bound on $R(s, t)$, inequality (5), was published in An upper bound for some Ramsey numbers, *J. Graph. Theory* **12** (1988) 509–518. The theorem of J. Paris and L. Harrington is from A mathematical incompleteness in Peano arithmetic, in *Handbook of Mathematical Logic* (ed. J. Barwise), North-Holland, pp. 1133–1142.

For numerous results concerning the approximation of small Ramsey numbers, see B.D. McKay and S. Radziszowski, Subgraph counting identities and Ramsey numbers, *J. Combinatorial Theory* Ser. B **69** (1997) 193–209.

The canonical theorem of P. Erdős and R. Rado is in A combinatorial theorem, *J. London Math. Soc.* **25** (1950) 249–255.

For graphical Ramsey theorems, see S.A. Burr, Generalized Ramsey theory for graphs – a survey, in *Graphs and Combinatorics* (R. Bari and F. Harary, eds) Springer-Verlag, 1974, pp. 52–75. For some of the deep results of Nešetřil and Rödl, one of which is Theorem 19, see J. Nešetřil and V. Rödl, The Ramsey property for graphs with forbidden subgraphs, *J. Combinatorial Theory* Ser. B, **20** (1976), 243–249, and Partitions of finite relational and set systems, *J. Combinatorial Theory* Ser. A **22** (1977), 289–312. Also, Theorem 12 is from V. Chvátal, Tree-complete Ramsey numbers, *J. Graph Theory* **1** (1977) p.93, and Theorem 13 is from Y. Li and C. Rousseau, Fan-complete graph Ramsey numbers, *J. Graph Theory* **23** (1996) 413–420. The paper by S.P. Radziszowski, Small Ramsey numbers, *The Electronic J. of Combinatorics* **1** (1994) 1–29, is a mine of up-to-date information.

B.L. van der Waerden's theorem is in Beweis einer Baudetschen Vermutung, *Nieuw Archief voor Wiskunde* **15** (1927) 212–216, and Theorem 23 is in A. Hales and R.I. Jewett, Regularity and positional games, *Trans. Amer. Math. Soc.* **106** (1963) 222–229. S. Shelah's extension of the Hales–Jewett theorem is in S. Shelah, Primitive recursive bounds for van der Waerden numbers, *J. Amer. Math. Soc.* **1** (1988) 683–697. The partition regularity theorem is from R. Rado, Studien zur Kombinatorik, *Math. Zeitschrift* **36** (1933) 424–480.

The first results of §5 are in F. Galvin and K. Prikry, Borel sets and Ramsey's theorem, *J. Symbolic Logic* (1973) 193–198; Glazer's proof of Hindman's theorem is in W.W. Comfort, Ultrafilters: some old and some new results, *Bull. Amer. Math. Soc.* **83** (1977) 417–455.

The foundations of partition calculus were laid down over thirty years ago by P. Erdős, A. Hajnal and R. Rado. Their main paper, popularly known as the 'giant triple paper', is Partition relations for cardinal numbers, *Acta Math. Acad. Sci. Hungar.* **16** (1965) 93–196; a detailed presentation of the theory is in the treatise by P. Erdős, A. Hajnal, A. Máté and R. Rado, *Combinatorial Set Theory: Partition Relations for Cardinals*, Studies in Logic and the Foundations of Mathematics, Akadémiai Kiadó/North Holland, Budapest/Amsterdam, 1984.

VII

Random Graphs

Although the theory of random graphs is one of the youngest branches of graph theory, in importance it is second to none. It began with some sporadic papers of Erdős in the 1940s and 1950s, in which Erdős used random methods to show the existence of graphs with seemingly contradictory properties. Among other results, Erdős gave an exponential lower bound for the Ramsey number $R(s, s)$; i.e., he showed that there exist graphs of large order such that neither the graph nor its complement contains a K_s. He also showed that for all natural numbers k and g there are k-chromatic graphs of girth at least g. As we saw in Chapters V and VI, the constructions that seem to be demanded by these assertions are not easy to come by. The great discovery of Erdős was that we can use probabilistic methods to demonstrate the existence of the desired graphs without actually constructing them. This phenomenon is not confined to graph theory and combinatorics: probabilistic methods have been used with great success in the geometry of Banach spaces, in Fourier analysis, in number theory, in computer science—especially in the theory of algorithms—and in many other areas. However, there is no area where probabilistic methods are more natural and lead to more striking results than in combinatorics.

In fact, random graphs are of great interest in their own right as well, not only as tools to attack problems that have nothing to do with probability theory or randomness.

We are asking the most basic questions: what do 'most' graphs in various families look like? Rather than being interested in the *extreme* values of our parameters, we wish to discover what happens *on average.* In addition to this, what makes the field so attractive and important is that more often than not the phenomena we discover are surprising and delicate.

The systematic study of random graphs for their own sake was started by Erdős and Rényi in 1959: in a series of papers they laid the foundations of a rich theory of random graphs, proving many of the fundamental results. Loosely speaking, Erdős and Rényi discovered that in the spaces they studied, there was a '*typical*' random graph: with high probability a random graph had certain sharply delineated properties. The other great discovery of Erdős and Rényi was that all the standard properties of graphs (being connected, having diameter at most 5, containing a complete graph of order 4, being Hamiltonian, etc.) arise rather suddenly: while a random graph with n vertices and a certain number of edges is unlikely to have the property at hand, a random graph with a few more edges is very likely to have the property. This phenomenon is described by a phrase borrowed from physics: there is a *phase transition.* The most dramatic example of a phase transition discovered by Erdős and Rényi concerns the order of the largest component of a random graph.

The contents of this chapter will reflect both aspects of the theory: we shall prove a number of basic results concerning the most frequently studied models of random graphs, and we shall use probabilistic methods to answer some important graph–theoretic questions that have nothing to do with randomness.

For many a problem one uses specifically tailored random models. For instance, remarkable successes have been achieved by arguments building on random colourings of graphs. Percolation theory is nothing more than the study of random subgraphs of various lattices. Also, many algorithms are based on the use of certain graphs, whose existence is most easily demonstrated by the use of random techniques. The importance of random graphs and random methods is due precisely to applications of this type.

For the sake of convenience, we state some simple inequalities that will be used in our calculations. For approximating factorials, we shall never need more than the following version of *Stirling's formula*:

$$\sqrt{2\pi s}(s/e)^s \le s! \le e^{1/12s}\sqrt{2\pi s}(s/e)^s. \tag{1}$$

In fact, in most cases it will suffice that $s! \ge 2\sqrt{s}(s/e)^s \ge (s/e)^s$, and so

$$\binom{n}{k} \le \frac{n^k}{k!} \le \frac{1}{2\sqrt{k}}\left(\frac{en}{k}\right)^k \le \left(\frac{en}{k}\right)^k. \tag{2}$$

We shall also use the inequality $1 - x \le e^{-x}$, so that

$$(1-x)^k \le e^{-kx} \tag{3}$$

for all $x < 1$ and $k \ge 0$.

VII.1 The Basic Models—The Use of the Expectation

Our first task is to make precise the notion of a 'random graph'. Rather trivially, every probability space whose points are graphs gives us a notion of a random graph. We shall concentrate on those probability spaces or models that arise most naturally and have been found to be most useful. Three closely related models stand out: $\mathcal{G}(n, M)$, $\mathcal{G}(n, p)$ and $\tilde{\mathcal{G}}^n$. In each case, the probability space consists of graphs on a fixed set of n distinguishable vertices: as usual, we take this set to be $V = [n] = \{1, 2, \ldots, n\}$. Note that the complete graph K_n on $[n]$ has $N = \binom{n}{2}$ edges and 2^N subgraphs.

For $0 \leq M \leq N$, the space $\mathcal{G}(n, M)$ consists of all $\binom{N}{M}$ subgraphs of K_n with M edges: we turn $\mathcal{G}(n, M)$ into a probability space by taking its elements to be equiprobable. Thus, writing $G_M = G_{n,M}$ for a random graph in the space $\mathcal{G}(n, M)$, the probability that G_M is precisely a fixed graph H on $[n]$ with M edges is $1/\binom{N}{M}$:

$$\mathbb{P}_M(G_M = H) = \binom{N}{M}^{-1}.$$

The space $\mathcal{G}(n, p)$, or $\mathcal{G}(n, \mathbb{P}(\text{edge}) = p)$, is defined for $0 \leq p \leq 1$. To get a random element of this space, we select the edges independently, with probability p. Putting it another way, the ground set of $\mathcal{G}(n, p)$ is the set of all 2^N graphs on $[n]$, and the probability of a graph H on $[n]$ with m edges is $p^m(1-p)^{N-m}$: each of the m edges of H has to be selected and none of the $N - m$ 'non-edges' of H is allowed to be selected. It is customary to write q for $1 - p$, the probability that an edge of K_n is not selected. Then, writing $G_p = G_{n,p}$ for a random element of $\mathcal{G}(n, p)$,

$$\mathbb{P}_p(G_p = H) = p^{e(H)} q^{N-e(H)}.$$

The space $\tilde{\mathcal{G}}^n$ is not a space of random graphs but a space of *sequences* of random graphs, one from each $\mathcal{G}(n, M)$. An element of $\tilde{\mathcal{G}}^n$ is a *graph process*, a nested sequence of graphs $G_0 \subset G_1 \subset \cdots \subset G_N$, with G_t having precisely t edges. Clearly, there are $N!$ graph processes $\tilde{G} = (G_t)_0^N$ on $[n]$, since every graph process $\tilde{G}$ is trivially identified with a permutation $(e_i)_1^N$ of the N edges of the complete graph K_n on $[n]$: this identification is given by $\{e_t\} = E(G_t) - E(G_{t-1})$. We turn $\tilde{\mathcal{G}}^n$, the set of all $N!$ graph processes, into a probability space by taking all processes to be equiprobable.

There is a pleasing interpretation of a random graph process $\tilde{G} = (G_t)_0^N \in \tilde{\mathcal{G}}^n$: it is a living organism that starts its life as the empty graph $G_0 = E_n$ and evolves by acquiring more and more edges, namely, at time t it acquires one more edge at random from among the $N - t$ possibilities.

In all these examples, we tend to be interested in what happens as $n \to \infty$. It is worth remarking that both $M = M(n)$ and $p = p(n)$ are functions of n. The space $\mathcal{G}(n, p)$ is of great interest for fixed values of p as well; in particular, $\mathcal{G}(n, 1/2)$ could be viewed as *the* space of random graphs of order n: it consists of all 2^N graphs on $[n]$, and all graphs are equiprobable. Thus $G_{n,1/2}$ is obtained by picking one of the 2^N graphs on $[n]$ at random. However, $\mathcal{G}(n, M)$ is not too exciting for a fixed value of M as $n \to \infty$, since then, with probability tending to 1, $G_{n,M}$ is just a set of M independent edges and $n - 2M$ isolated vertices (see Exercise 1).

The spaces $\mathcal{G}(n, M), \mathcal{G}(n, p)$ and $\tilde{\mathcal{G}}^n$ are closely related to each other. For example, the map $\tilde{\mathcal{G}}^n \to \mathcal{G}(n, M)$, given by $\tilde{G} = (G_t)_0^N \mapsto G_M$, is measure-preserving, so $\tilde{\mathcal{G}}^n$ 'couples' the spaces $\mathcal{G}(n, M)$, $M = 0, 1, \ldots, N$. Also, if in $\mathcal{G}(n, p)$ we condition on $e(G_p) = M$, then we obtain $\mathcal{G}(n, M)$. To get $\mathcal{G}(n, p)$ from $\tilde{\mathcal{G}}^n$, we pick a random element $\tilde{G} = (G_t)_0^N$ and take G_t, where t is a binomial random variable with parameters N and p, so that $\mathbb{P}(t = M) = \binom{N}{M} p^M q^{N-M}$. As we shall see later, for $M \sim pN$ the spaces $\mathcal{G}(n, M)$ and $\mathcal{G}(n, p)$ are close to each other.

Now that we have obtained a space of random graphs, every graph invariant becomes a random variable; the nature of such a random variable depends crucially on the space. For instance, the number $X_s(G)$ of complete graphs of order s in G is a random variable on our space of random graphs; whether it be $\mathcal{G}(n, M)$, $\mathcal{G}(n, p)$ or some other space.

In this section we shall confine ourselves to making use of the expectations of some basic random variables: it is surprising that even this minimal use of probability theory enables us to prove substantial results about graphs. As a first example, let us calculate the expectation of X_s. This will lead us quickly to the lower bound of Erdős on the Ramsey numbers. As earlier, we shall use the subscripts M and p to identify the space we are working in; thus $\mathbb{E}_M(X)$ denotes the expectation of the random variable X in the space $\mathcal{G}(n, M)$.

Analogously to $X_s(G)$, let $X_s'(G)$ be the number of independent sets of order s. Let us calculate very carefully the expectations of X_s and X_s' in $\mathcal{G}(n, M)$ and $\mathcal{G}(n, p)$. Let $\mathcal{S} = [n]^{(s)}$ be the set of s-subsets of $[n]$, and for $\alpha \in \mathcal{S}$ let Y_α be the indicator function of the complete graph K_α with vertex set α:

$$Y_\alpha(G) = \begin{cases} 1 & \text{if } G[\alpha] = K_\alpha, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$X_s(G) = \sum_{\alpha \in \mathcal{S}} Y_\alpha(G)$$

for every graph G on $[n]$. Similarly, writing Y_α' for the indicator function of E_α, the empty graph with vertex set α, we have $X_s'(G) = \sum_{\alpha \in \mathcal{S}} Y_\alpha'(G)$. Hence, no matter what probability space we take, by the additivity of the expectation,

$$\mathbb{E}(X_s) = \sum_{\alpha \in \mathcal{S}} \mathbb{E}(Y_\alpha) = \sum_{\alpha \in \mathcal{S}} \mathbb{P}(G[\alpha] = K_\alpha),$$

and a similar assertion holds for $\mathbb{E}(X_s')$.

Let us make use of this formula in the spaces $\mathcal{G}(n, M)$ and $\mathcal{G}(n, p)$. Starting with $\mathcal{G}(n, M)$,

$$\mathbb{E}_M(Y_\alpha) = \mathbb{P}_M(G_M[\alpha] = K_\alpha) = \binom{N - S}{M - S} \binom{N}{M}^{-1},$$

where $S = \binom{s}{2}$, since there are $\binom{N-S}{M-S}$ graphs of size M on $[n]$ that contain all S edges joining vertices of α. Similarly,

$$\mathbb{E}_M(Y'_\alpha) = \mathbb{P}_M(G_M[\alpha] = E_\alpha) = \binom{N-S}{M}\binom{N}{M}^{-1}.$$

The formulae are even simpler in $\mathcal{G}(n, p)$:

$$\mathbb{E}_p(Y_\alpha) = \mathbb{P}_p(G_p[\alpha] = K_\alpha) = p^S$$

and

$$\mathbb{E}_p(Y'_\alpha) = \mathbb{P}_p(G_p[\alpha] = E_\alpha) = q^S.$$

Since $|S| = \binom{n}{s}$, we have the following simple result.

Theorem 1 *Let $X_s = X_s(G)$ be the number of complete subgraphs of order s in G, and let $X'_s = X'_s(G) = X_s(\overline{G})$. Then*

$$\mathbb{E}_M(X_s) = \binom{n}{s}\binom{N-S}{M-S}\binom{N}{M}^{-1},$$

$$\mathbb{E}_M(X'_s) = \binom{n}{s}\binom{N-S}{M}\binom{N}{M}^{-1},$$

and

$$\mathbb{E}_p(X_s) = \binom{n}{s}p^S, \ \mathbb{E}_p(X'_s) = \binom{n}{s}q^S$$

where $S = \binom{s}{2}$ and $q = 1 - p$.

As a little diversion, let us remark that if instead of complete subgraphs we take subgraphs isomorphic to a fixed graph F then the arguments hardly change. Thus, writing $X_F = X_F(G_p)$ for the number of subgraphs of G_p isomorphic to F,

$$\mathbb{E}_p(X_F) = N_F p^{e(F)}, \tag{4}$$

where N_F is the number of subgraphs of K_n isomorphic to F.

At the danger of belabouring the point, note that (4) can be seen as above by taking an enumeration $F_1, F_2, \ldots, F_{N_F}$ of the subgraphs of K_n isomorphic to F and writing Y_i for the indicator function of F_i, that is, setting $Y_i(G_p) = 1$ if $F_i \subset G_p$ and $Y_i(G_p) = 0$ if $F_i \not\subset G_p$. Then $X_F = \sum_{i=1}^{N_F} Y_i$, and the summands are again identically distributed 0–1 random variables. Hence, by the additivity of expectation,

$$\mathbb{E}_p(X_F) = \sum_{i=1}^{N_F} \mathbb{E}_p(Y_i) = \sum_{i=1}^{N_F} \mathbb{P}_p(Y_i = 1) = N_F p^{e(F)},$$

as claimed by (4).

The number N_F is closely related to the automorphism group of F, that is, to the group of permutations of the vertices of F preserving adjacency. Indeed, if F has k vertices and its automorphism group has order a, so that K_k has $k!/a$ subgraphs isomorphic to F, then $N_F = \binom{n}{k}\frac{k!}{a} = \frac{(n)_k}{a}$, where $(n)_k$ is the kth *falling factorial*: $(n)_k = n(n-1)\cdots(n-k+1)$. Hence, in this case,

$$\mathbb{E}_p(X_F) = \frac{(n)_k}{a}p^{e(F)}. \tag{5}$$

In particular, if F is a k-cycle C_k, then

$$\mathbb{E}_p(X_{C_k}) = \frac{(n)_k}{2k}p^k. \tag{6}$$

Similar formulae hold for the number of *induced* subgraphs: for example, writing Y_{C_k} for the number of induced k-cycles,

$$\mathbb{E}_p(Y_{C_k}) = \frac{(n)_k}{2k}p^k q^{\binom{k}{2}-k} = \frac{(n)_k}{2k}p^k q^{k(k-3)/2}.$$

Let us return to our main thread. The simple Theorem 1 was all Erdős needed to get exponential lower bounds for the Ramsey numbers $R(s,t)$.

Theorem 2 (i) *If* $3 \le s \le n$ *are such that*

$$\binom{n}{s} < 2^{\binom{s}{2}-1},$$

then $R(s,s) \ge n+1$. *Also,*

$$R(s,s) > \frac{1}{e\sqrt{2}}s2^{s/2}. \tag{7}$$

(ii) *Suppose that* $3 \le s \le t \le n$ *and* $0 < p < 1$ *are such that*

$$\binom{n}{s}p^{\binom{s}{2}} + \binom{n}{t}q^{\binom{t}{2}} < 1,$$

where $q = 1-p$. *Then* $R(s,t) \ge n+1$.

Proof. (i) Consider $\mathcal{G}(n,1/2)$. With the notation above,

$$\mathbb{E}_{1/2}(X_s + X_s') = 2\binom{n}{s}2^{-\binom{s}{2}} < 1,$$

so there is a graph $G \in \mathcal{G}(n,1/2)$ with $(X_s + X_s')(G) = X_s(G) + X_s'(G) = 0$. This means precisely that neither G nor its complement contains a complete graph of order s. Hence $R(s,s) \ge n+1$, proving the first assertion.

Inequality (7) is an immediate consequence of this and inequality (1). Indeed, with $n = \lfloor \frac{s2^{s/2}}{e\sqrt{2}} \rfloor$, by (1) we have

$$\binom{n}{s}2^{-\binom{s}{2}+1} < \frac{n^s}{s!}2^{-\binom{s}{2}+1} < \frac{(e\sqrt{2})^{-s}s^s 2^{s^2/2}}{\sqrt{2\pi s}(s/e)^s}2^{-\binom{s}{2}+1} = \frac{2}{\sqrt{2\pi s}} < 1,$$

so $R(s,s) \ge n+1$.

(ii) This assertion is just a slight variant of the first: by our assumption, we have $\mathbb{E}_p(X_s + X'_t) < 1$, so $X_s(G) = X'_t(G) = 0$ for some graph $G \in \mathcal{G}(n, p)$. This means that G does not contain a complete graph of order s, and its complement $\overline{G}$ does not contain a complete graph of order t. Since G has n vertices, $R(s, t) \geq n + 1$, as claimed. □

The argument above can be applied to the space $\mathcal{G}(n, M)$ as well, instead of $\mathcal{G}(n, p)$: if

$$\mathbb{E}_M(X_s) + \mathbb{E}_M(X'_t) < 1,$$

then $R(s, t) \geq n+1$. In fact, in this way we get a *slightly* better result. For example, assuming that N is even, it is easily seen (cf. Exercise 11) that for $M = N/2$ we have

$$\mathbb{E}_M(X_s) \leq \mathbb{E}_{1/2}(X_s).$$

However, the improvement is negligible, and the calculations are considerably prettier in $\mathcal{G}(n, p)$ than in $\mathcal{G}(n, M)$.

Having seen the striking simplicity of the proof of Theorem 2 we do not think it unreasonable to expect that with more work we could improve on the bound $cs2^{s/2}$, where c is a constant. In fact, it seems that this is not the case: although the constant $c = 1/(e\sqrt{2})$ can be improved by a factor 2 to $\sqrt{2}/e$ (by a simple application of the Lovász local lemma, not discussed in this book), it is not even known whether the exponent 1 of s can be improved. Thus the Erdős–Szekeres upper bound (Theorem VI.1) and the Erdős lower bound (Theorem 2) tell us that

$$2^{s/2} \leq R(s, s) \leq 2^{2s},$$

and at the moment 1/2 and 2 are the best constants in the inequality above. It is very likely that in fact, $R(s, s) = 2^{(c+o(1))s}$ for some constant c, probably for $c = 1$, but a proof of this seems to be far in the future.

Concerning the off-diagonal Ramsey numbers $R(s, t)$, it is of particular interest to determine the order of $R_s(t) = R(s, t)$ as s is kept fixed and $t \to \infty$. After decades of improvements, it is now known that

$$\frac{c_1 t^2}{\log t} \leq R(3, t) \leq \frac{c_2 t^2}{\log t}$$

for some positive constants c_1 and c_2. The upper bound was proved by Shearer in 1983, making use of a method of Ajtai, Komlós and Szemerédi, while the lower bound was proved by Kim in 1995 by an ingenious and intricate probabilistic argument. As one of the first striking applications of the probabilistic method, Erdős had shown over 30 years before that $c_1 t^2/\log^2 t$ is a lower bound for $R(3, t)$.

Only a slightly more complicated argument is needed to give lower bounds in the problem of Zarankiewicz (cf. Theorem 11 and inequalities (6) and (7) of Chapter IV). We shall use an analogue of the model $\mathcal{G}(n, M)$, rather than an analogue of $\mathcal{G}(n, p)$, partly for the sake of variety, and also because it makes the second part of the argument a little easier.

Theorem 3 *Let $2 \le s \le n_1$, $2 \le t \le n_2$, $\alpha = (s-1)/(st-1)$ and $\beta = (t-1)/(st-1)$. Then there is a bipartite graph $G_2(n_1, n_2)$ of size*

$$\left\lfloor \left(1 - \frac{1}{s!t!}\right) n_1^{1-\alpha} n_2^{1-\beta} \right\rfloor$$

that does not contain a $K(s,t)$ (with s vertices in the first class and t vertices in the second class).

Proof. Let

$$\begin{aligned} n &= n_1 + n_2, \\ V_1 &= \{1, 2, \dots, n_1\}, \\ V_2 &= \{n_1 + 1, n_1 + 2, \dots, n_1 + n_2\}, \\ E &= \{ij : i \in V_1,\ j \in V_2\}, \\ M &= \lfloor n_1^{1-\alpha} n_2^{1-\beta} \rfloor. \end{aligned}$$

We shall consider the probability space $\mathcal{G}(K_{n_1,n_2}, M)$ consisting of the $\binom{|E|}{M}$ graphs with vertex set $V = V_1 \cup V_2$ having exactly M edges from E and none outside E. (Note that this is *not* the probability space considered in the previous theorems.) The expected number of $K_{s,t}$ subgraphs contained in a graph $G \in \mathcal{G}(K_{n_1,n_2}, M)$ is

$$E_{s,t} = \binom{n_1}{s}\binom{n_2}{t}\binom{|E| - st}{M - st}\binom{|E|}{M}^{-1},$$

where the first factor is the number of ways the first class of $K_{s,t}$ can be chosen, the second factor is the number of ways the second class can be chosen and the third factor is the number of ways the $M - st$ edges outside a $K_{s,t}$ can be chosen. Now,

$$\binom{|E| - st}{M - st}\binom{|E|}{M}^{-1} = \prod_{i=0}^{st-1} \frac{M - i}{n_1 n_2 - i} < \left(\frac{M}{n_1 n_2}\right)^{st},$$

so

$$E_{s,t} < \frac{1}{s!t!} n_1^s n_2^t \left(\frac{M}{n_1 n_2}\right)^{st} \le \frac{1}{s!t!} n_1^s n_2^t (n_1^{-\alpha} n_2^{-\beta})^{st} = \frac{1}{s!t!} n_1^{1-\alpha} n_2^{1-\beta}.$$

Thus there is a graph $G_0 \in \mathcal{G}(K_{n_1,n_2}, M)$ that contains fewer than $n_1^{1-\alpha} n_2^{1-\beta}/s!t!$ complete bipartite graphs $K_{s,t}$. Omit one edge from each $K_{s,t}$ in G_0. The obtained graph $G = G_2(n_1, n_2)$ has at least

$$\lfloor n_1^{1-\alpha} n_2^{1-\beta} \rfloor - \left\lfloor \frac{1}{s!t!} n_1^{1-\alpha} n_2^{1-\beta} \right\rfloor \ge \left\lfloor \left(1 - \frac{1}{s!t!}\right) n_1^{1-\alpha} n_2^{1-\beta} \right\rfloor$$

edges and contains no $K_{s,t}$. □

By similar methods one can construct a graph of order n and size

$$\left\lfloor \frac{1}{2}\left(1-\frac{1}{s!t!}\right) n^{2-(s+t-2)/(st-1)} \right\rfloor$$

that does not contain a $K_{s,t}$ (see Exercise 12).

It is very likely that the lower bound in Theorem 3 is also far from the truth: for example, for $s = t$ it gives $z(n, n; t, t) \geq cn^{2-2/(t+1)}$ while it is expected that the upper bound in Theorem IV.11, namely $cn^{2-1/t}$, is the correct order of $z(n, n; t, t)$. Nevertheless, for a fixed large value of $s = t$, Theorem 3 is essentially the best lower bound at the moment.

As our third, and final, application of random graphs to central problems of graph theory, we present the theorem of Erdős about the existence of graphs of large girth and large chromatic number.

Theorem 4 *Given natural numbers $g \geq 3$ and $k \geq 2$, there is a graph of order k^{3g}, girth at least g and chromatic number at least k.*

Proof. We may assume that $g \geq 4$, and $k \geq 4$ since otherwise the assertion is trivial. Set $n = k^{3g}$, $p = 2k^{2-3g} = 2k^2/n$, and consider the space $\mathcal{G}(n, p)$. Writing $Z_\ell = Z_\ell(G_p)$ for the number of ℓ-cycles in our random graph G_p, we know from (6) that

$$\mathbb{E}_p(Z_\ell) = \frac{(n)_\ell}{2\ell} p^\ell < \frac{(np)^\ell}{2\ell} \frac{2^\ell k^{2\ell}}{2\ell}.$$

Hence the expected number of cycles of length at most $g - 1$ is

$$\sum_{\ell=3}^{g-1} \mathbb{E}_p(Z_\ell) < \sum_{\ell=3}^{g-1} 2^\ell \frac{k^{2\ell}}{2\ell} < \frac{2^{g-1}k^{2g-2}}{3}, \tag{8}$$

where the last inequality is rather crude. Denote by Ω_1 the set of graphs in $\mathcal{G}(n, p)$ that contain at most $f = 2^{g-1}k^{2g-2}$ cycles of length less than g. Since

$$\sum_{\ell=3}^{g-1} \mathbb{E}_p(Z)_\ell \geq \mathbb{P}_p(\overline{\Omega}_1) f =\geq (1 - \mathbb{P}_p(\Omega_1)) f,$$

where $\overline{\Omega}_1 = \mathcal{G}(n, p) \backslash \Omega_1$, from (8) we see that

$$\mathbb{P}_p(\Omega_1) > \frac{2}{3}.$$

Now put $s = n/k = k^{3g-1}$, and write Ω_2 for the set of graphs in $\mathcal{G}(n, p)$ that do not contain a set of s vertices spanning at most f edges. Note that the assertion of the theorem follows if we show that $\Omega_1 \cap \Omega_2 \neq \emptyset$. Indeed, suppose that $G_0 \in \Omega_1 \cap \Omega_2$. In G_0, delete an edge from each cycle of length less than g to obtain a graph G of girth at least g. As $G_0 \in \Omega_1$, at most f edges have been deleted. Also, $G_0 \in \Omega_2$, so every s-set of vertices spans at least $f + 1$ edges. Hence in G every s-set spans at least one edge; i.e., an independent set has at most $s-1$ vertices: $\alpha(G) \le s-1$. Since by inequality (2) of Chapter V we have $\alpha(G)\chi(G) \ge n$, this gives $\chi(G) > k$, and we are done.

Now $\Omega_1 \cap \Omega_2 \ne \emptyset$ follows if we show that $\mathbb{P}_p(\Omega_2) \ge 1/3$. In fact, we shall be more generous and we shall show that $\mathbb{P}_p(\Omega_2)$ is very close to 1. In particular, $\mathbb{P}_p(\Omega_2) > 2/3$, so that $\mathbb{P}_p(\Omega_1 \cap \Omega_2) \ge \mathbb{P}_p(\Omega_1) + \mathbb{P}_p(\Omega_2) - 1 > 1/3$.

For $G_p \in \mathcal{G}(n,p)$ and $\ell \ge 0$, write $I_\ell(G_p)$ for the number of s-sets of vertices spanning precisely ℓ edges. Thus $I_\ell(G_p)$ is the number of s-sets of vertices that are independent but for ℓ edges. Setting $I(G_p) = \sum_{\ell=0}^{f} I_\ell(G_p)$, we have

$$\Omega_2 = \{G_p \in \mathcal{G}(n,p) : I(G_p) = 0\}.$$

Hence

$$\begin{aligned}\mathbb{E}_p(I) &= \sum_{m=0}^{\infty} \mathbb{P}_p(I=m)m \ge \sum_{m=1}^{\infty} \mathbb{P}_p(I=m) = \mathbb{P}_p(I \ge 1)\\ &= 1 - \mathbb{P}_p(I=0) = 1 - \mathbb{P}_p(\Omega_2);\end{aligned}$$

so it suffices to prove that $\mathbb{E}_p(I) < 1/3$.

This is only a matter of straightforward estimates. Indeed, with $S = \binom{s}{2}$,

$$\mathbb{E}_p(I_\ell) = \binom{n}{s}\binom{S}{\ell} p^\ell (1-p)^{S-\ell},$$

since we have $\binom{n}{s}$ choices for the s-set and $\binom{S}{\ell}$ choices for the ℓ edges spanned by the s-set; choosing these ℓ edges and none other that joins vertices in our s-set gives the factor $p^\ell(1-p)^{S-\ell}$. Note that, by (3), $(1-p)^{S-\ell}$ is at most $e^{-p(S-\ell)}$. It is easily checked that $\mathbb{E}_p(I_\ell)/\mathbb{E}_p(I_{\ell+1}) < 1/2$ for $0 \le l < f$. Recalling that $\binom{a}{b} \le (ea/b)^b$, $s = n/k = k^{3g-1}$, $S = s^2/2 - s/2 \le s^2/2$, $p = 2k^2/n$, $ps = 2k$, $ps^2 = 2n = 2k^{3g}$, $f = 2^{g-1}k^{2g-2}$ and $pf = 2^g k^{-g}$, we find that

$$\begin{aligned}\mathbb{E}_p(I) &= \sum_{\ell=0}^{f} \mathbb{E}_p(I_\ell) \le 2\mathbb{E}_p(I_f)\\ &\le \left(\frac{en}{s}\right)^s \left(\frac{eS}{f}\right)^f p^f e^{-pS+pf}\\ &< (ek)^{n/k} \left(\frac{eps^2}{2f}\right)^f e^{-ps^2/2+ps/2+pf}\\ &< (ek)^{n/k} k^{(g+2)f} e^{-n}\\ &= (ek)^{n/k} \left(\frac{ek^{g+2}}{2^{g-1}}\right)^f e^{-n+k+2^g k^{-g}}.\end{aligned}$$

In the last step above, we made use of the fact that $f \ge 8k^6$, and so $(e/2^{g-1})^f e^{k+2^g/k^{-g}} < e^{k+1-f} < 1$. Hence

$$\log \mathbb{E}_p(I) < -n\left\{1 - \frac{1+\log k}{k} - \frac{(g+2)2^{g-1}\log k}{k^{g+2}}\right\} < -n/4,$$

so

$$\mathbb{E}_p(I) < e^{-n/4} < 1/3,$$

as required. □

Theorem 4 raises a natural question which, at present, is far from being answered. About how large is $n_1(g, k)$, the minimal order of a graph of girth at least g and chromatic number at least k? Since a graph of chromatic number at least k has a subgraph of minimal degree at least $k - 1$, we have $n_1(g, k) \geq n_0(g, k - 1)$, where $n_0(g, \delta)$ is the function in Theorem IV.1. Hence $n_1(g, k)$ is roughly between $k^{g/2}$ and k^{3g}. With some more work the upper bound can be reduced a little, but it seems to be difficult to determine $\lim_{g,k\to\infty} \frac{\log n_1(g,k)}{g \log k}$, if the limit exists, which is most likely.

VII.2 Simple Properties of Almost All Graphs

In the first section we saw how useful it is to know that most graphs in a model have a certain property. Now we shall go a step further, namely we shall discuss properties shared by almost all graphs. Given a property Q, we shall say that *almost every* (a.e.) *graph* in a probability space Ω_n consisting of graphs of order n has property Q if $\mathbb{P}(G \in \Omega_n : G \text{ has } Q) \to 1$ as $n \to \infty$. In this section we shall always take $\Omega_n = \mathcal{G}(n, p)$, where $0 < p < 1$ may depend on n.

Let us assume first that $0 < p < 1$ is *fixed*; that is, p is independent of n.

There are many simple properties holding for almost every graph in $\mathcal{G}(n, p)$. For instance, if H is an arbitrary fixed graph, then almost every $G_p \in \mathcal{G}(n, p)$ contains H as a spanned subgraph. Indeed, if $|H| = h$, then the probability that the subgraph of G spanned by a given set of h vertices is isomorphic to H is positive, say $r > 0$. Since $V(G)$ contains $\lfloor n/h \rfloor$ disjoint subsets of h vertices each, the probability that no spanned subgraph of G is isomorphic to H is at most $(1 - r)^{\lfloor n/h \rfloor}$, which tends to 0 as $n \to \infty$. The following result is a strengthened version of this observation.

Theorem 5 *Let $1 \leq h \leq k$ be fixed natural numbers and let $0 < p < 1$ be fixed also. Then in $\mathcal{G}(n, p)$ a.e. graph G_p is such that for every sequence of k vertices $x_1, x_2, \dots, x_k$ there exists a vertex x such that $xx_i \in E(G_p)$ if $1 \leq i \leq h$ and $xx_i \notin E(G_p)$ if $h < i \leq k$.*

Proof. Let $x_1, x_2, \dots, x_k$ be a sequence of vertices. The probability that a vertex $x \in W = V(G) - \{x_1, \dots, x_k\}$ has the required properties is $p^h q^{k-h}$. Since for $x, y \in W$, $x \neq y$, the edges xx_i are chosen independently of the edges yx_i, the probability that no suitable vertex x can be found for this particular sequence is $(1 - p^h q^{k-h})^{n-k}$. There are $(n)_k = n(n - 1) \cdots (n - k + 1)$ choices for the sequence $x_1, x_2, \dots, x_k$, so the probability that there is a sequence $x_1, x_2, \dots, x_k$

for which no suitable x can be found is at most

$$\varepsilon = n^k(1 - p^h q^{k-n})^{n-k}.$$

Clearly, $\varepsilon \to 0$ as $n \to \infty$. □

By a result of Gaifman concerning first-order sentences, Theorem 5 implies that *for a fixed* $0 < p < 1$ *every first-order sentence about graphs is either true for a.e. graph in* $G \in \mathcal{G}(n, p)$ *or is false for a.e. graph.* Though this result looks rather sophisticated, it is in fact weaker than the shallow Theorem 5, for given any first-order sentence, Theorem 5 enables us to deduce without any effort whether the sentence holds for a.e. graph or it is false for a.e. graph. In particular, each of the following statements concerning the model $\mathcal{G}(n, p)$ for a fixed $p \in (0, 1)$ is an immediate consequence of Theorem 5.

1. For a fixed integer k, a.e. graph $G_{n,p}$ has minimal degree at least k.
2. Almost every graph $G_{n,p}$ has diameter 2.
3. Given a graph H, a.e. graph $G_{n,p}$ is such that whenever $F_0 \subset G_{n,p}$ is isomorphic to a subgraph F of H, there exists an H_0 isomorphic to H satisfying $F_0 \subset H_0 \subset G_{n,p}$.

Rather naturally, most statements we are interested in are not first-order sentences, since they concern large subsets of vertices. "For a given $\varepsilon > 0$, a.e. graph $G_{n,p}$ has at least $\frac{1}{2}(p - \varepsilon)n^2$ edges and at most $\frac{1}{2}(p + \varepsilon)n^2$ edges". "Almost no graph $G_{n,p}$ can be coloured with $n^{1/2}$ colours". "Almost every graph $G_{n,p}$ contains a complete graph of order $\log n/\log(1/p)$". "Given $\varepsilon > 0$, a.e. $G_{n,p}$ is $\frac{1}{2}(p - \varepsilon)n$-connected". These statements are all true for a fixed p and are easily proved (see Exercises 16–20); however, none of them is a first-order sentence.

Now we shall examine the model $\Omega = \mathcal{G}(n, p)$ under the assumption that $0 < p < 1$ *depends on* n, *but* $pn^2 \to \infty$ *and* $(1 - p)n^2 \to \infty$ *as* $n \to \infty$. In this case for every fixed m, a.e. G_p is such that $e(G_p) \geq m$ and $e(\overline{G_p}) \geq m$. As before, we put $N = \binom{n}{2}$, and for $M = 0, 1, \ldots, N$, we denote by Ω_M the set of graphs in $\mathcal{G}(n, M)$. Clearly, $\Omega = \bigcup_{M=0}^{N} \Omega_M$, and the elements of Ω_M have equal probability both in $\mathcal{G}(n, M)$ and $\mathcal{G}(n, p)$.

We shall show that the models $\Omega = \mathcal{G}(n, p)$ and $\mathcal{G}(n, M)$ are very close to each other, provided that M is about pN, the expected number of edges of a graph in Ω.

Clearly,

$$\mathbb{P}_p(\Omega_M) = \mathbb{P}_p(e(G_p) = M) = \binom{N}{M} p^M q^{N-M}.$$

Hence

$$\frac{\mathbb{P}_p(\Omega_M)}{\mathbb{P}_p(\Omega_{M+1})} = \frac{M+1}{N-M}\frac{q}{p}. \tag{9}$$

This shows that $\mathbb{P}_p(\Omega_M)/\mathbb{P}_p(\Omega_{M+1})$ increases with M, and $\mathbb{P}_p(\Omega_M)$ is maximal for some M satisfying $pN - p \leq M \leq pN + q$. Furthermore, if $0 < \varepsilon < 1$ and

n is sufficiently large then since $pn^2 \to \infty$ as $n \to \infty$,

$$\frac{\mathbb{P}_p(\Omega_M)}{\mathbb{P}_p(\Omega_{M+1})} < 1 - \varepsilon,$$

provided that $M < (1-\varepsilon)pN$; also, since $(1-p)n^2 \to \infty$ as $n \to \infty$,

$$\frac{\mathbb{P}_p(\Omega_{M+1})}{\mathbb{P}_p(\Omega_M)} < (1+\varepsilon)^{-1}$$

when $M > (1+\varepsilon)pN$. Putting $N_\varepsilon = \lfloor (1+\varepsilon)pN \rfloor$ and $N_{-\varepsilon} = \lceil (1-\varepsilon)pN \rceil$, we see from these inequalities that a.e. graph G_p satisfies $N_{-\varepsilon} \le e(G_p) \le N_\varepsilon$; that is,

$$\mathbb{P}_p\left(\bigcup_{M=N_{-\varepsilon}}^{N_\varepsilon} \Omega_M\right) \to 1 \text{ as } n \to \infty. \tag{10}$$

Another consequence of (9) is that there is an $\eta > 0$ (in fact, any $0 < \eta < \frac{1}{2}$ would do) such that

$$\mathbb{P}_p\left(\bigcup_{M=0}^{\lfloor pN \rfloor} \Omega_M\right) > \eta \tag{11}$$

if n is sufficiently large. Now (10) and (11) imply that if $\Omega^* \subset \Omega$ is such that $\mathbb{P}_p(\Omega^*) \to 1$ and $n \to \infty$, then for any $\varepsilon > 0$ there are M_1 and M_2, such that $(1-\varepsilon)pN \le M_1 \le pN \le M_2 \le (1+\varepsilon)pN$ and

$$\frac{|\Omega_{M_1 i} \cap \Omega^*|}{|\Omega_{M_i}|} \to 1 \text{ as } n \to \infty \quad (i = 1, 2). \tag{12}$$

We call a set $\Omega^* \subset \Omega$ *convex* if $G \in \Omega^*$ whenever $G_1 \subset G \subset G_2$ and $G_1, G_2 \in \Omega^*$; a convex property of graphs is defined analogously. It is easily seen that for a convex set Ω^* relation (12) implies that

$$\frac{|\Omega_M \cap \Omega^*|}{|\Omega_M|} \to 1 \text{ as } n \to \infty \tag{12'}$$

whenever $M_1 \le M \le M_2$ and, in particular, if $M = \lfloor pN \rfloor$. Let us restate the assertions above as a theorem about the connection between the models $\mathcal{G}(n, p)$ and $\mathcal{G}(n, M)$.

Theorem 6 *Let $0 < p = p(n) < 1$ be such that $pn^2 \to \infty$ and $(1-p)n^2 \to \infty$ as $n \to \infty$, and let Q be a property of graphs.*

(i) *Suppose $\varepsilon > 0$ is fixed and, if $(1-\varepsilon)N < M < (1+\varepsilon)pN$, then a.e. graph in $\mathcal{G}(n, M)$ has Q. Then a.e. graph in $\mathcal{G}(n, p)$ has Q.*

(ii) *If Q is a convex property and a.e. graph in $\mathcal{G}(n, P(\text{edge}) = p)$ has Q, then a.e. graph in $\mathcal{G}(n, \lfloor pN \rfloor)$ has Q.*

All this is rather simple and could be proved in a much sharper form, but even in this weak version it does show that $\mathcal{G}(n, p)$ and $\mathcal{G}(n, M)$ are practically interchangeable in many situations, provided $p = M/N$, $M \to \infty$ and $(N - M) \to \infty$.

VII.3 Almost Determined Variables—The Use of the Variance

If $X = X(G)$ is a non-negative variable on $\Omega = \mathcal{G}(n, M)$ or $\Omega = \mathcal{G}(n, p)$, and the expectation of X is $\mathbb{E}(X) = \mu$, then, for $c > 1$,

$$\mathbb{P}(X \geq c\mu) \leq \frac{1}{c} \text{ and } \mathbb{P}(X \leq c\mu) \geq \frac{c-1}{c},$$

since

$$\mu = \mathbb{E}(X) \geq \mathbb{P}(X \geq c\mu)c\mu.$$

Thus if the expectation of X is very small, then X is small for most graphs. This simple fact, Markov's inequality, was used over and over again in the first section. However, if we want to show that X is *large* or *non-zero* for almost every graph in Ω then the expected value itself can very rarely help us, so we have to try a slightly less trivial attack. In the first instance we turn to the *variance* for help. Recall that if $\mu = \mathbb{E}(X)$, is the expectation of X then

$$\operatorname{Var}(X) = \sigma^2(X) = \mathbb{E}((X - \mu)^2) = \mathbb{E}(X^2) - \mu^2$$

is the *variance* of X and $\sigma = \sigma(X) \geq 0$ is the *standard deviation.* Chebyshev's inequality, which is just Markov's inequality applied to $(X - \mu)^2$, states that if $a > 0$, then

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

In particular,

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mu| \geq \mu) \leq \frac{\sigma^2}{\mu^2}. \tag{13}$$

In many examples, $X = X(G)$ is the number of subgraphs of G contained in a family $\mathcal{F} = \{F_1, F_2, \ldots\}$. Here $\mathcal{F}$ depends on n, and $V(F_i) \subset V(G) = [n]$. For example, $\mathcal{F}$ may be the set of $\binom{n}{s}$ complete subgraphs of order s as in Section 1, the set of Hamilton cycles, or the set of complete matchings. As in Section 1, X can be written as $\sum_i Y_i$, where $Y_i = Y_{F_i} = Y_{F_i}(G)$ is the indicator function of F_i: it is 1 if $F_i \subset G$ and 0 otherwise. Then, clearly,

$$\mathbb{E}(X^2) = \mathbb{E}((\sum_i Y_i)^2) = \sum_i \sum_j \mathbb{E}(Y_i Y_j) = \sum_{(F_i, F_j)} \mathbb{P}(G \text{ contains } F_i \cup F_j), \tag{14}$$

where the summation is over all ordered pairs (F_i, F_j) with $F_i, F_j \in \mathcal{F}$.

Let us use these ideas to determine the values of $p = p(n)$ for which $G_{n,p}$ is likely to contain a subgraph F. Following Erdős and Rényi, we call a graph *balanced* if no subgraph of it has strictly larger average degree. Thus if $F = G(k, \ell)$, that is, F has k vertices and ℓ edges, then it is balanced if every subgraph with k' vertices has at most $k'\ell/k$ edges. Note that complete graphs, cycles and trees are balanced; see Fig. VII.1 for an illustration of the concept.

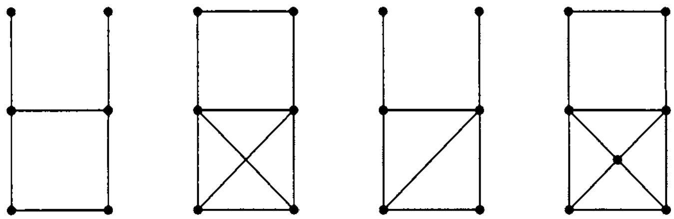


FIGURE VII.1. The first two graphs are balanced; the second two are not.

The following result of Erdős and Rényi shows that, as $p(n)$ increases, balanced subgraphs of $G_{n,p}$ appear rather suddenly.

Theorem 7 *Let $k \geq 2$, $k-1 \leq \ell \leq \binom{k}{2}$ and let $F = G(k, l)$ be a balanced graph (with k vertices and ℓ edges). If $p(n)n^{k/l} \to 0$ then almost no $G_{n,p}$ contains F, and if $p(n)n^{k/l} \to \infty$ then almost every $G_{n,p}$ contains F.*

Proof. Let $p = \gamma n^{-k/\ell}$, $0 < \gamma < n^{k/\ell}$, and denote by $X = X(G)$ the number of copies of F contained in $G_{n,p}$. Denote by k_F the number of graphs with a fixed set of k labelled vertices that are isomorphic to F. Clearly, $k_F \leq k!$. Then

$$\mu = \mathbb{E}_p(X) = \binom{n}{k} k_F p^\ell (1-p)^{\binom{k}{2}-\ell} \leq n^k(\gamma^\ell n^{-k}) = \gamma^\ell,$$

so $\mathbb{E}_p(X) \to 0$ as $\gamma \to 0$, showing the first assertion.

Now let us estimate the variance of X when γ is large. Note that there is a constant $c_1 > 0$ such that

$$\mu \geq c_1 \gamma^\ell \text{ for every } \gamma. \tag{15}$$

According to (14), we have to estimate the probability that G contains two fixed copies of F, say F' and F''. Put

$$A_s = \sum_s \mathbb{P}_p(G_{n,p} \supset F' \cup F''),$$

where $\sum_s$ means that the summation is over all pairs (F', F'') with s vertices in common. Clearly,

$$A_0 < \mu^2.$$

Furthermore, in a set of s vertices F' has $t \leq (\ell/k)s$ edges. Hence, counting first the choices for F' and then for F'' with $s \geq 1$ common vertices with F', we find that for some constants c_2 and c_3,

$$\begin{aligned}
\frac{A_s}{\mu} &\leq \sum_{t \leq \ell s/k} \binom{k}{s}\binom{n-k}{k-s} k!\, p^{\ell-t} q^{\binom{k}{2}-\binom{s}{2}-\ell+t} \\
&\leq \sum_{t \leq \ell s/k} c_2 n^{k-s} \left(\gamma n^{-k/l}\right)^{\ell-t} \\
&\leq c_2 n^{-s}\gamma^\ell + c_3\gamma^{\ell-1}.
\end{aligned}$$

Here in the last step we separated the term with $t = 0$ from the rest. Consequently, making use of (14), we find that

$$\frac{\mathbb{E}_p(X^2)}{\mu^2} = \frac{\sum_0^k A_s}{\mu^2} \leq 1 + c_4\gamma^{-1}$$

for some constant c_4. Therefore, by (13),

$$\mathbb{P}(X = 0) \leq \frac{\sigma^2}{\mu^2} \leq c_4\gamma^{-1},$$

so $\mathbb{P}(X = 0) \to 0$ as $\gamma \to \infty$. $\square$

One of the most striking examples of a graphic invariant being almost determined in a random graph is that of the *clique number*, the maximal order of a complete subgraph. It turns out that for a fixed p, $0 < p < 1$, the clique number of almost every graph in $\mathcal{G}(n, p)$ takes one of *two* possible values. In fact, as proved by Bollobás and Erdős in 1976, for most values of n (in a well-defined sense) the clique number of almost every graph is just a function of p and n. We shall confine ourselves to proving a simple result in this direction. As in Theorem 1, denote by $X_r = X_r(G_{n,p})$ the number of K_r subgraphs, so that

$$\mathbb{E}(X_r) = \binom{n}{r} p^{\binom{r}{2}}.$$

Let $d = d(n, p)$ be the greatest natural number for which

$$\mathbb{E}(X_d) = \binom{n}{d} p^{\binom{d}{2}} \geq \log n. \tag{16}$$

As $\mathbb{E}(X_1) = n$ and $\mathbb{E}(X_n) = p^{\binom{n}{2}} < 1$, there is such a d, with $1 \leq d \leq n - 1$. With the aid of Stirling's formula (1), it is easily checked that

$$\frac{n}{\log_b n} < \frac{n}{d} < p^{-d/2} < n, \tag{17}$$

where $b = 1/p$ and $\log_b n = \log n / \log b$. Also

$$d = 2\log_b n + O(\log\log n). \tag{18}$$

Theorem 8 *Let $0 < p < 1$ be fixed. Then the clique number of almost every $G_{n,p}$ is d or $d + 1$, where $d = d(n)$ is given by (16).*

Proof. The assertion is equivalent to the following:

$$\mathbb{P}(X_{d+2} > 0) \to 0,$$
$$\mathbb{P}(X_d > 0) \to 1.$$

Note that, by the definition of d, $\mathbb{E}_p(X_{d+1}) < \log n$ so, by (17),

$$\mathbb{E}(X_{d+2}) = \frac{n - d - 1}{d + 2} p^{d+1} \mathbb{E}(X_{d+1}) < p^{d/2} \log 2 < n^{-1/4} \to 0$$

implying the first assertion.

Let us turn to the main assertion that $\mathbb{P}(X_d > 0) \to 1$. Note first that $\mu_d = \mathbb{E}(X_d) \geq \log n \to \infty$ so, by (13), it suffices to prove that $\sigma_d/\mu_d \to 0$, where $\sigma_d = \sigma(X_d)$.

Let us use (14) to calculate the second moment of X_d, summing separately over pairs of K_d subgraphs with exactly ℓ vertices in common:

$$\mathbb{E}(X_d^2) = \sum_{\ell=0}^{d} \binom{n}{d}\binom{d}{\ell}\binom{n-d}{d-\ell} p^{2\binom{d}{2}-\binom{\ell}{2}} = \binom{n}{d} p^{2\binom{d}{2}} \sum_{l=0}^{d} \binom{d}{l}\binom{n-d}{d-l} p^{-\binom{l}{2}}.$$

Since

$$\mu_d^2 = \mathbb{E}(X_d)^2 = \binom{n}{d}^2 p^{2\binom{d}{2}} = \binom{n}{d} p^{2\binom{d}{2}} \sum_{\ell=0}^{d} \binom{d}{\ell}\binom{n-d}{d-\ell},$$

with $\sigma_d = \sigma(X_d)$ we have

$$\begin{aligned}
\frac{\sigma_d^2}{\mu_d^2} = \frac{\mathbb{E}(X_d^2) - \mu_d^2}{\mu_d^2} &\leq \sum_{l=0}^{d} \binom{d}{l}\binom{n-d}{d-\ell}(p^{-\binom{\ell}{2}} - 1)\binom{n}{d}^{-1} \\
&\leq \sum_{l=2}^{d} \binom{d}{l}\binom{n-d}{d-\ell} p^{-\binom{\ell}{2}}\binom{n}{d}^{-1} \\
&\leq 2\sum_{\ell=2}^{d} \frac{d!^2}{\ell!(d-\ell)!^2} n^{-\ell} p^{-\binom{\ell}{2}} \\
&= 2\sum_{\ell=2}^{d} \varepsilon_d,
\end{aligned}$$

say. The terms $\varepsilon_2, \varepsilon_3 \ldots, \varepsilon_\ell$ are first decreasing and then increasing. In fact, it suffices to check that

$$\varepsilon_\ell \leq \varepsilon_3 + \varepsilon_{d-1}$$

for $3 \leq \ell \leq d-1$. Hence

$$\frac{\sigma_d^2}{\mu_d^2} \leq 2(\varepsilon_2 + \varepsilon_d) + 2d(\varepsilon_3 + \varepsilon_{d-1}). \tag{19}$$

Now

$$\begin{aligned}
2\varepsilon_2 &\leq d^4 n^{-2} p^{-1} < n^{-1} \\
2d\varepsilon_3 &\leq 2d^7 n^{-3} p^{-3} < n^{-2}, \\
2\varepsilon_d &= 2d! n^{-d} p^{-\binom{d}{2}} \leq 2/\mu_d,
\end{aligned}$$

and

$$2d\varepsilon_{d-1} \leq d^2 n p^{\ell-1} \varepsilon_d < n^{-1/2}.$$

Putting these bounds into (19), we find that

$$\frac{\sigma_d^2}{\mu_d^2} \leq 2n^{-1/2} + 2/\mu_d = o(1),$$

as required. □

After these specific examples, let us say a few words about the broader picture. As before, a property of graphs is a class of graphs closed under isomorphism. In particular, a property Q_n of graphs of order n can be viewed as a subset of the set of graphs with vertex set $[n]$: all we have to require is that this set is invariant under permutations of $[n]$. A property Q of graphs is *monotone increasing* if Q is invariant under the addition of edges: if $G \in Q$, $G \subset H$ and $V(G) = V(H)$ then $H \in Q$. (Similarly, a property is monotone decreasing if it is invariant under the deletion of edges. Thus being connected or Hamiltonian is a monotone increasing property, the property of being at most 3-connected is monotone decreasing, but the property of containing an induced 6-cycle is neither increasing nor decreasing.

Given a property Q, we write $\mathbb{P}_p(Q) = \mathbb{P}_p(G_{n,p} \text{ has } Q) = \mathbb{P}_p(G_{n,p} \in Q)$ for the probability that $G_{n,p} \in \mathcal{G}(n, p)$ has property Q; the analogous notation is used in $\mathcal{G}(n, M)$. It sounds like a tautology, but it does need a proof that, for a monotone increasing property Q, the probability $\mathbb{P}_p(Q)$ is an increasing function of p, and $\mathbb{P}_M(Q)$ is an increasing function of M (see Exercises 21).

Theorems 7 and 8 illustrate the general principle discovered by Erdős and Rényi: many a monotone increasing property of graphs arises rather suddenly. To express this assertion precisely, it is convenient to introduce threshold functions. A function $p_\ell(n)$ is a *lower threshold function (ltf)* for a monotone increasing property Q if almost no $G_{n,p_\ell(n)}$ has Q, and $p_u(n)$ is an *upper threshold function (utf)* for Q if almost every $G_{n,p_u(n)}$ has Q. Threshold functions are defined similarly for the space $\mathcal{G}(n, M)$.

In terms of threshold functions, Theorem 7 says that if $\omega(n) \to \infty$ and F is a balanced graph of average degree $2\ell/k$ then $n^{-k/\ell}/\omega$ is an ltf and $\omega n^{-k/\ell}$ is a utf for the property of containing F as a subgraph. Although we did not prove it here, the converse of these assertions is also true: $p_\ell(n)$ is an ltf for containing F iff $p_\ell(n)n^{k/\ell} \to 0$ and $p_n(n)$ is a utf iff $p_u(n)n^{k/\ell} \to \infty$. In fact, in many cases the lower and upper threshold functions are much closer to each other than in this example. To illustrate this, we present a classical result of Erdős and Rényi.

Theorem 9 *Let $\omega(n) \to \infty$ and set $p_\ell = (\log n - \omega(n))/n$ and $p_u = (\log n + \omega(n))/n$. Then a.e. G_{p_ℓ} is disconnected and a.e. G_{p_u} is connected. Thus, in the model $\mathcal{G}(n, p)$, p_ℓ is an ltf and p_u is a utf for the property of being connected.*

Proof. In proving the theorem, we may and shall assume that $\omega(n)$ is not too large, say $\omega(n) \leq \log\log\log n$, and n is large enough to guarantee that $\omega(n) \geq 10$. For $k \in \mathbb{N}$, let $X_k = X_k(G)$ be the number of components of $G \in \mathcal{G}(n, p)$ having exactly k vertices.

(i) Let $p = p_\ell$ and write μ for the expected number of isolated vertices of G_p. Then

$$\mu = \mathbb{E}(X_1) = n(1-p)^{n-1} \sim ne^{-\log n + \omega(n)} = e^{\omega(n)} \to \infty. \tag{20}$$

Furthermore, the expected number of ordered pairs of isolated vertices is

$$\mathbb{E}\big(X_1(X_1-1)\big) = n(n-1)(1-p)^{2n-3},$$

since there are $n(n-1)$ ways of choosing an ordered pair of vertices, and two given vertices are isolated iff none of the $2n-3$ pairs of vertices incident with at least one of them is an edge of G_p.

Consequently,

$$\mathbb{E}(X_1^2) = n(n-1)(1-p)^{2n-3} + n(1-p)^{n-1}$$

and so the variance $\sigma^2 = \sigma^2(X_1)$ is

$$\begin{aligned}
\mathbb{E}\big((X_1-\mu)^2\big) &= \mathbb{E}(X_1^2) - \mu^2 \\
&= n(n-1)(1-p)^{2n-3} + n(1-p)^{n-1} - n^2(1-p)^{2n-2} \\
&\le n(1-p)^{n-1} + pn^2(1-p)^{2n-3} \\
&\le \mu + \big(\omega(n) + \log n\big)ne^{-2\log n + 2\omega(n)}(1-p)^{-3} \\
&\le \mu + \frac{2\log n}{n}e^{2\omega(n)} \le \mu + 1.
\end{aligned} \tag{21}$$

In the penultimate inequality we made use of the fact that p is small so $(1-p)^3 \ge 1/2$, with plenty to spare. Therefore, by (20) and (21) we have

$$\mathbb{P}(G_p \text{ is connected}) \le \mathbb{P}(X_1 = 0) \le \frac{1}{\mu^2}\mathbb{E}\big((X_1-\mu)^2\big) \le \frac{\mu+1}{\mu^2} = \mu^{-1} + \mu^{-2},$$

showing that almost every G_p is disconnected.

(ii) Set $p = p_u = (\log n + \omega(n))/n$. Clearly,

$$\begin{aligned}
\mathbb{P}(G_p \text{ is disconnected}) &= \mathbb{P}\left[\sum_{k=1}^{\lfloor n/2\rfloor} X_k \ge 1\right] \\
&\le \mathbb{E}\left[\sum_{k=1}^{\lfloor n/2\rfloor} X_k\right] = \sum_{k=1}^{\lfloor n/2\rfloor} \mathbb{E}(X_k) \\
&\le \sum_{k=1}^{\lfloor n/2\rfloor} \binom{n}{k}(1-p)^{k(n-k)}
\end{aligned} \tag{22}$$

since we have $\binom{n}{k}$ choices for the vertex set of a component with k vertices and we have to guarantee that there are no edges joining this set to the rest of the graph (in addition to having some edges to guarantee that the component is connected, but we do not make use of this condition). Let us split the sum above into two parts.

First we take the case when k is small:

$$\begin{aligned}\sum_{1\le k\le n^{3/4}} \binom{n}{k}(1-p)^{k(n-k)} &\le \sum_{1\le k\le n^{3/4}} \left(\frac{en}{k}\right)^k e^{-knp} e^{k^2 p}\\ &\le \sum_{1\le k\le n^{3/4}} e^{(1-\omega(n))k} k^{-k} e^{2k^2(\log n)/n}\\ &\le 3e^{-\omega(n)}, \end{aligned} \tag{23}$$

if n is sufficiently large. When k is large, we argue slightly differently:

$$\begin{aligned}\sum_{n^{3/4}\le k\le n/2} \binom{n}{k}(1-p)^{k(n-k)} &\le \sum_{n^{3/4}\le k\le n/2} \left(\frac{en}{k}\right)^k e^{-knp/2}\\ &\le \sum_{n^{3/4}\le k\le n/2} (en^{1/4})^k n^{-k/2}\\ &\le \sum_{n^{3/4}\le k\le n/2} (e/n^{1/4})^k \le n^{-n^{3/4}/5}. \end{aligned} \tag{24}$$

Putting together (22), (23) and (24), we find that

$$\mathbb{P}(G_p \text{ is disconnected}) \le 4e^{-\omega(n)}$$

if n is sufficiently large. This shows that a.e. G_p is connected. □

What about the chromatic number of a random graph $G_{n,p}$ for a fixed value of p? Theorem 8 immediately gives us a lower, bound since $\chi(G) \ge |G|/\alpha(G)$ for every graph G. Also, the complement of a random graph $G_{n,p}$ is a random graph $G_{n,q}$ with $q = 1 - p$, so the distribution of the independence number $\alpha(G_{n,p})$ is precisely the distribution of the clique number $\omega(G_{n,q})$. Since, by Theorem 8, $\omega(G_{n,q}) = (\frac{1}{2} + o(1)) \log n/\log(1/q)$, we see that

$$\chi(G_{n,p}) \ge \left(\frac{1}{2} + o(1)\right) \frac{\log n}{\log(1/q)} \tag{25}$$

for almost every $G_{n,p}$.

How far is this trivial lower bound from the truth? A natural way of getting an upper bound for $\chi(G_{n,p})$ is to analyse the result of a colouring algorithm run on $G_{n,p}$. Now, the easiest colouring algorithm is the greedy colouring algorithm discussed in Section V.1. As shown by Bollobás and Erdős in 1976, this algorithm does produce a colouring that, with high probability, uses only about twice as many colours as the lower bound in (25). In 1988, Bollobás used a different, non-algorithmic, approach to prove that, in fact, (25) is almost best possible.

Theorem 10 *Let* $0 < p < 1$ *be constant. Then*

$$\chi(G_{n,p}) = \left(\frac{1}{2} + o(1)\right) \frac{\log n}{\log(1/q)}$$

for a.e. $G_{n,p}$, *where* $q = 1 - p$.

What Theorem 10 claims is that if $\varepsilon > 0$ then

$$\lim_{n\to\infty} \mathbb{P}_p\left(|\chi(G_{n,p})\log(1/q)/\log n - \frac{1}{2}| < \varepsilon\right) = 1.$$

The optimal threshold functions tell us a considerable amount about a property, but in order to obtain an even better insight into the emergence of a property, we should look at *hitting times.* Given a monotone increasing property Q, the time τ at which Q appears in a graph process $\tilde{G} = (G_t)_0^N$ is the *hitting time* of Q:

$$\tau = \tau_Q = \tau(\tilde{G}; Q) = \min\{t : G_t \text{ has } Q\}.$$

The threshold functions in the model $\mathcal{G}(n, M)$ are easily characterized in terms of hitting times. Indeed, m_ℓ is a lower threshold function for a property Q if, and only if,

$$\tau(\tilde{G}; Q) > m_\ell$$

for almost every $\tilde{G}$, and an upper threshold function is characterized analogously.

There are several striking results concerning hitting times stating that two properties that seem to be far from each other are almost the same in our space of random graphs. A beautiful example of this is the property of being connected, considered in Theorem 9.

What is a simple obstruction to being connected? The existence of an isolated vertex. Putting it another way, if Q_1 is the property of being connected (Q_1 = "conn") and Q_2 is the property of having minimal degree at least 1 (Q_2 = "$\delta \geq 1$"), then $\tau(\tilde{G}; Q_1) \geq \tau(\tilde{G}; Q_2)$ for *every* graph process $\tilde{G}$. Rather surprisingly, equality holds for almost every graph process.

Theorem 11 *For almost every graph process $\tilde{G}$ we have $\tau(\tilde{G}; \mathrm{conn}) = \tau(\tilde{G}; \delta \geq 1)$.*

Although the proof is only a little more complicated than that of Theorem 9, we shall not give it here. However, let us expand on the assertion. What Theorem 10 tells us is that if we start with an empty graph on a large set of vertices and keep adding to it edges at random until the graph has no isolated vertices then, with high probability, the graph we obtain is connected: the very edge that gets rid of the last isolated vertex makes the graph connected. At first sight this is a most unexpected result indeed.

Note that it is easy to deduce Theorem 9 from Theorem 11, since Theorem 11 implies that the property of being connected has the same threshold functions as the property of having minimal degree at least 1. It is easily proved that $\mathbb{P}_p(\delta(G_{n,p}) \geq 1) \to 0$ if, and only if, $\mathbb{E}_p(X_{n,p}) \to \infty$, where $X_{n,p}$ is the number of isolated vertices of $G_{n,p}$. Also, $\mathbb{P}_p(\delta(G_{n,p}) \geq 1) \to 1$ if, and only if, $\mathbb{E}_p(X_{n,p}) \to 0$. Similar assertions hold for the properties of being k-connected and having minimal degree at least k.

VII.4 Hamilton Cycles—The Use of Graph Theoretic Tools

In the proofs so far we always adopted a more or less head-on attack. We hardly needed more from graph theory than the definitions of the concepts involved, the emphasis was on the use of elementary probability theory. This section is devoted to a beautiful theorem of Pósa, concerning Hamilton cycles, the proof of which is based on an elegant result in graph theory. Of course, the ideal use of probabilistic methods in graph theory would have a mixture of all the ideas presented in the four sections. Thus we would prepare the ground by using non-trivial graph theoretic results and would apply probability theory to get information about graphs in a probability space tailor-made for the problem. We could then select an appropriate graph which we would afterwards alter with the aid of powerful graph-theoretic tools.

As we saw in Chapter IV, the study of Hamilton cycles has been an important part of graph theory for many years, and by now we know a good many sufficient conditions for a graph to be Hamiltonian. Here we are interested in a rather different aspect of the Hamilton cycle problem: what happens in the average case? Given n, for what values of m does a *typical* graph of order n and size m have a Hamilton cycle?

A Hamiltonian graph of order n has at least n edges, and a non-Hamiltonian graph of order n has at most $\binom{n-1}{2}+1$ edges. This leaves a rather large 'uncertainty' window: for

$$n \le m \le \binom{n-1}{2} + 1$$

some graphs $G(n, m)$ are Hamiltonian, and some others are non-Hamiltonian. Changing the restriction from the size to the minimal degree, we are only slightly better off: a Hamiltonian graph has minimal degree at least 2 and, by Dirac's theorem (Theorem III.2), a non-Hamiltonian graph has minimal degree at most $\lfloor (n-1)/2 \rfloor$.

What is fascinating is that if we do not demand *certainty*, only *high probability*, then the window above becomes very small indeed.

Instead of fixing the size, we shall fix the probability: as we know, there is very little difference between the two approaches, and it is easier to work with $\mathcal{G}(n, p)$ then $\mathcal{G}(n, M)$. Thus the problem we wish to tackle is the following. For what values of $p = p(n)$ is $G_{n,p}$ likely to be Hamiltonian? From the results in the previous section we do know a lower bound, albeit a rather weak one: if $p = (\log n - \omega(n))/n$, where $\omega(n) \to \infty$, then almost every $G_{n,p}$ is disconnected and so, a fortiori, almost no $G_{n,p}$ is Hamiltonian. In 1976 Pósa achieved a breakthrough when he proved that the same order of the probability guarantees that almost every $G_{n,p}$ is Hamiltonian.

The basis of the proof of this result is Theorem IV.15. Let S be a longest x_0-path in a graph H and write L for the set of endvertices of the transforms of S. Denote by N the set of neighbours of vertices of L *on* S and put $R = V(H) - L \cup N$. Then Theorem IV.15 states that H has no L–R edge. All we shall need from this is that if $|L| = \ell \le |H|/3$ then there are disjoint sets of size ℓ and $|H| - 3\ell + 1$ that are joined by no edge of H.

We start with a simple lemma in the vein of Theorem 4. Denote by D_t the number of pairs (X, Y) of disjoint subsets of V such that $|X| = t$, $|Y| = n - 3t$, and G has no X–Y edge.

Lemma 12 *Let $c > 3$ and $0 < \gamma < \frac{1}{3}$ be constants and let $p = (c \log n)/n$. Then in $\mathcal{G}(n, p)$ we have*

$$\mathbb{P}_p(D_t > 0 \text{ for some } t,\ 1 \le t \le \gamma n) = O(n^{3-c}).$$

Proof. Put $\beta = \frac{(c-3)}{4c}$. Clearly,

$$\begin{aligned}\sum_{t=1}^{\lfloor \gamma n \rfloor} \mathbb{E}_p(D_t) &= \sum_{t=1}^{\lfloor \gamma n \rfloor} \binom{n}{t}\binom{n-t}{n-3t}(1-p)^{t(n-3t)} \\ &\le n\binom{n-1}{2}(1-p)^{n-3} + \sum_{t=2}^{\lfloor \beta n \rfloor} \frac{1}{t!} n^{3t}(1-p)^{t(n-3t)} \\ &\quad + \sum_{t=\lfloor \beta n \rfloor + 1}^{\lfloor \gamma n \rfloor} 2^{2n}(1-p)^{t(n-3t)}.\end{aligned}$$

Now, since $(1-p)^n < n^{-c}$, we have

$$n^3(1-p)^{n-3} < (1-p)^{-3} n^{3-c};$$

if $2 \le t \le \beta n$, then

$$n^{3t}(1-p)^{t(n-3t)} < n^{t(3-(c(n-3t)/n))} \le n^{3-c};$$

and if $\beta n \le t \le \gamma n$, then

$$2^{2n}(1-p)^{t(n-3t)} < n^{2n/\log n - (n-3t)t/n} = O(n^{-\beta(1-3\gamma)n}).$$

Consequently,

$$\sum_{t=1}^{\lfloor \gamma n \rfloor} \mathbb{E}_p(D_t) = O(n^{3-c}),$$

implying the assertion of the lemma. □

Theorem 13 *Let $p = (c \log n)/n$ and consider the space $\mathcal{G}(n, p)$. If $c > 3$ and x and y are arbitrary vertices, then almost every graph contains a Hamilton path from x to y. If $c > 9$ then almost every graph is Hamiltonian connected: every pair of distinct vertices is joined by a Hamilton path.*

Proof. Choose $\gamma < \frac{1}{3}$ in such a way that $c\gamma > 3$ if $c > 9$ and $c\gamma > 1$ if $c > 3$.

Let us introduce the following notation for certain events in $\mathcal{G}(n, p)$.

$D = \{D_t = 0 \text{ for every } t,\ 1 \le t \le \lfloor \gamma n \rfloor\}$, where D_t is as before

$E(W, x) = \{G_{n,p}[W]$ has a path of maximal length whose end vertex is joined to $x\}$

$$\begin{aligned}
E(W, x|w) &= \{G_{n,p}[W] \text{ has a } w\text{-path of maximal length among the } w\text{-paths,} \\
&\qquad \text{whose endvertex is joined to } x\} \\
F(x) &= \{\text{every path of maximal length contains } x\} \\
H(W) &= \{G_{n,p}[W] \text{ has a Hamilton path}\} \\
H(x, y) &= \{G_{n,p} \text{ has a Hamilton } x\text{–}y \text{ path}\} \\
HC &= \{G_{n,p} \text{ is Hamiltonian connected}\}.
\end{aligned}$$

We identify an event with the corresponding subset of $\mathcal{G}(n, p)$, so that the complement of an event A is $\overline{A} = \mathcal{G}(n, p)\backslash A$.

Note that by Lemma 12 we have

$$\mathbb{P}_p(\overline{D}) = 1 - \mathbb{P}_p(D) = O(n^{3-c}).$$

Let us fix a vertex x and a set $W \subset V\backslash\{x\}$ with $|W| = n - 2$ or $n - 1$. Our first aim is to show that $\mathbb{P}_p(D \cap \overline{E}(W, x))$ is rather small. Let $G \in D \cap \overline{E}(W, x)$ and consider a path $S = x_0x_1 \cdots x_k$ of maximal length in $G[W]$. (By introducing an ordering in W, we can easily achieve that S is determined by $G[W]$.) Let $L = L(G[W])$ be the set of endvertices of the transforms of the x_0-path S and let R be as in Theorem IV.17 (applied to $G[W]$). Recall that $|R| \geq |W| + 1 - 3|L|$ and there is no L–R edge, so no L–$(R \cup \{x\})$ edge either. Since $G \in D$ and $|R \cup \{x\}| \geq n - 3|L|$, we find that $|L| \geq \gamma n$. Now, L depends only on the edges of $G[W]$, so it is independent of the edges incident with x. Hence,

$$\mathbb{P}_p(D \cap \overline{E}(W, x)) \leq \mathbb{P}_p(|L(G[W])| \geq \gamma n)\mathbb{P}_p(\Gamma(x) \cap L' = \emptyset),$$

where L' is a fixed set of $\lceil \gamma n \rceil$ vertices of W. Therefore,

$$\mathbb{P}_p(D \cap \overline{E}(W, x)) \leq (1 - p)^{\gamma n} < n^{-c\gamma},$$

so the probability in question is indeed small. Exactly the same proof implies that

$$\mathbb{P}_p(D \cap \overline{E}(W, x|w)) < n^{-c\gamma},$$

provided $|W| = n - 2$ or $n - 1$, $w \in W$ and $x \notin W$.

Note now that $\overline{F}(x) \subset \overline{E}(V - \{x\},\ x)$, so

$$\begin{aligned}
\mathbb{P}_p(\overline{H}(V)) &= \mathbb{P}_p\left(\bigcup_{x\in V} \overline{F}(x)\right) \leq \mathbb{P}_p\left(D \cap \bigcup_{x\in V} \overline{F}(x)\right) + \mathbb{P}_p(\overline{D}) \\
&\leq \sum_{x\in V} \mathbb{P}_p(D \cap \overline{F}(x)) + \mathbb{P}_p(\overline{D}) \\
&\leq n\mathbb{P}_p(D \cap \overline{E}(V - \{x\},\ x)) + \mathbb{P}_p(\overline{D}) \\
&\leq n^{1-c\gamma} + O(n^{3-c}).
\end{aligned}$$

This proves that if $c > 3$ then almost every graph has a Hamilton path.

Now let x and y be distinct vertices and put $W = V - \{x, y\}$. By the first part,

$$\mathbb{P}_p(\overline{H}(W)) \leq 2n^{1-c\gamma} + O(n^{3-c}).$$

Since

$$H(x, y) \supset H(W) \cap E(W, y) \cap E(W, x|y),$$

we have

$$\begin{aligned}\mathbb{P}_p(\overline{H}(x, y)) &\leq \mathbb{P}_p(\overline{H}(W)) + \mathbb{P}_p(D \cap \overline{E}(W, y)) \\ &\quad + \mathbb{P}_p(D \cap \overline{E}(W, x|y)) + \mathbb{P}_p(\overline{D}) \\ &\leq 2n^{1-c\gamma} + 2n^{-c\gamma} + O(n^{3-c}).\end{aligned}$$

Therefore, if $c > 3$ then almost every graph contains a Hamilton path from x to y.

Finally, as there are $\binom{n}{2}$ choices for an unordered pair $(x, y), x \neq y$,

$$\mathbb{P}_p(\overline{HC}) \leq \sum_{x \neq y} \mathbb{P}_p(\overline{H}(x, y)) \leq n^{3-c\gamma} + n^{2-c\gamma} + O(n^{5-c}).$$

Thus if $c > 9$ then almost every graph is Hamiltonian connected. □

Since every Hamiltonian connected graph is Hamiltonian, by Theorem 13 we have in particular that if $c > 9$ and $p = c \log n/n$ then almost every $G_{n,p}$ is Hamiltonian. Independently of Pósa, Korshunov proved the essentially best possible result that this assertion holds for every $c > 1$. More importantly, in 1983 Komlós and Szemerédi determined the best threshold functions for the property of being Hamiltonian.

Theorem 14 *Let $\omega(n) \to \infty$ and set $p_\ell = (\log n + \log\log n - \omega(n))/n$ and $p_u = (\log n + \log\log n + \omega(n))/n$. Then p_ℓ is a lower threshold function for the property of being Hamiltonian and p_u is an upper threshold function.*

In fact, analogously to Theorem 11, there is a hitting time result connecting the property of being Hamiltonian to its obvious obstruction. We remarked that being disconnected is an obvious obstruction to being Hamiltonian. In fact, there is an even more obvious obstruction, which is easy to detect: having minimal degree at most 1. As shown by Bollobás in 1983, in a graph process this is the main obstruction.

Theorem 15 *Almost every graph process $\tilde{G}$ is such that $\tau(\tilde{G}; \mathrm{Ham}) = \tau(\tilde{G}; \delta \geq 2)$, where "Ham" is the property of being Hamiltonian and "$\delta \geq 2$" is the property of having minimal degree at least* 2.

Thus if we stop a random graph process as soon as we get rid of the last vertex of degree at most 1 then, with high probability, we have a Hamiltonian graph. Theorem 15 easily implies Theorem 14; in fact, with only a little additional work it implies the following sharper form of Theorem 14.

Let $c \in \mathbb{R}$ be fixed and set $p = (\log n + \log\log n + c)/n$. Then

$$\lim_{n\to\infty} \mathbb{P}_p(G_{n,p} \text{ is Hamiltonian}) = e^{-e^{-c}}.$$

Needless to say, the story does not stop here: there are many further questions concerning random graphs and Hamilton cycles. For example, having discovered the 'primary' obstruction, namely the existence of vertices of degree at most 1, we find it natural to rule them out and ask for the probability that a random graph of order n and size m, *conditional* on having minimal degree at least 2, is Hamiltonian. As it happens, the secondary obstruction is a 'spider': three vertices of degree 2 having a common neighbour. However results of this type are better suited for a specialist treatise on random graphs, rather than for this book on graph theory.

VII.5 The Phase Transition

What does a 'typical' random graph G_M look like? Better still, what does a 'typical' graph process $(G_t)_0^N$ look like? In particular, how does the component structure of G_t change as t increases?

It is fairly easy to see that if t is rather small then G_t tends to have only tree-components, with the orders depending on the size of t.

Theorem 16 *Almost every random graph process is such that if $k \geq 2$ is fixed and $t = o(n^{(k-1)/k})$ then every component of G_t is a tree of order at most t. Furthermore, if s is constant and $t/n^{(k-2)/(k-1)} \to \infty$ then G_t has at least s components of order k.*

The proof of this assertion goes along the lines of the proof of Theorem 7 and is rather vapid: we do not even need that there are k^{k-2} trees of order k (see Exercise I.41 and Theorem VIII.20). All we have to do is to estimate $\mathbb{E}(X_k)$ and $\mathbb{E}(X_k^2)$, where X_k is the number of trees of order k in G_t, using that there are some $t(k)$, $1 \leq t(k) \leq \binom{\binom{k}{2}}{k-1}$, trees on k distinguished vertices (see Exercise 34).

The growth of the maximal order of a component described in Theorem 16 is fairly steady and regular, without any unexpected changes. What Erdős and Rényi discovered is that around $t = n/2$ this growth becomes frantic: taking a bird's-eye view of the graph we see a sudden qualitative change in the component structure. This qualitative change is the *phase transition* of a random graph process. Vaguely speaking, before time $n/2$ every component has $O(\log n)$ vertices, but after time $n/2$ there is a unique largest component of order n (i.e., containing a constant proportion of the vertices). Even more, all other components are still of order $\log n$; in fact, as t increases, they are getting smaller.

In order to formulate a result precisely, for a graph G let us write $L^{(1)}(G) \geq L^{(2)}(G) \geq \cdots$ for the orders of the components of G, so that $\sum_i L^{(i)}(G) = |G|$. We see from Theorem 16 that $L^{(1)}(G_t) = k$ for almost every $G_t = G_{n,t}$ if $t/n^{(k-1)/k} \to 0$ and $t/n^{(k-2)/(k-1)} \to \infty$. For $t = \lfloor cn/2 \rfloor$, with c constant, the following is a sharper version of the celebrated result of Erdős and Rényi about phase transition.

Theorem 17 *Let $c > 0$ and $h \geq 1$ be fixed and let $\omega(n) \to \infty$. Set $\alpha = c - 1 - \log c$ and $t = t(n) = \lfloor cn/2 \rfloor$.*

(i) If $c < 1$ then, for almost every random graph G_t,

$$|L^{(i)}(G_t) - \frac{1}{\alpha}\left\{\log n - \frac{5}{2}\log\log n\right\}| \leq \omega(n)$$

for every i, $1 \leq i \leq h$.

(ii) There are constants $0 < c_1 < c_2$ such that, for every i, $1 \leq i \leq h$,

$$c_1 n^{2/3} < L^{(i)}(G_{\lfloor n/2 \rfloor}) < c_2 n^{2/3}$$

for almost every $G_{\lfloor n/2 \rfloor}$.

(iii) If $c > 1$ then, for almost every random graph G_t,

$$|L^{(1)}(G_t) - \gamma n| \leq \omega(n) n^{1/2},$$

where $0 < \gamma = \gamma(c) < 1$ is the unique solution of

$$e^{-c\gamma} = 1 - \gamma.$$

Furthermore,

$$\left| L^{(i)}(G_t) - \frac{1}{\alpha}\left\{\log n - \frac{5}{2}\log\log n\right\} \right| \leq \omega(n)$$

for every i, $2 \leq i \leq h$.

To appreciate the striking nature of Theorem 17, observe what happens to $L^{(1)}(G_t)$, the maximal order of a component, in a typical random graph process, as t increases from $\lfloor cn/2 \rfloor$ to $\lfloor c'n/2 \rfloor$, where $c' = \frac{100}{99}c$, say. If $0 < c < c' < 1$ or $1 < c < c'$, then $L^{(1)}(G_t)$ increases by a constant factor. However, if $c < 1 < c'$ then $L^{(1)}(G_t)$ grows dramatically, from order $\log n$ to order n. Passing through the critical point $c = 1$, the component structure changes completely, and a so-called *phase transition* occurs: before the critical point there are many components of about maximal size, and this maximal size is $O(\log n)$, but after the critical point there is a unique maximal component, which is much larger than the second largest. Passing through the critical point, a *giant component* emerges, with about γn vertices for some positive constant $\gamma = \gamma(c)$, while the second largest component still has order $O(\log n)$.

To see that $\gamma = \gamma(c)$ is well-defined for $c > 1$, set $f_c(\gamma) = c\gamma$ and $g(\gamma) = -\log(1-\gamma) = \gamma + \gamma^2/2 + \gamma^3/3 + \cdots$. Since for $\gamma \geq 0$ the function $g(\gamma)$ is strictly convex, $g'(0) = 1$ and $g(\gamma) \to \infty$ as $g \to 1-$, there is indeed a unique $\gamma \in (0, 1)$ such that $f_c(\gamma) = g(\gamma)$, i.e. $e^{-c\gamma} = 1 - \gamma$. In fact, it is easily checked that for $c = 1 + \varepsilon > 1$ we have $\gamma(c) = \gamma(1+\varepsilon) = 2\varepsilon - \frac{8}{3}\varepsilon^2 + \frac{28}{9}\varepsilon^3 + O(\varepsilon^4)$ (see Fig. VII.2).

What happens near the critical point remained a mystery for over 20 years, with many natural questions unanswered. In a typical graph process, for what values of t is the largest component at least twice as large as the second? How large can the second largest component become? Can it grow to $n/\log n$, say? Once we see the giant component of a graph process, at what speed does it grow?

Questions like these were answered by Bollobás in 1984, greatly clarifying the phase transition, and much more detailed results were proved by Łuczak in 1990,

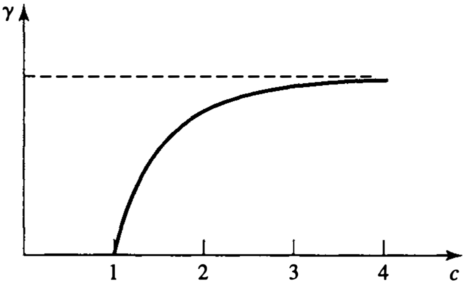


FIGURE VII.2. The curve $\gamma(c)$ for $c > 1$.

and by Janson, Knuth, Łuczak and Pittel in 1993. For example, if $t = \frac{n}{2} + s$, $s/n^{2/3} \to \infty$ and $s/n \to 0$ then

$$L^{(1)}(G_t) = 4s + o(s)$$

for almost every G_t; that is, on average, the addition of one edge adds 4 new vertices to the giant component. Also, $L^{(2)}(G_t)$ does not grow substantially above $n^{2/3}$: for example, in $\tilde{\mathcal{G}}^n$ we have

$$\lim_{n\to\infty} \mathbb{P}(L^{(2)}(G_t) \geq n^{2/3+\varepsilon} \text{ for some } t) = 0$$

for every $\varepsilon > 0$.

We shall not give a proper proof of Theorem 17, nor shall we do more than outline a very elegant approach due to Karp that can lead to a proof of Theorem 17. Karp's idea is to ignore the structure of a component and concentrate on the correspondence $x \to |C(x)|$, where $C(x)$ is the vertex set of the component of the vertex x in our random graph, and to exploit the similarity with a branching process. As usual, it is more convenient to work with the model $\mathcal{G}(n, p)$ rather than $\mathcal{G}(n, M)$: as in Section 2, it is then easy to pass from one model to another.

For a graph G and vertex $x \in V(G)$, let $C_G(x)$ be the vertex set of the component of x in G. To construct $C = C_G(x)$, proceed as follows. Set $x_1 = x$, $A_0 = \emptyset$ and $B_0 = \{x_1\}$. Then set $A_1 = \{x_1\}$, and add to B_0 all the neighbours of x_1 in G to get B_1. If $B_1 = A_1$, then $C = B_1 = A_1$; otherwise, pick a vertex $x_2 \in B_1 - A_1$, set $A_2 = A_1 \cup \{x_2\}$, and add to B_1 all the neighbours of x_2 in G to get B_2. If $B_2 = A_2$ then $C = B_2 = A_2$; otherwise, pick a vertex $x_3 \in B_2 - A_2$, set $A_3 = A_2 \cup \{x_3\}$ and add to B_2 all the neighbours of x_3 in G to get B_3. Proceeding in this way, we get a set $C = B_\ell = A_\ell$: this is precisely $C_G(x)$. Note that this set depends only on G, and not on the choices we made during the construction. At 'time' i, A_i is the set of vertices in C that we have tested for neighbours, so that at time ℓ we run out of new vertices to be tested, and thus $A_\ell = B_\ell$ is precisely the component $C_G(x)$.

Note that in the construction above, $|A_i| = i$ and $A_i \subset B_i$ for $i = 0, 1, \ldots, \ell$, and ℓ is the first index with $A_\ell = B_\ell$. Now if G is a random graph $G_{n,p}$ then, $A_i \subset B_i$, $A_i \neq B_i$, having constructed the probability that a vertex in $V - B_i$ is put into B_{i+1} is precisely p, independently of all other vertices and of all earlier choices. Hence we can run the process without any reference to the sets A_i (of vertices tested in the first i rounds): having constructed B_i, select the vertices of $V - B_i$ independently of each other, and add the selected vertices of B_i to obtain B_{i+1}. This sequence $B_0 \subset B_1 \subset \cdots$ can be run ad infinitum. However, the connection with $C(x)$ is easy to recover: write ℓ for the smallest index with $|B_\ell| = \ell$, the set B_ℓ is distributed precisely as $C(x)$. In particular, the probability that $C(x)$ has exactly k vertices is at most the probability that $|B_k|$ is precisely k.

The crunch comes now: each $|B_i|$ has a very simple distribution. Indeed, what is the probability that a vertex $y \in [n] - \{x\}$ is not put into B_i? We make i attempts at adding y to B_i, with each attempt succeeding with probability p, so $\mathbb{P}(y \notin B_i) = (1-p)^i$. As all choices are independent, $|B_i|$ has binomial distribution with parameters $n-1$ and $1-(1-p)^i$:

$$\mathbb{P}(|B_i| = k) = \binom{n-1}{k}(1-(1-p)^i)^k(1-p)^{i(n-k-1)}. \tag{26}$$

As noted above, this probability is an upper bound for $\mathbb{P}_p(|C(x)| = k)$.

This relation enables us to show that some values are extremely unlikely to occur as orders of components of $G_{n,p}$: with very high probability there is a gap in the orders of components.

Theorem 18 *Let $a \geq 2$ be fixed. If n is sufficiently large, $\varepsilon = \varepsilon(n) < 1/3$ and $p = p(n) = \frac{1+\varepsilon}{n}$ then, with probability at least $1 - n^{-a}$, $G_{n,p}$ has no component whose order k satisfies*

$$\frac{8a}{\varepsilon^2}\log n \leq k \leq \frac{\varepsilon^2}{12}n.$$

Proof. Set $k_0 = \lceil 8a\varepsilon^{-2}\log n \rceil$ and $k_1 = \lceil \varepsilon^2 n/12 \rceil$. Writing p_k for the probability that the component of $G_{n,p}$ containing a fixed vertex has k vertices, the probability that $G_{n,p}$ has a component of order k is at most np_k. Hence, it suffices to prove that

$$\sum_{k=k_0}^{k_1} p_k \leq n^{-a-1}.$$

We may assume that $k_0 \leq k_1$, so $\varepsilon^4 \geq 96a(\log n)/n \geq 1/n$, since otherwise there is nothing to prove. Now, by (26),

$$p_k \leq \mathbb{P}(|B_k| = k) \leq \frac{n^k}{k!}e^{-k^2/2n}(kp)^k(1-p)^{k(n-k-1)},$$

since

$$\frac{n-1}{k} = \frac{n^k}{k!}\prod_{j=1}^{k}(1-\frac{j}{n}) \leq \frac{n^k}{k!}e^{-k^2/2n},$$

and

$$(1-p)^k \geq 1 - kp.$$

Noting that

$$1+\varepsilon \leq e^{\varepsilon-\varepsilon^2/3}$$

for $|\varepsilon| \leq 1/3$, and recalling (1), Stirling's formula, we have

$$\begin{aligned} p_k &\leq \exp\{-k^2/2n - \varepsilon^2 k/3 + k^2(1+\varepsilon)/n\} \\ &\leq \exp\{-\varepsilon^2 k/3 + k^2/n\} \leq e^{-\varepsilon^2 k/4}. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{k=k_0}^{k_1} p_k &\leq \sum_{k=k_0}^{k_1} e^{-\varepsilon^2 k/4} = e^{-\varepsilon^2 k_0/4}(1-e^{-\varepsilon^2/4}) \\ &\leq \frac{5}{\varepsilon^2} e^{-\varepsilon^2 k_0/4} \leq nn^{-2a} \leq n^{-a-1}, \end{aligned}$$

as required. □

With more work, Theorem 18 could be proved in a much stronger version, giving us that a steadily growing, much larger gap arises soon after time $n/2$ and lasts till the end of the process. Nevertheless, even in this form it tells us a great deal about the components in a random graph process. For example, given $0 < \varepsilon_1 < \varepsilon_2 < 1/3$, there are positive constants α, β such that almost every graph process $\tilde{G} = (G_t)_0^N$ is such that for $(1+\varepsilon_1)n/2 \leq t \leq (1+\varepsilon_2)n/2$ the graph G_t has no component whose order is between $\alpha \log n$ and βn. Call a component *small* if it has at most $\alpha \log n$ vertices, and *large* if it has at least βn vertices. Also, set $t_1 = \lceil (1+\varepsilon_1)n/2 \rceil$ and $t_2 = \lfloor (1+\varepsilon_2)n/2 \rfloor$. Then a typical graph process $(G_t)_0^N$ is such that for $t_1 \leq t \leq t_2$ every component of G_t is either small or large.

Let us observe the changes in the component structure in a typical process $(G_t)_0^N$ as t grows from t_1 to t_2. What is the effect of the addition of an edge to G_t to produce G_{t+1}? If the new edge is added to a component, there is no change. If the new edge joins a large component to another, then the two components are replaced by a single component. Most importantly, what happens if the new edge joins two small components? The union of these small components is certainly not large, as $2\alpha \log n < \beta n$, with plenty to spare, so it has to be small. In particular, G_{t+1} has at most as many large components, as G_t; even more, if the new edge joins two large components then G_{t+1} has one fewer large component than G_t.

Conditioning on $\tilde{G} = (G_t)_0^N$ being a typical graph process and G_t containing at least two large components, the probability that the $(t+1)$st edge joins two large components is at least $\beta^2 n^2/\binom{n}{2} > 2\beta^2$. Now, G_{t_1} has at most $1/\beta$ large components, so if $\omega(n) \to \infty$ then with probability $1-o(1)$, after the addition of the next $\omega(n)$ edges all large components have been united. In particular, in a typical graph process G_{t_2} has a unique large component, the giant component of G_{t_2}, and all other components have at most $c_1 \log n$ vertices. This is precisely the qualitative version of the most interesting part of Theorem 17, part (iii).

The value of γ in Theorem 17(iii) can be determined by exploiting the similarity between a branching process and the growth of a component containing a vertex. Let us give a brief heuristic description of the similarity.

Let $p = c/n$, $c > 1$, and suppose we know that almost every $G_{n,p}$ has a unique giant component with $(\gamma + o(1))n$ vertices, where $\gamma > 0$, and all other components are small, with no more than $\alpha \log n$ vertices. How can we find γ? Clearly, γ is the limit of the probability that the component $C(x)$ of a fixed vertex x is small: $|C(x)| \le \alpha \log n$. To estimate this probability, let us 'grow' $C(x)$, starting from x, as in the proof of Theorem 18, but keeping track of the neighbours we put into $C(x)$. To be precise, let U_ℓ be the set of vertices at distance ℓ from x, and let $V_\ell = \bigcup_{i=1}^{\ell} U_i$ be the set of vertices at distance at most ℓ from x. We stop this process if $|V_\ell| > \alpha \log n$ for some ℓ, since then $C(x)$ is large, and also if $U_{\ell+1} = \emptyset$ and $|V_\ell| \le \alpha \log n$, since then $C(x) = V_\ell$ is small.

Let us take a close look at the way $U_{\ell+1}$ arises from (U_ℓ, V_ℓ). Letting $U_\ell = \{u_1, u_2, \dots\}$, first take all new neighbours of u_1, then all new neighbours of u_2, and so on, stopping the process if we ever reach $\alpha \log n$ vertices. Suppose then that we have reached $h \le \alpha \log n$ vertices from x when we test for the new neighbours of u_i. What is the distribution of the number of new neighbours of u_i? Clearly, for $k \le \alpha \log n$, we have

$$\begin{aligned}\mathbb{P}(u_i \text{ has } k \text{ new neighbours}) &= \binom{n-h}{k} p^k (1-p)^{n-h-k} \\ &= \frac{c^k}{k!} e^{-c}(1 + O((\log n)^2/n)).\end{aligned}$$

This means that the distribution of the number of descendents of u_i, that is, the number of vertices we add to $C(x)$ because of u_i, is close to the Poisson distribution with mean c. Therefore, the distribution of $|C(x)|$ is close to the distribution of the total population in a Poisson branching process, provided $C(x)$ is small.

To define this process, let Z_{ij}, $i = 0, 1, \dots$, $j = 1, 2, \dots$, be independent Poisson random variables, each with mean c:

$$\mathbb{P}(Z_{ij} = k) = \frac{c^k}{k!} e^{-c}.$$

Set $Z_0 = 1$. Having defined Z_n, set $Z_{n+1} = Z_{n1} + Z_{n2} + \cdots + Z_{nZ_n}$, where the empty sum (for $Z_n = 0$) is 0, as always. The interpretation is that Z_n is the size of the population in the nth generation and Z_{ni} is the number of descendents of the ith member of the nth generation.

Theorem 19 *Let $(Z_n)_0^\infty$ be as above, with $c > 1$, and write p_∞ for the probability that $Z_n > 0$ for every n. Then p_∞ is the unique root of*

$$e^{-cp_\infty} = 1 - p_\infty$$

in the interval $(0, 1)$.

Proof. Let p_n be the probability that $Z_n > 0$, so that $p_0 = 1$ and $p_\infty = \lim_{n\to\infty} p_n$. First we check, by induction on n, that $p_n \ge \gamma$ for every n, where γ is the unique root of $e^{-\gamma c} = 1 - \gamma$ in $(0, 1)$. This holds for $n = 0$ since $p_0 = 1 \geq \gamma$. Assume then that $n \geq 0$ and $p_n \geq \gamma$. Conditioning on $Z_1 = k \geq 1$, the process is the sum of k independent processes with the same distribution. Since $1 - p_t$ is the probability that the process dies out by time t,

$$\begin{aligned} 1 - p_{n+1} &= \mathbb{P}(Z_1 = 0) + \sum_{k=1}^{\infty} \mathbb{P}(Z_1 = k)(1 - p_n)^k \\ &= \sum_{k=0}^{\infty} \frac{c^k}{k!} e^{-c}(1 - p_n)^k \\ &= e^{-cp_n} \sum_{k=0}^{\infty} \frac{(c(1 - p_n))^k}{k!} e^{-c(1-p_n)} \\ &= e^{-cp_n} \leq e^{-c\gamma}. \end{aligned}$$

Hence $p_{n+1} \geq \gamma$, as claimed, and so $p_\infty = \lim_{n\to\infty} p_n \geq \gamma$.

By applying the argument above to $1 - p_\infty$ rather than $1 - p_{n+1}$ and $1 - p_n$, we see that

$$1 - p_\infty = \mathbb{P}(Z_1 = 0) + \sum_{k=1}^{\infty} \mathbb{P}(Z_1 = k)(1 - p_\infty)^k = e^{-cp_\infty}.$$

Hence p_∞ is a root of $e^{-cp_\infty} = 1 - p_\infty$ satisfying $0 < p_\infty \leq 1$, and we are done. □

Returning to the size $(\gamma + o(1))n$ of the giant component in Theorem 17(iii), we know that γ is the limit of the probability that our fixed vertex $x \in [n]$ belongs to the giant component. Therefore, γ is the probability that Z_n in Theorem 19 does not die out, so $\gamma = p_\infty$. Hence $e^{-c\gamma} = 1 - \gamma$, as claimed in Theorem 17(iii).

To make all this rigorous, we have to do more work, but it is clear that this approach can be used to establish the principal features of the phase transition. In fact, the method above is only one of several ways of investigating the phase transition. In particular, Erdős and Rényi, Bollobás, and Łuczak made use of the finer structure of the components, and Janson, Knuth, Łuczak and Pittel relied on generating functions and hard analysis to obtain very detailed results about the emergence of the giant component.

VII.6 Exercises

1.⁻ Show that the complement $\overline{G_p}$ of a random graph G_p is precisely G_q, where $q = 1 - p$.

2.⁻ Let $M \geq 0$ be fixed and for $n \geq 2M$ let $H_{n,M}$ consist of M independent edges and $n - 2M$ isolated vertices. Show that

$$\lim_{n\to\infty} \mathbb{P}_{n,M}(G_{n,M} \cong H_{n,M}) = 1.$$

3.$^-$ Prove that there is a tournament of order n (see Exercise 12 of Chapter I) that contains at least $n!2^{-n+1}$ directed Hamiltonian paths.

4.$^-$ Let $\vec{G} = (V, \vec{E})$ be a directed graph with m edges and without loops. Use expectation to show that V can be partitioned into sets V_1 and V_2 such that $\vec{G}$ contains more than $m/4$ edges from V_1 to V_2.

5.$^-$ Show that a graph of size m has a k-partite subgraph with at least $(k-1)m/k$ edges. [*Hint.* Consider a random k-colouring of the vertices. What is the expected number of edges joining vertices of distinct colours?]

6.$^-$ Show that a graph of order n and size m has a bipartite subgraph with at least $2\lfloor n^2/4\rfloor m/n(n-1)$ edges. [*Hint.* Consider random bipartitions into as equal classes as possible.]

7. Let $G = G(n, m)$ be a graph with chromatic number r. Show that G has a bipartite subgraph with at least

$$\frac{m}{2}\left(1 + \frac{1}{2\lfloor (r-1)/2\rfloor + 1}\right) \geq \left(1 + \binom{1}{r}\right)\binom{m}{2}$$

edges. [*Hint.* Let $V_1, \ldots, V_r$ be the colour classes. Consider partitions of the colour classes.]

8. Show that a.e. graph $G_{n,1/2}$ has maximal degree at least $n/2+\sqrt{n}$ and minimal degree at most $n/2 - \sqrt{n}$.

9. Show that a.e. graph $G_{n,1/2}$ has at least $n^{1/3}$ vertices of degree precisely $\lfloor n/2\rfloor$. [*Hint.* Compute the expectation and variance of the number of these vertices.]

10.$^-$ Let G and H be graphs of order n. Show that G has a subgraph with at least $e(G)e(H)/\binom{n}{2}$ edges that is isomorphic to a subgraph of H.

11. Let F be a fixed graph and let $X_F = X_F(G)$ be the number of subgraphs of G isomorphic to F. Suppose that $N = \binom{n}{2}$ is even and set $M = N/2$. Show that

$$\mathbb{E}_M(X_F) \leq \mathbb{E}_{1/2}(X_F).$$

12. Show that there is a graph of order n and size

$$\left\lfloor \frac{1}{2}(1 - 1/s!t!)\, n^{2-(s+t-2)/(st-1)} \right\rfloor$$

that does not contain a $K_{s,t}$.

13. Given $2 \leq s \leq n$, let d be the maximal integer for which there is a $G_3(n, n, n)$ without a $K_3(s, s, s)$, in which every vertex is joined by at least d edges to each of the other two classes. Prove a lower bound for d.

14.$^+$ Use Theorem 3 to prove that if $r > 2$, $0 < \varepsilon < \frac{1}{2}(r-1)^{-2}$ and

$$d_r^* > -\frac{2}{\log(2(r-1)^2\varepsilon)},$$

then for every sufficiently large n there is a graph $G(n, m)$ not containing a $K_r(t)$, where $m \geq \{(r-2)/(2(r-1)) + \varepsilon\}n^2$ and $t = \lfloor d_r^* \log n \rfloor$. (Note that this shows that Theorem IV.20 is essentially the best possible.)

15. Show that a fixed vertex is isolated in about $1/e^2$ of the graphs in $\mathcal{G}(n, n)$ and has degree 1 in about $2/e^2$ of the graphs in $\mathcal{G}(n, n)$.
In Exercises 16–20 the model $\mathcal{G}(n, P(\text{edge}) = p)$ is used and $0 < p < 1$ is assumed to be fixed.

16. Show that for $\varepsilon > 0$ a.e. graph has at least $\frac{1}{2}(p-\varepsilon)n^2$ edges and at most $\frac{1}{2}(p+\varepsilon)n^2$ edges.

17.$^+$ Prove that a.e. graph G satisfies.

$$\delta(G) = \lambda(G) = \kappa(G) = pn - (2pqn\log n)^{1/2} + o(n\log n)^{1/2},$$

where $q = 1 - p$.

18.$^+$ Estimate the maximal value of t for which a.e. graph contains a spanned $K_{t,t}$. Estimate the corresponding value for $K_r(t) = K_{t,\ldots,t}$.

19. Let $0 < c < 1$. Prove that a.e. graph has the property that for every set W of $k = \lfloor c \log_2 n \rfloor$ vertices there is a vertex x_Z for each subset Z of W such that x_Z is joined to each vertex in Z and to none in $W - Z$. Check that for $c = 1$ it is impossible to find even a set of 2^k vertices disjoint from W. [*Hint.* Refine the proof of Theorem 7.]

20. Let H be a fixed graph. Show that a.e. $G_{n,p}$ is such that whenever an induced subgraph F_0 of $G_{n,p}$ is isomorphic to an induced subgraph F of H, then $G_{n,p}$ has an induced subgraph $H_0 \cong H$ containing F_0.

21.$^-$ Let Q be a monotone increasing property of graphs of order n. Show that if $p_0 < p_1$ and $M_0 < M_1$ then $\mathbb{P}_{p_0}(Q) \leq \mathbb{P}_{p_1}(Q)$ and $\mathbb{P}_{M_0}(Q) \leq \mathbb{P}_{M_1}(Q)$.

22.$^+$ Let $x \in \mathbb{R}$ be fixed and $p = p(n) = (\log n)/n + x/n$. Show that $\mathbb{P}_p(G_{n,p}$ has no isolated vertices$) \to e^{-e^{-x}}$. [*Hint.* Write $X = X(G_{n,p})$ for the number of isolated vertices. Show that, for every fixed $k \geq 1$, the kth factorial moment $\mathbb{E}_p((X_k)) = \mathbb{E}_p(X(X-1)\cdots(X-k+1))$ tends to e^{-kx}. Apply the Inclusion -Exclusion Formula to prove the result.]

23.$^+$ Sharpen Theorem 9 to the following result. If $p = (\log n)/n + x/n$ then the probability that $\mathcal{G}_{n,p}$ is connected is $e^{-e^{-x}}$. [*Hint.* Show first that a.e. $\mathcal{G}_{n,p}$ consists of a component and isolated vertices. Apply the result from Exercise 22].

24. Let $p = \log n/n + \omega(n)/n$, where $\omega(n) \to \infty$ arbitrarily slowly. Prove that a.e. $\mathcal{G}_{n,p}$ contains a 1-factor. [*Hint.* Use Tutte's theorem, Theorem III,12, ignoring the parity of the components.]

25. Show that $1/n$ is a threshold function for F_1 in Fig. VII.3; that is, if $pn \to 0$ then almost no graph contains F_1, and if $pn \to \infty$ then a.e. graph does.

26. What is the threshold function for F_2 in Fig. VII.3?

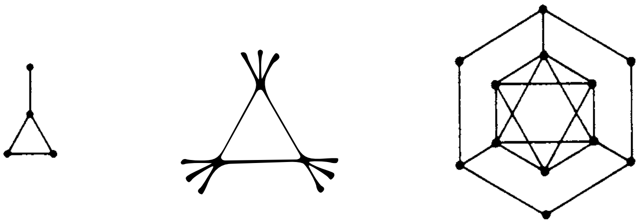


FIGURE VII.3. The graphs F_1, F_2, and F_3.

27.$^+$ Let $\varepsilon > 0$. Prove that if $p = n^{-(1/2)-\varepsilon}$ then almost no $G_{n,p}$ contains F_3 in Fig. VII.3 but if $p = n^{-(1/2)+\varepsilon}$ then a.e. $G_{n,p}$ does. [*Hint.* Find a suitable graph F_3^* that has average degree $2 + \varepsilon$.]

28.$^+$ Consider the random bipartite graph $G_{n,n;p}$ with two vertex classes of n vertices each, in which vertices in different classes are joined with probability p, independently of each other. Show that, for all fixed $k \geq 1$ and $0 < p < 1$, almost every $G_{n,n;p}$ is k-connected, has a k-factor, and has diameter 3.

29.$^+$ Consider random directed graphs in which all edges are chosen independently and with the same probability p. Prove that there is a constant c such that if $p = c((\log n)/n)^{1/2}$ then a.e. directed graph contains a directed Hamilton cycle. [*Hint.* What is the probability that a graph contains both edges $\overrightarrow{ab}$ and $\overrightarrow{ba}$? Apply Theorem 14 to the random graph formed by the double edges.]

30. Note that the suggested solution of Exercise 29 gives two directed Hamilton cycles with the same underlying (non-directed) edge set. Show that with $p = (1 - \varepsilon)((\log n)/n)^{1/2}$ almost no directed graph contains such a pair of Hamilton cycles.

31.$^+$ Show that there are at least $(2^N/n!) + o(2^N/n!)$ non-isomorphic graphs of order n. [*Hint.* Show that a.e. graph in $\mathcal{G}(n, P(\text{edge}) = \frac{1}{2})$ has trivial automorphism group; for the automorphism group see Section VIII.3.]

32.$^+$ Construct a *random interval graph* G_n with vertex set $[n] = \{1, 2, \ldots, n\}$ as follows: partition $[2n]$ into n pairs, $\{a_1, b_1\}, \ldots, \{a_n, b_n\}$, say, with $a_i < b_i$, and join i to j if $[a_i, b_i] \cap [a_j, b_j] \neq \emptyset$. What is the expected number of edges of G_n? Show that almost every G_n is connected.

33. Show that every tournament of order 2^k contains a transitive subtournament of order $k+1$ (i.e., there are vertices $x_1, \ldots, x_k + 1$ such that x_i dominates x_j whenever $i < j$). Show also that a tournament on $\lfloor 2^{k/2} \rfloor$ vertices need not contain a transitive subtournament on $k+1$ vertices.

34. Give a detailed proof of Theorem 16 calculating the expectation and variance of the number of tree-components of orders $k+1$ and k. Concerning the number $t(k)$ of trees on k distinguishable vertices, use only the fact that $t(k) \geq 1$ depends only on k.

35.+ The *conjugate* of $c > 1$ is $0 < c' < 1$, satisfying

$$c' e^{-c'} = ce^{-c}.$$

Show that every $c > 1$ has a unique conjugate. Show also that $\gamma(c) = 1 - c'/c$, where $\gamma(c)$ is the function in Theorem 17(iii), so that the giant component of a typical $G_{n,c/n}$ has about $\gamma(c)n$ vertices.

36. Given the space Ω_n of random permutations of $[n]$, with the permuations taken to be equiprobable. Write a permutation $\pi \in \Omega_n$ as a sequence $\pi(1), \pi(2), \cdots, \pi(n)$. If $1 \leq i_1 < \cdots < i_k \leq n$ and $\pi(i_1) < \ldots < \pi(i_k)$, then $\pi(i_1), \ldots, \pi(i_k)$ is said to be an *increasing subsequence of length k* in π. Show that almost no subsequence $\pi \in \Omega_n$ contains an increasing subsequence of length at least $e\sqrt{n}$. [*Hint.* Let $I_k(\pi)$ be the number of increasing subsequences of length k contained in π. Estimate the expectation of I_k.]

37.+ (Exercise 36 contd.) Find a constant $c < e$ such that almost no permutation $\pi \in \Omega_n$ contains an increasing subsequence of length at least $c\sqrt{n}$. [*Hint.* for $d < c < e$, set $k = \lceil c\sqrt{n} \rceil$, $\ell = \lceil d\sqrt{n} \rceil$, and note that $\mathbb{E}(I_\ell) \geq \binom{k}{\ell} \mathbb{P}(I_k \geq 1)$.]

38.+ (Exercise 36 contd.) Show that almost every permutation $\pi \in \Omega_n$ contains an increasing subsequence of length at least $\sqrt{n}/e$. [*Hint.* Write $I(\pi)$ for the maximal length of an increasing subsequence of π, and $D(\pi)$ for the maximal length of a decreasing subsequence. Recall the result of Erdős and Szekeres stated in Exercise II.54 that $I(\pi)D(\pi) \geq n$ for every $\pi \in \Omega_n$. The assertion is easily deduced from this inequality and the assertion in Exercise 36. In fact, $I(\pi) = (2 + o(1))\sqrt{n}$ for almost every $\pi \in \Omega_n$, but the proof of this is quite substantial.]

39. Let A be the area of a triangle formed by three points selected at random and independently from a convex set $D \subset \mathbb{R}^2$ of area 1. Show that for $0 \leq a \leq 1$ we have $\mathbb{P}(A \leq a) \geq a$.

40.+ (Exercise 39 contd.) Let A be the area of a triangle formed by three points selected at random and independently from a unit disc (of area π). Prove that for $a > 0$ we have $\mathbb{P}(A \leq a) \leq 4a$. Can you prove a sharper inequality? [*Hint.* Let x, y and z be the three points. Show first that

$$\mathbb{P}(A \leq a \mid d(x, y) = t) \leq 4a/t\pi.$$

Deduce from $\mathbb{P}(d(x, y) \le t) \le t^2$ that

$$\mathbb{P}(A \le a) \le \int_0^1 2t\frac{4a}{t\pi}\,dt + \frac{4a}{\pi}.$$

Note that all these inequalities are very crude, and are easily improved.]

41.$^+$ (Exercise 40 contd.) Prove that, for every $n \ge 1$, there are n points in the unit disc such that no three points form a triangle of area less than $1/6n^2$.
[*Hint.* Select $2n$ points at random, and delete a point from every triple whose triangle has area less than $1/6n^2$. Use the assertion of Exercise 40 to prove the result. In fact, Heilbronn conjectured in the 1930s that the assertion in this exercise is essentially best possible: no matter how we arrange n points in the unit disc, some three of the points form a triangle of area $O(1/n^2)$. In 1981, this conjecture was shown to be false by Komlós, Pintz and Szemerédi.]

42. By imitating the proof of Theorem 4, show that for $p = \frac{1}{2}n^{-1/2}$ almost every $G_{n,p}$ is such that no maximal triangle-free subgraph of it contains more than $2n^{1/2}\log n$ vertices.

43.$^+$ (Exercise 42 contd.) Let $p = \frac{1}{2}n^{-1/2}$. Show that almost every $G_{n,p}$ is such that if H is a maximal triangle-free subgraph of it then
(i) $n^{3/2}/6 < e(H) < n^{3/2}/3$,
(ii) H does not contain an induced bipartite subgraph with more than $30n^{1/2}(\log n)^2$ edges.

VII.7 Notes

Perhaps the first non-trivial combinatorial result proved by probabilistic methods is the assertion of Exercise 3, proved by T. Szele in Combinatorial investigations concerning directed complete graphs (in Hungarian), *Mat. Fiz. Lapok* **50** (1943) 223–256; for a German translation see Kombinatorische Untersuchungen über gerichtete vollständige Graphen, *Publ. Math. Debrecen* **13** (1966) 145–168. However, the theory of random graphs really started with a number of papers of P. Erdős, including Some remarks on the theory of graphs, Bull. Amer. Soc. **53** (1947) 292–294, Graph theory and probability, *Canad. J. Math.* **11** (1959) 34–38, and Graph theory and probability II, *Canad. J. Math.* **13** (1961) 346–352. These papers contain Theorems 2 and 4, and the bound on $R(3, t)$ mentioned after Theorem 2.

The result about first order sentences that we mentioned after Theorem 7 is due to R. Fagin, Probabilities on finite models, *J. Symb. Logic* **41** (1976) 50–58.

The sharpest results in the direction of Theorem 8 are in B. Bollobás and P. Erdős, Cliques in random graphs, *Math. Proc. Cambridge Phil. Soc.* **80** (1976) 419–427, and Theorem 10 is from B. Bollobás, The chromatic number of random graphs, *Combinatorica* **8** (1988) 49–55.

Pósa's theorem (Theorem 13) is in L. Pósa, *Discrete Math.* **14** (1976) 359–364, its sharper form is in A.D. Korshunov, Solution of a problem of Erdős and Rényi, on Hamilton cycles in nonoriented graphs, *Soviet Mat. Doklady* **17** (1976) 760–764. The exact solution of the Hamilton cycle problem for random graphs (Theorem 14) is from J. Komlós and E. Szemerédi, Limit distributions for the existence of Hamilton cycles in a random graph, *Discrete Math.* **43** (1983) 55–63, and its hitting time version (Theorem 15) is from B. Bollobás, Almost all regular graphs are Hamiltonian, *Europ. J. Comb.* **4** (1983) 97–106. For the result on 'spiders' mentioned at the end of §4, see B. Bollobás, T.I. Fenner and A. M. Frieze, Hamilton cycles in random graphs of minimal degree k, in *A Tribute to Paul Erdős*, (A. Baker, B. Bollobás and A. Hajnal, eds), Cambridge University Press, 1990, pp. 59–95.

The fundamental paper on the growth of random graphs is P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960) 17–61. This paper contains a detailed discussion of sparse random graphs, covering amongst other phenomena the distribution of their components, the occurrence of small subgraphs (Theorem 7), and the phase transition (Theorem 18). The real nature of the phase transition was revealed in B. Bollobás, The evolution of random graphs, *Trans. Amer. Math. Soc.* **286** (1984) 257–274. For more detailed results see T. Łuczak, Component behaviour near the critical point of the random graph process, *Random Structures and Algorithms* **1** (1990) 287–310, T. Łuczak, B. Pittel, and J.C. Wierman, The structure of a random graph at the point of the phase transition, *Trans. Amer. Math. Soc.* **341** (1994) 721–748, and S. Janson, D.E. Knuth, T. Łuczak and B. Pittel, The birth of the giant component, *Random Structures and Algorithms* **4** (1993) 233–358.

For Heilbronn's conjecture, mentioned in Exercise 41, see J. Komlós, J. Pintz and E. Szemerédi, On Heilbronn's triangle problem, *J. London Math. Soc.* (2) **24** (1981) 385–396.

This chapter was based on B. Bollobás, *Random Graphs*, Academic Press, London, 1985, xvi+447 pp.

VIII

Graphs, Groups and Matrices

This chapter provides a brief introduction to *algebraic graph theory*, which is a substantial subject in its own right. We shall deal with only two aspects of this subject: the interplay between graphs and groups, and the use of matrix methods.

Graphs arise naturally in the study of groups, in the form of Cayley and Schreier diagrams, and also as objects whose automorphisms help us to understand finite simple groups. On an elementary level, a graph is hardly more than a visual or computational aid, but it does help to make the presentation clearer and the problems more manageable. The methods are useful both in theory and in practice: they help us to prove general results about groups and particular results about individual groups. The first section, about Cayley and Schreier diagrams, illustrates both these aspects. It also contains an informal account of group presentations.

The second section is about the use of the adjacency matrix of a graph, and its close relative, the Laplacian. Elementary linear algebra methods enable one to establish close links between eigenvalue distributions and basic combinatorial properties of graphs.

Matrix methods are especially powerful when the graphs to be studied have particularly pleasant symmetry properties. The third section is about such classes of graphs. Among other results, we shall present the theorem of Hoffman and Singleton, stating that a natural class of highly symmetric graphs has only few members.

The last section is about enumeration. As we shall see, some classes of labelled graphs are easily enumerated, while other enumeration problems, such as counting isomorphism classes of graphs, lead us to the study of orbits of permutation groups. The highlight of the section is Pólya's classical theorem, proved in 1937, which is the fundamental theorem for enumerating such orbits.

VIII.1 Cayley and Schreier Diagrams

Let A be a group generated by $a, b, \ldots$ The *graph* of A, also called its *Cayley diagram*, with respect to these generators is a *directed multigraph* whose edges are *coloured* with the generators: there is an edge from x to y *coloured* with a generator g iff $xg = y$. To illustrate this concept, in Fig. VIII.1 we show the Cayley diagrams of three small groups.

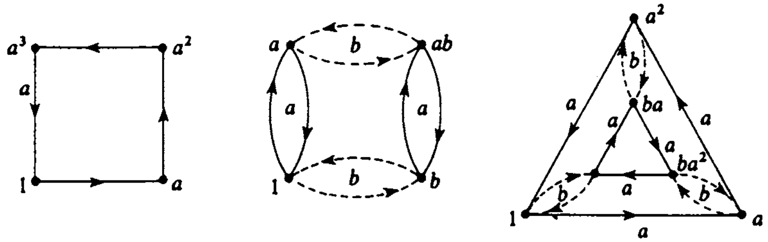


FIGURE VIII.1. The Cayley diagrams of (i) the cyclic group C_4 generated by a, (ii) the Klein four-group with generators a, b and (iii) the symmetric group S_3 with generators $a = (123)$ and $b = (12)$.

A Cayley diagram of a group is *regular*, and so is its colouring, in the following sense: for each vertex x and each generator (colour) g there is exactly one edge of colour g starting at x and exactly one edge of colour g ending at x. Furthermore, at most one edge goes from x to another vertex y. If we know the Cayley diagram of a group then we can easily answer questions posed in terms of the generators. What is the element aba^2b in S_3? It is the end of the directed walk starting at 1 whose first edge has colour a, the second has colour b, the third a, the fourth a and, finally, the fifth b. By following this walk in the third picture in Fig. VIII.1, we find that $aba^2b = a^2$. In general, two elements expressed as products of some generators are equal iff the corresponding walks starting at 1 end at the same vertex.

The Schreier diagram is a slight extension of the Cayley diagram. This time we have a group A, a set S of elements of A and a subgroup B of A. The *Schreier diagram of* A mod B describes the effect of the elements of S on the right cosets of B: it is a directed multigraph whose vertices are the right cosets of B, in which an edge of colour $s \in S$ goes *from* a coset H *to* a coset K iff $Hs = K$. (Thus a Cayley diagram is a Schreier diagram mod B, where B is the trivial subgroup $\{1\}$.) In most cases S is chosen to be a set of generators or a set which together with B generates A. Instead of giving a separate illustration, note that if A is the symmetric group on $\{1, 2, 3, 4\}$, B is the subgroup of elements fixing 1 and $S = \{a\}$ with $a = (1234)$ then the Schreier diagram is exactly the first picture in Fig. VIII.1, that is, the Cayley diagram of C_4. Once again we note that for each vertex H and each colour $g \in S$ exactly one edge coloured g starts at H and exactly one edge coloured g ends at H. However, some of the edges may be *loops*, that is, they may start and end at the same coset. Furthermore, there may be many edges of different colours joining two vertices.

Group diagrams do not tell us anything about groups that cannot be expressed algebraically. However, the disadvantage of the algebraic approach is that many lines of print are needed to express what is conveyed almost instantaneously by a single diagram. These diagrams are especially helpful when we have a concrete problem to be solved by hand, and moreover, the purely mechanical techniques involved are ideal for direct use on a computer. Since the advent of fast electronic computers, many otherwise hopeless problems have been solved in this way.

Group diagrams are particularly useful in attacking problems concerning groups given by means of their *presentations*. For the convenience of the reader we recall the basic facts about group presentations. We aim throughout for an intuitive description, rather than a rigorous treatment; the interested readers may fill in the details themselves or turn to some specialist books on the subject. A *word* W in the symbols a, b, c, ... is a finite sequence such as $ba^{-1}ccaa^{-1}b^{-1}a$; the empty sequence is denoted by 1. We call two words *equivalent* if one can be obtained from the other by repeatedly replacing xx^{-1} or $x^{-1}x$ by 1 (the empty word) or vice versa. Thus $abb^{-1}a^{-1}c^{-1}$ and $cc^{-1}c^{-1}dd^{-1}$ are both equivalent to c^{-1}. In fact, we shall use the same notation for a word and its equivalence class and so we write simply $abb^{-1}a^{-1}c^{-1} = cc^{-1}c^{-1}dd^{-1} = c^{-1}$. Furthermore, for simplicity $abbc^{-1}c^{-1}c^{-1} = ab^2c^{-3}$, etc. The (equivalence classes of) words form a group if multiplication is defined as juxtaposition: $(ab^{-1}c)(c^{-1}ba) = ab^{-1}cc^{-1}ba = a^2$. Clearly, a^{-1} is the inverse of a and $(a^{-1}b^{-1}c)^{-1} = c^{-1}ba$. This group is the *free group* generated by a, b, c ... and it is denoted by $\langle a, b, c \ldots \rangle$.

Let $R_\mu, R_\nu, \ldots$ be words in the symbols $a, b, c, \ldots$, let $F = \langle a, b, c, \ldots \rangle$ and let K be the normal subgroup of F generated by $R_\mu, R_\nu, \ldots$. Then the quotient group $A = F/K$ is said to be the group generated by a, b, c, ... and the *relators* $R_\mu, R_\nu, \ldots$; in notation $A = \langle a, b, c, \ldots \mid R_\mu, R_\nu, \ldots \rangle$.

Once again we use a word to denote its equivalence class and write equality to express equivalence. More often than not, a group presentation is written with *defining relations* instead of the more pedantic relators. Thus $\langle a, b \mid a^2 = b^3 \rangle$ denotes the group $\langle a, b \mid a^2b^{-3} \rangle$. A group is *finitely presented* if in its presentation there are finitely many generators and relations. It is easily seen that two words W_1 and W_2 are equivalent in A iff W_2 can be obtained from W_1 by repeated insertions or deletions of $aa^{-1}, a^{-1}a, bb^{-1}, \ldots$, the relators $R_\mu, R_\nu, \ldots$ and their inverses $R_\mu^{-1}, R_\nu^{-1}, \ldots$. As an example, note that in $A = \langle a, b \mid a^3b, b^3, a^4 \rangle$ we have $a = aa^3b = a^4b = b$. Hence $1 = a^3b(b^3)^{-1} = a = b$ and so A is the trivial group of order 1.

Even the trivial example above illustrates our difficulties when faced with a group given in terms of defining relations. However, groups defined by means of a presentation arise naturally in diverse areas of mathematics, especially in knot theory, topology and geometry, so we have to try to overcome these difficulties. The fundamental problems concerning group presentations were formulated by Max Dehn in 1911. These problems ask for general and effective methods for deciding in a finite number of steps (i) whether two given words represent the same group element (*the word problem*), (ii) whether they represent conjugate elements (*the conjugacy problem*) and (iii) whether two finitely presented groups are isomorphic (*the isomorphism problem*). All these problems have turned out to be problems in logic, and cannot be solved in general. Explicit solutions of these problems are always based on specific presentations and often make use of group diagrams. (Dehn himself was a particularly enthusiastic advocate of group diagrams.)

Let $A = \langle a, b, \dots \mid R_\mu, R_\nu, \dots \rangle$. We shall attempt to construct the Cayley diagram of A with respect to the generators a, b, ... Having got the Cayley diagram, we clearly have a solution to the word problem for this presentation.

The Cayley diagram of a group has the following two properties.

(a) The (directed) edges have a *regular* colouring with a, b, ...; that is, for each vertex x and generator g there is exactly one edge coloured g starting at x and exactly one edge coloured g ending at x.

(b) Every relation is satisfied at *every vertex*, that is, if x is a vertex and R_μ is a relator then the walk starting at x corresponding to R_μ ends at x.

How shall we go about finding the Cayley diagram? We try to satisfy (a) and (b) *without ever identifying two vertices unless we are forced to do so.* Thus at each stage we are free to take a *new* vertex and an edge into it (or from it). We identify two vertices when (a) or (b) forces us to do so. If the process stops, we have arrived at the Cayley diagram. Note that until the end we do *not* know that distinct vertices represent distinct group elements.

As an example, let us see how we can find the Cayley diagram of $A = \langle a, b \mid a^3 = b^2 = (ab)^2 = 1\rangle$. We replace each double edge corresponding to b by a single undirected edge; this makes the Cayley diagram into a graph, with some oriented edges. We start with the identity and with a triangle 123 corresponding to $a^3 = 1$; for simplicity we use numbers 1, 2, ... to denote the vertices, reserving 1 for the identity element. An edge coloured b must start at each of the vertices 1, 2 and 3, giving vertices 4, 5 and 6. Now $a^3 = 1$ must be satisfied at 6, giving another triangle, say 678, whose edges are coloured a, as in Fig. VIII.2. At this stage we may care to bring in the relation $(ab)^2 = abab = 1$. Checking it at 8, say, we see that the walk 86314 must end at 8, so the vertices denoted so far by 8 and 4 have to coincide. Next we check $abab = 1$ at 7: the walk $7(8 \equiv 4)125$ must end at 7 so 5 and 7 are identical. All that remains to check is that the diagram we

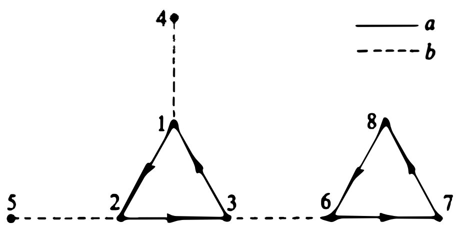


FIGURE VIII.2. Construction of a Cayley diagram.

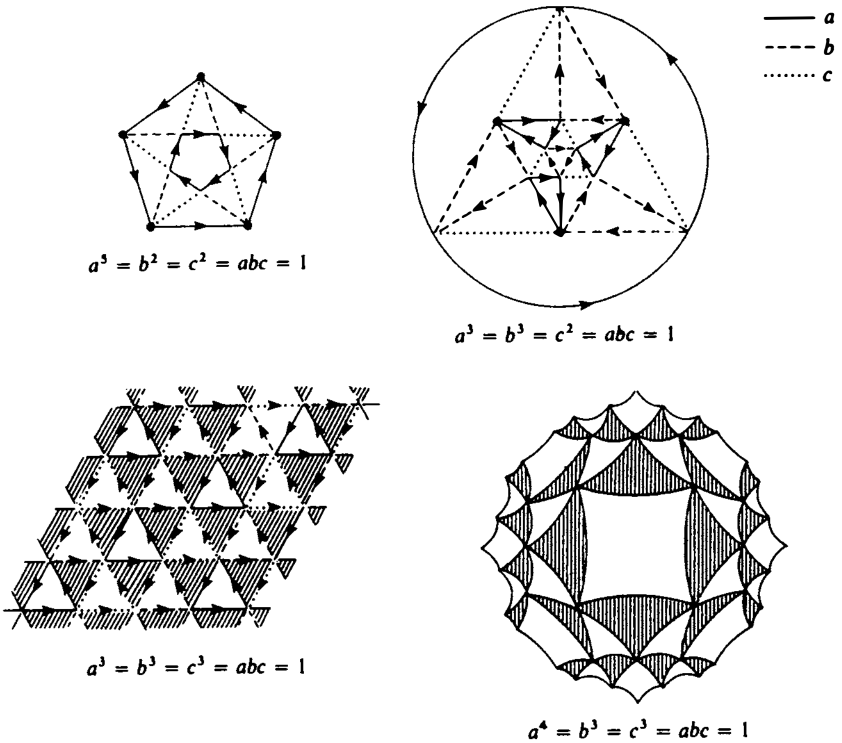


FIGURE VIII.3. Some Cayley diagrams. The shaded regions correspond to $abc = 1$.

obtained satisfies (a) and (b), so it is the Cayley diagram of the group in question. In fact, the diagram is exactly the third picture in Fig. VIII.1, so the group is S_3.

For $p \geq q \geq r \geq 2$ denote by (p, q, r) the *group* $\langle a, b, c \mid a^p = b^q = c^r = abc = 1\rangle$. Given specific values of p, q and r, with a little effort the reader can find the Cayley diagram of the group (p, q, r) with respect to the generators a, b and c. Fig. VIII.3 shows some of these diagrams.

The diagrams above indicate some connection with tessellations. The beauty of the use of Cayley diagrams is that one can make good use of this connection. Indeed, the reader who is slightly familiar with tessellations of the sphere, the Euclidean plane and the hyperbolic plane, can easily prove the following result.

Theorem 1 *If* $(1/p) + (1/q) + (1/r) > 1$ *then the group* (p, q, r) *is finite and has order* $2s$ *where* $1/s = (1/p) + (1/q) + (1/r) - 1$. *The Cayley diagram is a tessellation of the sphere (as in the first two pictures in Fig. VIII.3).*

If $(1/p) + (1/q) + (1/r) \leq 1$ *then the group* (p, q, r) *is infinite. If equality holds, then the Cayley diagram is a tessellation of the Euclidean plane, while otherwise it is a tessellation of the hyperbolic plane (as in the last two pictures in Fig. VIII.3).* □

As we remarked earlier, groups given by means of their presentations arise frequently in knot theory. In particular, Dehn showed how a presentation of the group of a (tame) *knot* (that is, the fundamental group of $\mathbb{R}^3$ after the removal of the knot) can be read off from a projection of the knot into a plane. The projection of the knot forms the boundary of certain bounded domains of the plane. To each of these domains there corresponds a generator (the identity corresponds to the unbounded domain) and to each cross-over there corresponds a relator. The general form of these relators is easily deduced from the two examples shown in Fig. VIII.4. (Indeed, readers familiar with the fundamentals of algebraic topology can easily prove the correctness of this presentation.) The group of the *trefoil* (or *clover leaf*) *knot* is $\langle a, b, c, d \mid ad^{-1}b, cd^{-1}a, cbd^{-1}\rangle$ and the group of the *figure of eight knot* is $\langle a, b, c, d, e \mid ab^{-1}c, ad^{-1}eb^{-1}, ed^{-1}cb^{-1}, acd^{-1}\rangle$. In Section X.5 we shall study knots in a completely different way, namely by means of polynomial invariants rather than groups.

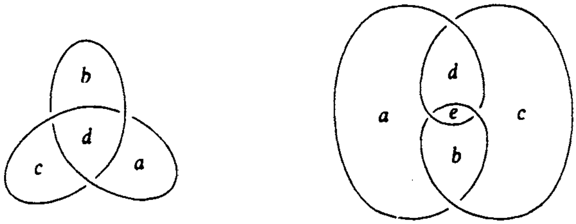


FIGURE VIII.4. The (right-handed) trefoil knot and the figure of eight.

Of course, before embarking on an investigation of the group, it is sensible to attempt to simplify the presentation. For example, $cbd^{-1} = 1$ means that $d = cb$ or $bd^{-1}c = 1$. Thus the group of the trefoil knot is

$$\langle a, b, c, d \mid ad^{-1}b, bd^{-1}c, cd^{-1}a\rangle$$

or, equivalently,

$$\langle a, b, c \mid cb = ba = ac\rangle.$$

We invite the reader to check that the Cayley diagram of this group is made up of replicas of the ladder shown in Fig. VIII.5. (Exercise 4). At each edge three ladders are glued together in such a way that when looking at these ladders from above, we see an infinite cubic tree (Fig. VIII.5). Having obtained the Cayley diagram, we can read off the properties we are interested in. In the case of this group the method does not happen to be too economical, but this is the way Dehn proved in 1910 that the group of the trefoil knot is not the group of a circle, which is the infinite cyclic group.

Schreier diagrams can be constructed analogously to Cayley diagrams. In fact, in order to determine the structure of a largish group given by means of a presentation, it is often advantageous to determine first the Schreier diagram of a subgroup. In order to show this, we work through another example, once again due to Dehn. What is the group

$$A = \langle a, b \mid a^2 = b^5 = (ba)^3 = 1\rangle?$$

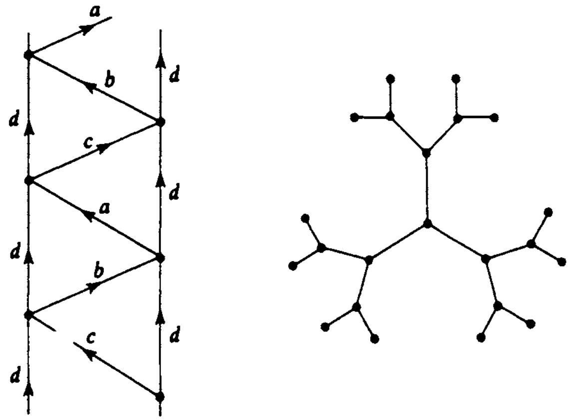


FIGURE VIII.5. The ingredients of the Cayley diagram of the trefoil knot.

Let us construct the Schreier diagram of the cosets of the subgroup B generated by b. As before, we take a vertex 1 for B and try to construct as big a diagram as conditions (a) and (b) allow us. (Recall that (a) requires the colouring to be regular and (b) requires that *each* defining relation is satisfied at *each* vertex.) However, in this case there is one more condition: the edge coloured b starting at 1 must end in 1 (so it is a loop) since $Bb = B$. Thus after two steps we have the diagram shown in Fig. VIII.6. (Once again the edges coloured a will not be directed since $a^2 = 1$.)

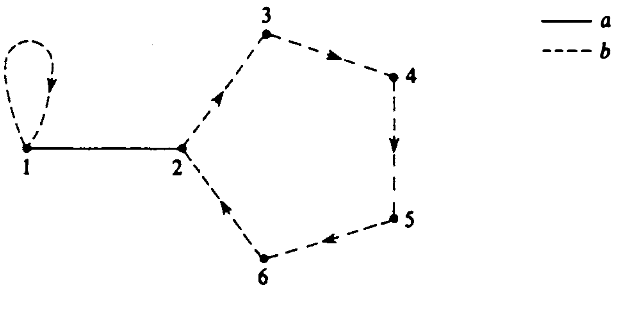


FIGURE VIII.6. The initial part of the Schreier diagram of $A \bmod B$.

Now let us check the condition $bababa = 1$ at vertex 6. The walk $babab$ takes us from 6 to 3, so there must be an edge coloured a from 3 to 6. In order to have edges coloured a starting at 4 and 5, we take up new vertices 7 and 8, together with edges 47 and 58 coloured a. Next we check the condition $ab^{-1}ab^{-1}ab^{-1} = 1$, which is equivalent to $(ba)^3 = 1$ at 7, and find that there is an edge from 7 to 8 coloured b. To satisfy $b^5 = 1$ at 7 we take three new vertices, 9, 10, and 11. Checking $(ba)^3 = 1$ at 11 we find that there is an edge from 9 to 11 coloured a. At this stage we are almost home, but no edge coloured a begins at 10, so we take a new vertex 12 joined to 10 by an edge coloured a. What does the condition $ab^{-1}ab^{-1}ab^{-1} = 1$ tell us at vertex 12? The walk $ab^{-1}ab^{-1}a$ starting at 12 ends

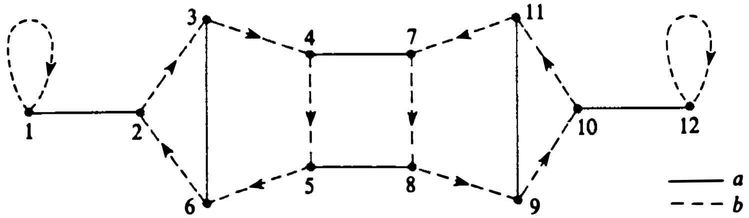


FIGURE VIII.7. The complete Schreier diagram.

at 12, so there must be an edge coloured b *starting* at 12 and *ending* at 12, giving us Fig. VIII.7. This is the Schreier diagram we have been looking for, since the colouring is clearly regular and it is a simple matter to check that *each* defining relation is satisfied at *each* vertex. In fact, $a^2 = 1$ and $b^5 = 1$ are obviously satisfied; since $(ba)^3 = 1$ holds at 6 by construction, it also holds at each vertex of the walk 621236, etc.

This detailed and cumbersome description fails to do justice to the method which, when performed on a piece of paper or on a blackboard, is quick and efficient. The reader is encouraged to find this out for himself.

What the Schreier diagram certainly tells us is the *index* of B: it is simply the number of vertices. Indeed, Schreier diagrams are often constructed on computers just to determine the index of a subgroup. In some cases, as in the example above, it tells us considerably more. The Schreier diagram is essentially a shorthand for the representation of A as a group of permutations of the cosets of B. In this case $a \to (12)(36)(47)(58)(9\ 11)(10\ 12)$ and $b \to (1)(2\ 3\ 4\ 5\ 6)(7\ 8\ 9\ 10\ 11)(12)$. Since the permutation corresponding to b has order 5, which is exactly the order of b in A, we see that A is a group of order $5 \cdot 12 = 60$ and is, in fact, isomorphic to A_5, the alternating group on 5 symbols.

If we want our Schreier diagram to carry more information, then we fix certain representatives of the cosets and keep track of the effect of the generators on these representatives. We *decorate* each *coset* by its representative: if H is decorated by $[h]$ then $H = Bh$. Now, if there is an edge coloured a from $H = Bh$ to $K = Bk$, then we *decorate* this *edge* by $[\alpha]$ if $ha = \alpha k$. Since $K = Bk = Ha = Bha$, we see that $\alpha \in B$, so the edges are decorated with elements of B. Furthermore, if H, K and L are decorated with h, k and l, and there are edges coloured a, b, c and decorated $[\alpha]$, $[\beta]$ and $[\gamma]$ joining them, as in Fig. VIII.8, then $habc = \alpha kbc = \alpha\beta lc = \alpha\beta\gamma h$. In particular, if $h = 1$ then we have $abc = \alpha\beta\gamma$. Of course, an analogous assertion holds for arbitrary walks starting and ending at B: the product of the colours equals the product of the decorations.

One of the simplest ways of decorating the vertices and edges makes use of spanning trees. Select a spanning tree of the Schreier diagram. Decorate B, the subgroup itself, by 1 (the identity) and decorate the edges of the spanning tree also by 1. This determines the decoration of *every vertex* (that is, every coset) and *every edge*. Indeed, for each vertex H the spanning tree contains a unique path from B to H; clearly, H has to be decorated with the product $abc \cdots$ corresponding to this

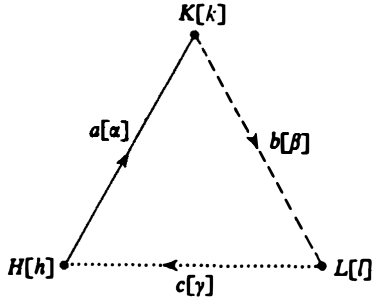


FIGURE VIII.8. A cycle in a decorated Schreier diagram.

path. These coset representatives are said to form a *Schreier system for* B mod A. What is the decoration of a chord HK, an edge not in the tree? By the remark above it is the product of the colours on the B–H path, the edge HK and the K–B path, as in Fig. VIII.9.

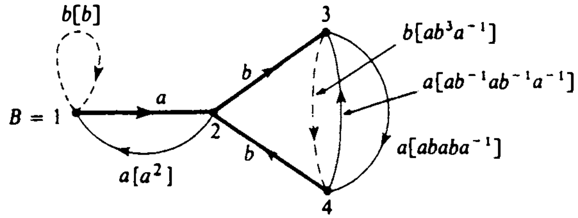


FIGURE VIII.9. The decorations induced by a spanning tree.

Since each element of B is the product of the colours on a closed walk from B to B in the Schreier diagram, the decorations of the chords generate B. Thus from a Schreier diagram we can read off a set of generators of B, *independently* of the structure of A.

Theorem 2 *The subgroup B of A is generated by the decorations of the chords.* $\square$

In particular, the subgroup B in Fig. VIII.9 is generated by b, ab^3a^{-1}, $ab^{-1}ab^{-1}a^{-1}$ and $ababa^{-1}$.

It is equally simple to find a presentation of B, provided that we have a presentation of A. This is obtained by the *Reidemeister–Schreier rewriting process*; we give a quick and loose description of it. The generators of the presentation are the chords of the spanning tree; to distinguish chords of the same colour we write c_i for the edge coloured c starting at vertex i. For each vertex i and each relator R_μ denote by R^i_μ the (word of the) walk starting at i given by R_μ expressed as a product of the c_j, say $R^i_\mu = b_i c_j \cdots$. The reader can easily fill in the missing details in the proof of the following beautiful result, due to Reidemeister and Schreier.

Theorem 3 *The subgroup B has a presentation*

$$\langle a_1, b_1, \dots, a_2, b_2, \dots \mid R^1_\mu, R^1_\nu, \dots, R^2_\mu, R^2_\nu, \dots \rangle. \qquad \square$$

Now, if we wish to preserve the connection between this presentation of B and the original presentation of A, we simply equate c_i with the decoration of the edge coloured c starting at vertex i.

If A is a free group, its presentation contains no relations. Hence the above presentation of B contains no relations either, so B is also a free group. This is a fundamental result of Nielsen and Schreier.

Theorem 4 *A subgroup of a free group is free. Furthermore, if A is a free group of rank k (that is, it has k free generators) and B is a subgroup of index n, then B has rank $(k-1)n+1$.*

Proof. The presentation of B given in Theorem 3 is a free presentation on the set of chords of the Schreier diagram. Altogether there are kn edges of which $n-1$ are tree edges; hence there are $(k-1)n+1$ chords. $\square$

It is amusing to note that Theorem 4 implies that the rank of a free group is well defined. Indeed, suppose that A is freely generated by a, b, ... Then *every* directed multigraph with a vertex (corresponding to the subgroup) and a regular colouring (by a, b, ...) is the Schreier diagram of some subgroup B of A, for there are no relations to be satisfied. Hence subgroups of index n are in 1-to-1 correspondence with regularly coloured multigraphs of order n. In particular, if A has $k \geq 2$ generators, it has $2^k - 1$ subgroups of index 2, for there are 2^k multigraphs of order 2 regularly coloured with k colours, but one of those is disconnected. The number of subgroups of index 2 is clearly independent of the presentation, so k is determined by A.

VIII.2 The Adjacency Matrix and the Laplacian

Recall from Section II.3 that the vertex space $C_0(G)$ of a graph G is the complex vector space of all functions from $V(G)$ into $\mathbb{C}$. Once again we take $V(G) = \{v_1, v_2, \dots, v_n\}$, so that $\dim C_0(G) = n$, and we write the elements of $C_0(G)$ in the form $\mathbf{x} = \sum_{i=1}^n x_i v_i$ or $\mathbf{x} = (x_i)_1^n$; here x_i is the value of $\mathbf{x}$ at v_i, also called the *weight at* v_i. The space $C_0(G)$ is given the natural inner product associated with the basis $(v_i)_1^n$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i \overline{y}_i$. The *norm* of $\mathbf{x}$ is $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$.

As in the paragraph above, bold letters ($\mathbf{x}$, $\mathbf{y}$, ...) will be used for vectors only if we wish to emphasize that they are vectors; for example, v_i denotes both a vertex and the corresponding basis vector. This slight inconsistency is unlikely to lead to confusion.

First we shall consider the adjacency matrix $A = A(G) = (a_{ij})$ of G, the $0-1$ matrix where $a_{ij} = 1$ iff $v_i v_j$ is an edge. As usual, A is identified with a linear endomorphism of $C_0(G)$. To start with, we recollect some simple facts from linear algebra. The matrix A is real and symmetric, so it is *hermitian*, that is,

$\langle A\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A\mathbf{y}\rangle$. Hence its *numerical range*

$$V(A) = \{\langle A\mathbf{x}, \mathbf{x}\rangle : \|\mathbf{x}\| = 1\}$$

is a closed interval of the real line. The *distinct* eigenvalues of A are *real*, say $\mu_1 > \mu_2 > \cdots > \mu_t$, and $V(A)$ is exactly the interval $[\mu_t, \mu_1]$. (Our notation here is not entirely standard: it is customary to write $\lambda_1, \lambda_2, \ldots$ for the eigenvalues rather than $\mu_1, \mu_2, \ldots$. However, we reserve the λ_i for the eigenvalues of the combinatorial Laplacian.) For simplicity, an eigenvalue of A is said to be an *eigenvalue of* G. We shall write $\mu_{\max}(G)$ for the maximal eigenvalue μ_1 and $\mu_{\min}(G)$ for the minimal eigenvalue μ_t. If G has at least one edge, say $v_1v_2 \in E(G)$, then $\langle A\mathbf{x}, \mathbf{x}\rangle > 0$ if $\mathbf{x} = (1, 1, 0, \ldots)$, and $\langle A\mathbf{x}, \mathbf{x}\rangle < 0$ if $\mathbf{x} = (1, -1, 0, \ldots, 0)$. Hence $\mu_{\min} < 0 < \mu_{\max}$, unless G is empty.

The inner product space $C_0(G)$ has an orthonormal basis consisting of eigenvectors of A. In particular, writing $m(\mu)$ for the (geometric or algebraic) *multiplicity* of an eigenvalue μ, we have $\sum_{i=1}^t m(\mu_i) = n$. Since $a_{ii} = 0$ for every i, the *trace* of A is 0: $\operatorname{tr} A = \sum_{i=1}^n a_{ii} = 0$. In an orthonormal basis consisting of eigenvectors of A, the trace of A is $\sum_{i=1}^t m(\mu_i)\mu_i$; as a change of basis does not alter the trace, $\sum_{i=1}^t m(\mu_i)\mu_i = 0$.

We collect some further basic properties of the eigenvalues in the following theorem.

Theorem 5 *Let G be a connected graph of order n with adjacency matrix A.*

(i) Every eigenvalue μ of G satisfies $|\mu| \le \Delta = \Delta(G)$.

(ii) The maximal degree Δ is an eigenvalue of G iff G is regular; if Δ is an eigenvalue then $m(\Delta) = 1$.

(iii) If $-\Delta$ is an eigenvalue of G then G is regular and bipartite.

(iv) If G is bipartite and μ is an eigenvalue of G then so is $-\mu$, and $m(\mu) = m(-\mu)$.

(v) The maximal eigenvalue satisfies $\delta(G) \le \mu_{\max}(G) \le \Delta(G)$.

(vi) If H is an induced subgraph of G then $\mu_{\min}(G) \le \mu_{\min}(H) \le \mu_{\max}(H) \le \mu_{\max}(G)$.

Proof. (i) Let $\mathbf{x} = (x_i)$ be a non-zero eigenvector with eigenvalue μ. Let x_p be a weight with maximum modulus: $|x_p| \ge |x_i|$ for every i; we may assume without loss of generality that $x_p = 1$. Then

$$|\mu| = |\mu x_p| = \left|\sum_{l=1}^n a_{pl}x_\ell\right| \le \sum_{l=1}^n a_{pl}|x_\ell| \le |x_p| d(v_p) \le |x_p|\Delta = \Delta,$$

showing $|\mu| \le \Delta$.

(ii) If $\mu = \Delta$ is an eigenvalue and $\mathbf{x}$, x_p are chosen as in (i), then

$$\Delta = \Delta x_p = \sum_{\ell=1}^n a_{p\ell}x_\ell$$

and $x_\ell \le 1$ imply that $d(v_p) = \Delta$ and $x_\ell = x_p = 1$ whenever v_ℓ is adjacent to v_p. In turn, this implies that $d(v_\ell) = \Delta$ and $x_k = x_\ell = 1$ whenever v_k is adjacent to

v_ℓ, and so on; as G is connected, $d(v_i) = \Delta$ and $x_i = 1$ for every i. Hence G is Δ-regular and $\mathbf{x}$ is $\mathbf{j}$, the vector all of whose entries are equal to 1.

Conversely, if G is Δ-regular, then $(A\mathbf{j})_i = \sum_{l=1}^n a_{il} = \Delta$, so $A\mathbf{j} = \Delta\mathbf{j}$.

(iii) If $\mu = -\Delta$ is an eigenvalue then, as in (ii), we find that $d(v_p) = \Delta$ and $x_\ell = -x_p = -1$ whenever v_ℓ is adjacent to v_p. As in (ii), this implies that G is Δ-regular. Furthermore, at each vertex v_k adjacent to v_ℓ the weight is 1, at each neighbour of v_k it is -1, and so on. The weight is 1 at the vertices at an even distance from v_p and it is -1 at the other vertices; also, every edge joins vertices of different weights. Thus G is bipartite, say $V = V_1 \cup V_2$, where $v_p \in V_1$.

(iv) Suppose G is bipartite with vertex classes V_1 and V_2. Let $\mathbf{b}$ be the function (vector) that is, 1 on V_1 and -1 on V_2. Then $\mathbf{x} \to \mathbf{bx} = (b_i x_i)_1^n$ is an automorphism of the vector space $C_0(G)$. Now, if $A\mathbf{x} = \mu\mathbf{x}$ and $v_i \in V_1$, say, then

$$\begin{aligned}(A(\mathbf{bx}))_i &= \sum_{j=1}^n a_{ij} b_j x_j = \sum_{v_j \in V_1} a_{ij} x_j - \sum_{v_j \in V_2} a_{ij} x_j = -\sum_{v_j \in V_2} a_{ij} x_j \\ &= -\sum_{v_j \in V_1} a_{ij} x_j - \sum_{v_j \in V_2} a_{ij} x_j = -\sum_{j=1}^n a_{ij} x_j = -\mu x_i = -\mu(\mathbf{bx})_i.\end{aligned}$$

Hence, writing I_n for the n by n identity matrix, we find that $\mathbf{b}$ gives an isomorphism between $\ker(A - \mu I_n)$ and $\ker(A + \mu I_n)$. In particular, $m(\mu) = m(-\mu)$.

(v) We know already that $\mu_{\max}(G) \le \Delta(G)$. Note that for $\mathbf{j} = (1, 1, \dots, 1)$ we have $\langle \mathbf{j}, \mathbf{j} \rangle = n$, so $V(A)$ contains

$$\frac{1}{n}\langle A\mathbf{j}, \mathbf{j} \rangle = \frac{1}{n}\sum_{k=1}^n \sum_{l=1}^n a_{kl} = \frac{1}{n}\sum_{k=1}^n d(v_k) \ge \delta(G).$$

Hence $\mu_{\max}(G) = \max V(A) \ge \delta(G)$.

(vi) It suffices to prove the result for an induced subgraph H of order $n - 1$, say with $V(H) = \{v_1, v_2, \dots, v_{n-1}\}$.

Let A' be the adjacency matrix of H. Then there is a vector $\mathbf{y} \in C_0(H)$ such that $\langle \mathbf{y}, \mathbf{y} \rangle = 1$ and $\langle A'\mathbf{y}, \mathbf{y} \rangle = \mu_{\max}(H)$. Let $\mathbf{x} = (y_1, y_2, \dots, y_{n-1}, 0)$. Then $\mathbf{x} \in C_0(G)$, $\langle \mathbf{x}, \mathbf{x} \rangle = 1$ and $\langle A\mathbf{x}, \mathbf{x} \rangle = \langle A'\mathbf{y}, \mathbf{y} \rangle = \mu_{\max}(H) \in V(A)$. Consequently, $\mu_{\max}(G) \ge \mu_{\max}(H)$. The other inequality is proved analogously. □

Let us note an immediate consequence of Theorem 5 (v) and (vi), concerning the chromatic number.

Corollary 6 *Every graph G satisfies $\chi(G) \le \mu_{\max}(G) + 1$.*

Proof. For every induced subgraph H of G we have

$$\delta(H) \le \mu_{\max}(H) \le \mu_{\max}(G),$$

so we are done by Theorem V.1. □

In fact, with a little work one can also give a lower bound on the chromatic number in terms of the eigenvalues.

Theorem 7 *Let G be a non-empty graph. Then*

$$\chi(G) \geq 1 - \frac{\mu_{\max}(G)}{\mu_{\min}(G)}.$$

Proof. As before, we take $V = \{v_1, \ldots, v_n\}$ for the set of vertices, so that $(v_1, \ldots, v_n)$ is the canonical basis of $C_0(G)$. Let $c : V(G) \to [k]$ be a (proper) colouring of G with $k = \chi(G)$ colours. Then, writing $\langle \mathbf{a}, \mathbf{b}, \ldots \rangle$ for the space spanned by the vectors $\mathbf{a}, \mathbf{b}, \ldots$, the space $C_0(G)$ is the orthogonal direct sum of the 'colour spaces' $U_i = \langle v_j : c(v_j) = i \rangle$, $i = 1, \ldots, k$. Since no edge joins vertices of the same colour, the adjacency matrix $A = A(G)$ is such that if $u, w \in U_i$ for some i then $\langle Au, w \rangle = 0$. In particular, $\langle Au, u \rangle = 0$ for $u \in U_i$, $i = 1, \ldots, k$.

Let $\mathbf{x} \in C_0(G)$ be an eigenvector of the adjacency matrix A with eigenvalue $\mu_{\max}$, and let $\mathbf{x} = \sum_{i=1}^k \xi_i u_i$, where $u_i \in U_i$ and $\|u_i\| = 1$. Let $U = \langle u_1, \ldots, u_k \rangle$, so that $(u_1, \ldots, u_k)$ is an orthonormal basis of U, and let $S : U \to C_0(G)$ be the inclusion map.

For $\mathbf{u} \in U$, $\|\mathbf{u}\| = 1$, we have $\|S\mathbf{u}\| = \|\mathbf{u}\| = 1$, so $\langle S^*AS\mathbf{u}, \mathbf{u} \rangle = \langle AS\mathbf{u}, S\mathbf{u} \rangle = \langle A\mathbf{u}, \mathbf{u} \rangle \in V(A)$. Hence the numerical range of the hermitian operator S^*AS is contained in the numerical range of A:

$$V(S^*AS) \subset V(A) = [\mu_{\min}, \mu_{\max}].$$

In fact, $\mu_{\max}$ is an eigenvalue of S^*AS as well, with eigenvector x:

$$\langle S^*ASx, u_i \rangle = \langle Ax, u_i \rangle = \mu_{\max} \langle x, u_i \rangle = \mu_{\max} \xi_i,$$

so $S^*ASx = \mu_{\max} x$.

Also, $\langle S^*ASu_i, u_i \rangle = \langle Au_i, u_i \rangle = 0$ for every i, so $\operatorname{tr}(S^*AS) = 0$. Therefore, since every eigenvalue of S^*AS is at least $\mu_{\min}$,

$$\mu_{\max} + (k-1)\mu_{\min} \leq \operatorname{tr}(S^*AS) = 0.$$

As G is non-empty, $\mu_{\min} < 0$, so the result follows. □

The quadratic form $\langle A\mathbf{x}, \mathbf{x} \rangle$ appearing in the definition of the numerical range and in the proofs of Theorems 5 and 7 is sometimes called the *Lagrangian* of G, and is denoted by $f_G(\mathbf{x})$:

$$f_G(\mathbf{x}) = \langle A\mathbf{x}, \mathbf{x} \rangle = \sum_{i,j=1}^n a_{ij} x_i x_j = \sum_{v_i \sim v_j} x_i x_j.$$

Note that every edge $v_i v_j$ contributes $2x_i x_j$ to the Lagrangian: $x_i x_j$ for $v_i \sim v_j$ and $x_j x_i = x_i x_j$ for $v_j \sim v_i$.

Before we illustrate the use of the Langrangian, let us note some additional simple facts from linear algebra. Let W be an n-dimensional complex inner product space, and let $T : W \to W$ be a hermitian operator on W. Let W_+ be the subspace of W spanned by the eigenvectors of T with strictly positive eigenvalues; define W_- similarly, for negative eigenvalues, and set $W_0 = \ker T$. Then W is the orthogonal direct sum of these three subspaces: $W = W_+ \oplus W_- \oplus W_0$. Set $n_+(T) = \dim W_+$, $n_-(T) = \dim W_-$, and $n_0(T) = \dim W_0$, so that $n = n_+ + n_- + n_0$.

Let $q(\mathbf{x}) = q_T(\mathbf{x}) = \langle T\mathbf{x}, \mathbf{x}\rangle$ be the quadratic form associated with T. Note that q is *positive definite* on W_+ and *negative definite* on W_-, that is, $q(\mathbf{x}) > 0$ and $q(\mathbf{y}) < 0$ for all $\mathbf{x} \in W_+$, $\mathbf{y} \in W_-$, $\mathbf{x}, \mathbf{y} \neq 0$. Similarly, q is *positive semi-definite* on $W_+ \oplus W_0$ and *negative semi-definite* on $W_- \oplus W_0$: if $\mathbf{x} \in W_+ \oplus W_0$ and $\mathbf{y} \in W_- \oplus W_0$ with $\mathbf{x}, \mathbf{y} \neq 0$ then $q(\mathbf{x}) \geq 0$ and $q(\mathbf{y}) \leq 0$. The subspaces above also have maximal dimensions with respect to these properties; for example, if q is positive semi-definite on a subspace $U \subset W$ then

$$\dim U \leq n_+ + n_0 = \dim(W_+ \oplus W_0). \tag{1}$$

Indeed, if $\dim U > n_+ + n_0$ then $U \cap W_-$ contains a non-zero vector $\mathbf{x}$; as $\mathbf{x} \in W_-$, we have $q(\mathbf{x}) < 0$.

These simple facts imply the following connection between the independence number $\beta(G)$ and the distribution of the eigenvalues.

Theorem 8 *The adjacency matrix of a graph G has at least $\beta(G)$ non-negative and at least $\beta(G)$ non-positive eigenvalues, counted with multiplicity.*

Proof. The Lagrangian $f_G(\mathbf{x})$ is identically 0 on every subspace spanned by a set of independent vertices. In particular, f_G is positive semi-definite and negative semi-definite on a subspace of dimension $\beta(G)$. Hence we are done by the analogues of (1). □

Ideally, one would like to determine the entire *spectrum* of a graph, that is, all the eigenvalues and their multiplicities. Needless to say, in most cases this is out of the question, and we have to be satisfied with various bounds. However, in some simple cases it is easy to determine the spectrum. For example, it is trivial that the empty graph $E_n = \overline{K}_n$ has one eigenvalue, 0, with multiplicity n. More generally, adding an isolated vertex to a graph G just increases by one the multiplicity of 0. It is only a little less trivial that the complete graph K_n has eigenvalues $\mu_1 = n - 1$ and $\mu = -1$, with multiplicities $m(n-1) = 1$ and $m(-1) = n - 1$. Indeed, $\sum_{i=1}^n x_i v_i$ is an eigenvector with eigenvalue -1 if $\sum_{i=1}^n x_i = 0$.

The complete bipartite graph $K_{k,n-k}$ has three eigenvalues: $(k(n-k))^{1/2}$ and $-(k(n-k))^{1/2}$, each with multiplicity one, and 0, with multiplicity $n - 2$. If $U = \{v_1, \ldots, v_k\}$ and $W = \{w_1, \ldots, w_{n-k}\}$ are the two classes then $\sum_{i=1}^k x_i v_i + \sum_{j=1}^{n-k} y_j w_j$ is an eigenvector with eigenvalue 0 if

$$\sum_{i=1}^k x_i = \sum_{j=1}^{n-k} y_j = 0.$$

These simple facts about spectra suffice to give us the following theorem of Graham and Pollak.

Theorem 9 *The complete graph K_n is not the edge-disjoint union of $n - 2$ complete bipartite graphs.*

Proof. Suppose that, contrary to the assertion, K_n is the edge-disjoint union of complete bipartite graphs $G_1, \ldots, G_{n-2}$. For each i, let H_i be obtained from G_i

by adding to it isolated vertices so that $V(H_i) = V(K_n)$. Note that the Lagrangians of these graphs are such that $f_{K_n} = \sum_{i=1}^{n-2} f_{H_i}$.

We know that each f_{H_i} is positive semi-definite on some subspace $U_i \subset C_0(K_n)$ of dimension $n-1$. But then $U = \bigcap_{i=1}^{n-2} U_i$ is a subspace of dimension at least 2, on which each f_{H_i} is positive semi-definite. Hence $f_{K_n} = \sum_{i=1}^{n-2} f_{H_i}$ is positive semi-definite on U, contradicting the fact that f_{K_n} is not positive semi-definite on any subspace of dimension 2. □

Clearly, the simple argument above proves the following more general assertion. Suppose a graph G of order n is the edge-disjoint union of $n-r$ complete bipartite graphs. Then the quadratic form of G is positive semi-definite on some subspace of dimension r, and negative semi-definite on some subspace of dimension r.

In 1965, Motzkin and Straus showed that one can use the Lagrangian to give yet another proof of a slightly weaker form of Turán's theorem (Theorem IV. 8); this is our final application of the adjacency matrix and the Lagrangian. Consider the simplex $S = S_n = \{\mathbf{x} = (\mathbf{x}_i)_1^n \in \mathbb{R}^n : \sum_{i=1}^n \mathbf{x}_i = 1 \text{ and } \mathbf{x}_i \geq 0 \text{ for every } n\}$, and set

$$f(G) = \max_{\mathbf{x} \in S} f_G(\mathbf{x}).$$

It is immediate from the definition of f that it is an increasing function: if $H \subset G$ then $f(H) \leq f(G)$. Furthermore, if $f_G(\mathbf{x})$ attains its supremum at $\mathbf{x} = \mathbf{y}$, and $H = G[W]$ is the subgraph of G induced by the support of $\mathbf{y}$:

$$W = \operatorname{supp} \mathbf{y} = \{v_i : y_i > 0\},$$

then $f(G) = f(H)$.

As the theorem of Motzkin and Straus below shows, $f(G)$ is intimately related to the complete subgraphs of G: in fact, it depends only on the *clique number* $\omega(G)$ of G, the maximal order of a complete subgraph. Note first that if G is a complete graph of order n then $f(G) = (n-1)/n$. Indeed,

$$\begin{aligned} f(G) &= \max\left\{2 \sum_{1 \leq i < j \leq n} x_i x_j : \mathbf{x} \in S\right\} \\ &= \max\left\{\sum_{i=1}^n x_i(1-x_i) : \mathbf{x} \in S\right\} \\ &= \max\left\{1 - \sum_{i=1}^n x_i^2 : \mathbf{x} \in S\right\} \\ &= 1 - n(1/n)^2 = (n-1)/n. \end{aligned}$$

Theorem 10 *Let G be a graph with clique number k_0. Then $f(G) = (k_0 - 1)/k_0$.*

Proof. Let $\mathbf{y} = (y_i)_1^n \in S$ be a point at which $f_G(\mathbf{x})$ attains its maximum and $\operatorname{supp} \mathbf{y} = \{v_i : y_i > 0\}$ is as small as possible. We claim that the support of $\mathbf{y}$ spans a complete subgraph of G. Indeed, suppose $y_1, y_2 > 0$ and $v_1 \not\sim v_2$. Assuming, as we may, that $\sum_{v_i \sim v_1} y_i \geq \sum_{v_i \sim v_2} y_i$, set $\mathbf{y}' = (y_1 + y_2, 0, y_3, y_4, \ldots, y_n) \in S$.

Then $f_G(\mathbf{y}') \geq f_G(\mathbf{y})$ and supp $\mathbf{y}'$ is strictly smaller than supp $\mathbf{y}$, contradicting our choice of $\mathbf{y}$.

Writing K for the complete subgraph of G spanned by the support of $\mathbf{y}$, we have $f(G) = f(K) = (k-1)/k$, where $k = |K| = |\text{supp}\, \mathbf{y}|$. Hence k is as large as possible, namely k_0, and we are done. □

The result above implies the following assertion, which is only slightly weaker than Turán's theorem.

Corollary 11 *Let $G = G(n, m)$, with $m > \frac{r-2}{2(r-1)}n^2$. Then G contains a complete graph of order r.*

Proof. Writing $k_0 = \omega(G)$ for the clique number of G, we know that $f(G) = (k_0 - 1)/k_0$. On the other hand, with $\mathbf{x} = (1/n, 1/n, \ldots, 1/n)$ we see that

$$f(G) \geq f_G(\mathbf{x}) = \frac{2m}{n^2} > \frac{r-2}{r-1}.$$

Hence $k_0 \geq r$, as claimed. □

Recall from Chapter II that, for a graph G with vertex set $\{v_1, \ldots, v_n\}$ and adjacency matrix A, the (combinatorial) *Laplacian* of G is $L = D - A$, where $D = (D_{ij})$ is the diagonal matrix in which D_{ii} is the degree $d(v_i)$ of v_i. The Laplacian is an even more powerful tool than the Lagrangian, although for a regular graph they are just two sides of the same coin.

In our study of the Laplacian, we shall need a simple and useful characterization of the spectrum of a hermitian operator T on an n-dimensional complex inner product space V. Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be the eigenvalues of T, enumerated with multiplicities, and let $q(\mathbf{x}) = \langle T\mathbf{x}, \mathbf{x}\rangle$ be the quadratic form of T. We know that the numerical range of T is $[\lambda_1, \lambda_n]$:

$$\lambda_1 = \min\{q(\mathbf{x}) : \|\mathbf{x}\| = 1\} \text{ and } \lambda_n = \max\{q(\mathbf{x}) : \|\mathbf{x}\| = 1\}.$$

In fact, if $q(\mathbf{x}_1) = \lambda_1$ and $\|\mathbf{x}_1\| = 1$ then $\mathbf{x}_1$ is an eigenvector of T with eigenvalue λ_1, and

$$\lambda_2 = \min\{q(\mathbf{x}) : \langle \mathbf{x}, \mathbf{x}_1\rangle = 0 \text{ and } \|\mathbf{x}\| = 1\}. \tag{2}$$

It is easily seen that the other eigenvalues have similar characterizations (see Exercises 47-48). As we shall see, the second smallest eigenvalue of the Laplacian is especially important.

The quadratic form $q(\mathbf{x}) = \langle (D-A)\mathbf{x}, \mathbf{x}\rangle$ associated with the Laplacian has a particularly pleasing form, emphasizing the intimate connection between the Laplacian and the structure of the graph: for $\mathbf{x} = \sum_{i=1}^n x_i v_i$ we have

$$q(\mathbf{x}) = \sum_{i=1}^n \{d(v_i)x_i^2 - \sum_{v_j \sim v_i} x_i x_j\} = \sum_{v_i v_j \in E(G)} (x_i - x_j)^2. \tag{3}$$

We shall write $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ for the sequence of eigenvalues of L, so that $C_0(G)$ has an orthonormal basis $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ with $L\mathbf{x}_i = \lambda_i \mathbf{x}_i$. If G is r-regular then μ is an eigenvalue of the adjacency matrix A iff $\lambda = r - \mu$ is an eigenvalue of the Laplacian L, so the spectrum of L is just the spectrum of A 'reversed and shifted'. In particular, if G is also connected then $\lambda_1 = r - \mu_1 = 0$ and $\lambda_2 = r - \mu_2 > 0$. In general, the connection between the spectra of A and L is a little less straightforward.

We know from Theorem II.10 that $L = BB^t$, where B is the (signed) incidence matrix of G. Consequently, L is positive semi-definite. Furthermore, as $L\mathbf{j} = \mathbf{0}$ for the vector $\mathbf{j}$ with all 1 coordinates, $\lambda_1 = 0$. However, $\lambda_2 = \lambda_2(G)$, the second smallest eigenvalue of the Laplacian, is far from trivial: in fact, it is difficult to overemphasize its importance. Roughly, the larger $\lambda_2(G)$ is, the more difficult it is to cut G into pieces, and the more G 'expands'. We present two results illustrating this assertion.

Before we turn to these results, let us adapt (2) to the case of the Laplacian. Since $L\mathbf{j} = (D - A)\mathbf{j} = 0$, with $q(\mathbf{x}) = \langle L\mathbf{x}, \mathbf{x}\rangle$ we have

$$\begin{aligned}\lambda_2(G) &= \min\left\{\frac{q(\mathbf{x})}{\|\mathbf{x}\|^2} : \langle \mathbf{x}, \mathbf{j}\rangle = 0, \mathbf{x} \neq 0\right\}\\ &= \min\left\{\frac{\langle (D-A)\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle} : \langle \mathbf{x}, \mathbf{j}\rangle = 0, \mathbf{x} \neq 0\right\}.\end{aligned} \tag{4}$$

If $G = K_n$ then $L = (n-1)I - A = nI - J$, so $\lambda_1 = \lambda_2 = \ldots = \lambda_{n-1} = n$ and $\lambda_n = 0$. In particular, if $n \geq 2$ then $\lambda_2(K_n) = n > \kappa(K_n) = n - 1$. However, if G is incomplete, this inequality cannot hold.

Theorem 12 *The vertex connectivity of an incomplete graph G is at least as large as the second smallest eigenvalue $\lambda_2(G)$ of the Laplacian of G.*

Proof. If $G = K_n$ then $\lambda_2 = n - 1 = \kappa(G)$. Suppose then that G is not a complete graph, and let $V' \cup S \cup V''$ be a partition of the vertex set $\{v_1, \ldots, v_n\}$ of G such that $|S| = \kappa(G)$, V' and V'' are non-empty, and G has no V'–V'' edge. Thus S is a vertex cut with $k = \kappa(G)$ vertices.

Our aim is to construct a vector $\mathbf{x}$ orthogonal to $\mathbf{j}$ such that $q(\mathbf{x})/\|\mathbf{x}\|^2$ is small, namely at most k. To this end, set $a = |V'|$, $b = |V''|$, and let $\mathbf{x} = \sum_{i=1}^n x_i v_i \in C_0(G)$ be the vector with

$$x_i = \begin{cases} b & \text{if } v_i \in V', \\ 0 & \text{if } v_i \in S, \\ -a & \text{if } v_i \in V''. \end{cases}$$

Then $\langle \mathbf{x}, \mathbf{j}\rangle = 0$ and $\|\mathbf{x}\|^2 = ab^2 + ba^2$.

What are the coordinates of $(D-A)\mathbf{x} = \mathbf{y} = \sum_{i=1}^n y_i v_i$? Since $(D-A)b\mathbf{j} = 0$, we have $\mathbf{y} = (D-A)(\mathbf{x} - b\mathbf{j})$, so if $v_i \in V'$ then y_i is precisely b times the number of neighbours of v_i in S. Hence $y_i \leq kb$. Similarly, $y_i \geq -ka$ for $v_i \in V''$. Therefore, as $|V'| = a$ and $|V''| = b$, it follows from (2) that

$$\lambda_2\|\mathbf{x}\|^2 \leq q(\mathbf{x}) = \langle (D-A)\mathbf{x}, \mathbf{x}\rangle \leq kab^2 + kba^2 = k\|\mathbf{x}\|^2,$$

completing the proof. □

It is easily seen that for every connectivity there are infinitely many graphs for which the bound on λ_2 in Theorem 12 is sharp (see Exercise 49).

The next result is the basic reason why $\lambda_2(G)$ is such an important parameter. Given a subset U of the vertex set of a graph G, the *edge-boundary* $\partial U = \partial_G U$ is the set of edges of G from U to $V \setminus U$.

Theorem 13 *Let G be a graph of order n. Then for $U \subset V = V(G)$ we have*

$$|\partial U| \geq \frac{\lambda_2(G)|U||V \setminus U|}{n}.$$

Proof. We may assume that $\emptyset \neq U \neq V = \{v_1, \ldots, v_n\}$. Set $k = |U|$, and define $\mathbf{x} = \sum_{i=1}^{n} x_i v_i$ as follows:

$$x_i = \begin{cases} n-k & \text{if } v_i \in U, \\ -k & \text{if } v_i \in V \setminus U. \end{cases}$$

Then $\langle \mathbf{x}, \mathbf{j} \rangle = 0$ and $\|\mathbf{x}\|^2 = kn(n-k)$. By (3),

$$\langle (D - A)\mathbf{x}, \mathbf{x} \rangle = |\partial U| n^2$$

and so, by (2),

$$\lambda_2(G) \leq |\partial U| n^2 / kn(n-k),$$

as claimed. □

For another connection between the expansion of a graph G and $\lambda_2(G)$, see Exercise 50.

VIII.3 Strongly Regular Graphs

It is reasonable to expect that a graph with many automorphisms will have particularly pleasing properties and that these will be reflected in the adjacency matrix. The *automorphism group* of a graph G is the group, Aut$G$, of permutations of the vertices preserving adjacency. Every abstract group can be represented as the automorphism group of some graph. For instance, if F is any finite group, consider its Cayley diagram with respect to some set of generators. The automorphism group of this *coloured* and *directed multigraph* is exactly F. It only remains to replace each edge of this diagram by a suitable subgraph that bears the information previously given by the direction and colour. This produces a graph G with automorphism group isomorphic to F. An example is shown in Fig. VIII.10.

Each $\pi \in$ Aut$G$ induces an endomorphism of $C_0(G)$, and this endomorphism is given by a permutation matrix P. In fact, an arbitrary permutation matrix Q corresponds to an automorphism of G precisely when it commutes with the adjacency matrix A, that is, $AQ = QA$. The group of these matrices therefore faithfully represents Aut$G$. Regarding A as an endomorphism of $C_0(G)$, we find that the *eigenspaces* of A are *invariant* under P. In particular, if an eigenvalue of A

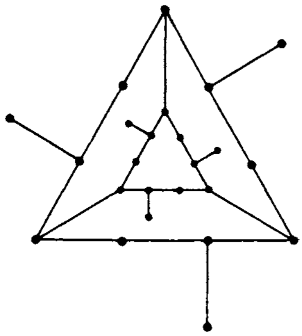


FIGURE VIII.10. A graph with automorphism group S_3, constructed from the Cayley diagram in Fig. VIII.1.

is simple (that is, it has multiplicity 1) then P must map an eigenvector to a multiple of itself. Thus if all eigenvalues are simple, we must have $P^2 = I$. This is a strong restriction on those permutations which might correspond to automorphisms of G. For example; if G has at least 3 vertices and every eigenvalue is simple, then Aut$G$ cannot be vertex-transitive, so not every pair of vertices can be interchanged by an automorphism.

These remarks indicate how the methods of representation theory may be used to deduce restrictions on the adjacency matrix of graphs which have extensive automorphism groups. Lack of space prevents us from exploring this further. Instead, we shall use algebraic methods to study graphs which are highly regular, although this regularity is not expressed in terms of the automorphism group.

Algebraic methods are particularly useful if we want to prove that certain regularity conditions cannot be satisfied except perhaps for a small set of parameters. A problem of this type arose in Chapter IV: for which values of k is there a k-regular graph of order $n = k^2 + 1$ and girth 5? We shall show later that if there is such a graph then k is 2, 3, 7 or 57. Group theory is particularly rich in problems of this type: at the end of this section we shall mention some examples.

Regularity, that is, the condition that all vertices have the same degree k, is not too restrictive, although the adjacency matrix of a connected regular graph does satisfy a very pleasant condition. Let $J = J_n$ be the n by n matrix with all n^2 entries 1 and, as before, let $\mathbf{j} = \mathbf{j}_n \in \mathbb{C}^n$ be the vector with all coordinates 1. Note that J has two eigenvalues: n, with multiplicity 1 and eigenvector $\mathbf{j}$, and 0, with multiplicity $n - 1$. Thus J/n is the orthogonal projection onto the 1-dimensional subspace $\langle \mathbf{j} \rangle$.

Theorem 14 *Let G be a connected k-regular graph of order n, with adjacency matrix A and distinct eigenvalues $k, \mu_1, \mu_2, \ldots, \mu_r$. Then*

$$\prod_{i=1}^{r} \frac{A - \mu_i I}{k - \mu_i} = \frac{J}{n}.$$

Proof. Each side is the orthogonal projection onto $\langle \mathbf{j} \rangle$. □

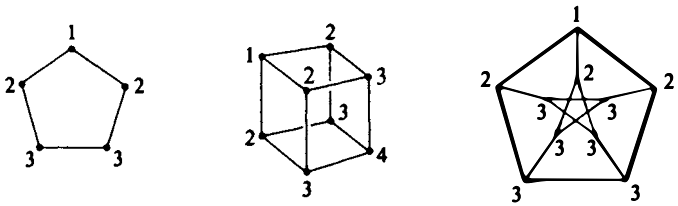
Theorem 14 readily implies that J is a polynomial of A iff G is connected and regular (see Exercise 46).

The above algebraic characterization of regular graphs is mildly interesting, but it does not come close to showing the power of algebraic methods in graph theory. This is not surprising since for a fixed k and large n there are simply too many k-regular graphs of order n (provided that kn is even), so that we cannot even contemplate a meaningful characterization of them. In order to give algebraic methods a chance to work their magic, we have to impose more restrictive regularity conditions on our graphs.

Call a connected graph G *highly regular with collapsed adjacency matrix* $C = (c_{ij})$ if for every vertex $x \in V = V(G)$ there is a partition of V into non-empty sets $V_1 = \{x\}, V_2, \dots, V_p$ such that each vertex $y \in V_j$ is adjacent to exactly c_{ij} vertices in V_i (see Fig. VIII.11). It is immediate from the definition that G is regular, say every vertex has degree k. In this case each *column sum* in the collapsed matrix is k. The collapsed adjacency matrix C can be obtained from the adjacency matrix A as follows:

$$c_{ij} = \sum_{v_s \in V_i} a_{st}, \quad \text{where } v_t \in V_j.$$

The point is exactly that the above sum is *independent* of the representative v_t of V_j. We are especially interested in the collapsed adjacency matrix C if it is of a much smaller size than A.



$$\begin{pmatrix} 0 & 1 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 2 & 0 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 2 & 2 \end{pmatrix}$$

FIGURE VIII.11. The pentagon, the cube and the Petersen graph together with their collapsed adjacency matrices.

At the risk of being too pedantic, in the arguments below we shall be particularly careful to identify the various spaces and maps. Let P be the p-dimensional complex vector space with basis $(w_1, \dots, w_p)$, and identify C with the linear map $P \to P$ with matrix C in this basis. For $v_r \in V = \{v_1, \dots, v_n\}$, let $V_1^{(r)} = \{v_r\}, V_2^{(r)}, \dots, V_p^{(r)}$ be a partition belonging to the vertex v_r. (We consider v_r the root of this partition.) Also, let $\pi_r : C_0(G) \to P$ be the linear map given by

$$\pi_r\left(\sum_{j=1}^{n} x_j v_j\right) = \sum_{i=1}^{p}\left(\sum_{v_j\in V_i^{(r)}} x_j\right) w_i.$$

Theorem 15 *Let G, A, C, P and π_r be as above.*
(i) $\pi_r A = C\pi_r$, that is, the diagram below commutes.

$$\begin{array}{ccc} C_0(G) & \xrightarrow{A} & C_0(G) \\ \pi_r\downarrow & & \downarrow\pi_r \\ P & \xrightarrow{C} & P \end{array}$$

(ii) The adjacency matrix A and the collapsed adjacency matrix C have the same minimal polynomial. In particular, μ is an eigenvalue of A iff it is a root of the characteristic polynomial of C.

Proof. (i) Let us show that $\pi_r(Av_t) = C(\pi_r v_t)$, where v_t is the basis vector corresponding to an arbitrary vertex $v_t \in V_j^{(r)}$. To do this it suffices to check that the ith coordinates of the two sides are equal. Clearly, $\pi_r v_t = w_j$ so

$$(C(\pi_r v_t))_i = (Cw_j)_i = c_{ij}$$

and

$$(\pi_r(Av_t))_i = \sum_{v_s\in V_i^{(r)}} a_{st}$$

and these are equal by definition.

(ii) Let q be the minimal polynomial of C. In order to prove that $q(A) = 0$, let $\mathbf{x} \in C_0(G)$ and set $q(A)\mathbf{x} = \sum_{i=1}^{n} y_i v_i$. Then for each r, $1 \le r \le n$, we have

$$y_r = (q(A)\mathbf{x})_r = (\pi_r(q(A)\mathbf{x}))_1 = (q(C)(\pi_r\mathbf{x}))_1 = (\mathbf{0})_1 = 0.$$

Conversely, the minimal polynomial of A annihilates C since $\pi_r C_0(G) = P$. □

This result enables us to restrict rather severely the matrices C that may arise as collapsed adjacency matrices.

Theorem 16 *Let G be a connected highly regular graph of order n with collapsed adjacency matrix C. Let $\mu_1, \mu_2, \ldots, \mu_r$ be the roots of the characteristic polynomial of C different from k, the degree of the vertices of G. Then there are natural numbers $m_1, m_2, \ldots, m_r$ such that*

$$\sum_{i=1}^{r} m_i = n - 1$$

and

$$\sum_{i=1}^{r} m_i \mu_i = -k.$$

Proof. We know from Theorem 5 that $\mu_1, \mu_2, \ldots, \mu_r$ are the eigenvalues of A in addition to k, which has multiplicity 1. Thus if $m(\mu_i)$ is the multiplicity of μ_1 then

$$1 + \sum_{i=1}^{r} m(\mu_i) = n,$$

since $C_0(G)$ has an orthonormal basis consisting of eigenvectors of A. Furthermore, since the trace of A is 0 and a change of basis does not alter the trace,

$$\operatorname{tr} A = k + \sum_{i=1}^{r} m(\mu_i)\mu_i = 0. \qquad \square$$

The condition expressed in Theorem 16 is not easily satisfied, especially if $\mu_1, \mu_2, \ldots, \mu_r$ are not rational numbers, so it rules out the possibility of constructing highly regular graphs with many seemingly feasible parameters. For the so-called strongly regular graphs we shall rewrite the condition in a more attractive form. A graph G is said to be *strongly regular with parameters* (k, a, b) if it is a k-regular incomplete graph such that any two adjacent vertices have exactly $a \geq 0$ common neighbours and any two non-adjacent vertices have $b \geq 1$ common neighbours. In other words, G is highly regular with collapsed adjacency matrix

$$C = \begin{pmatrix} 0 & 1 & 0 \\ k & a & b \\ 0 & k-a-1 & k-b \end{pmatrix}.$$

Putting it yet another way: if G is a connected incomplete graph with adjacency matrix A then G is strongly regular iff

$$A^2 \in \langle I, J, A \rangle,$$

where, as before, J is the matrix with every entry 1. More precisely, G has parameters (n, k, a, b) iff

$$A^2 = kI + \alpha A + b(J - I - A),$$

where A, I and J are n by n matrices. Indeed, the last equation is equivalent to

$$(A^2)_{ij} = \begin{cases} k & \text{if } i = j, \\ a & \text{if } v_i v_j \in E(G), \\ b & \text{otherwise.} \end{cases}$$

As G is neither complete nor empty, $b \geq 1$, so this is just the statement that G is strongly regular, with parameters (n, k, a, b). From Theorems 14 and 15 we can read off another simple characterization of strongly regular graphs.

Theorem 17 *Let G be a connected imcomplete regular graph. Then G is strongly regular iff it has precisely three distinct eigenvalues.*

Proof. Suppose G is a strongly regular graph with adjacency matrix A. As its collapsed adjacency matrix has order 3, by Theorem 15 it has at most three distinct eigenvalues. Furthermore, if G had only two distinct eigenvalues then, by Theorem 14 we would have $A \in \langle I, J\rangle$, which would imply that G is complete or empty.

Conversely, if A has three distinct eigenvalues then, again by Theorem 14, we have $A^2 \in \langle I, J, A\rangle$. □

Theorem 18 *If there is a strongly regular graph of order n with parameters (k, a, b) then*

$$m_1, m_2 = \frac{1}{2}\left\{n - 1 \pm \frac{(n-1)(b-a) - 2k}{\left\{(a-b)^2 + 4(k-b)\right\}^{1/2}}\right\}$$

are natural numbers.

Proof. The characteristic polynomial of the collapsed adjacency matrix C is

$$x^3 + (b - a - k)x^2 + ((a-b)k + b - k)x + k(k-b).$$

On dividing by $x - k$, we find that the roots different from k are

$$\mu_1, \mu_2 = \frac{1}{2}\left\{a - b \pm \left\{(a-b)^2 + 4(k-b)\right\}^{1/2}\right\}.$$

By Theorem 16 there are natural numbers m_1 and m_2 satisfying

$$m_1 + m_2 = n - 1$$

and

$$m_1\mu_1 + m_2\mu_2 = -k.$$

Solving these for m_1 and m_2 we arrive at the assertion of the theorem. □

Theorem 18 is sometimes called the *rationality condition* for strongly regular graphs. It is also easily proved without invoking Theorem 16. Indeed, if A is the adjacency matrix of a strongly regular graph with parameters (k, a, b) then, as we know,

$$A^2 = kI + aA + b(J - I - A).$$

Therefore J is a quadratic polynomial in A, so J and A are simultaneously diagonalizable. Noting that J has only two eigenvalues, namely n, with multiplicity 1, and 0, with multiplicity $n-1$, one can easily find that μ_1 and μ_2 are as above (cf. Exercise 30).

From the rationality condition it is but a short step to the beautiful result of Hoffman and Singleton, proved in 1960, concerning Moore graphs of diameter 2 (or girth 5).

Theorem 19 *Suppose there is a k-regular graph G of order $n = k^2 + 1$ and diameter 2. Then $k = 2, 3, 7$ or 57.*

Proof. We know from Theorem IV.1 that G is strongly regular with parameters $(k, 0, 1)$. By the rationality condition at least one of the following two conditions has to hold:

(i): $(n-1) - 2k = k^2 - 2k = 0$ and $n - 1 = k^2$ is even,

(ii): $1 + 4(k-1) = 4k - 3$ is a square, say $4k - 3 = s^2$.

Now, if (i) holds then $k = 2$.

If (ii) holds then $k = \frac{1}{4}(s^2 + 3)$; on substituting this into the expression for the multiplicity m_1 we find that

$$m_1 = \frac{1}{2}\left\{\frac{1}{16}(s^2+3)^2 + \frac{[(s^2+3)^2/16] - [(s^2+3)/2]}{s}\right\},$$

that is,

$$s^5 + s^4 + 6s^3 - 2s^2 + (9 - 32m_1)s - 15 = 0.$$

Hence s divides 15, so s is one of the values 1, 3, 5 and 15, giving $k = 1, 3, 7$ or 57. The case $k = 1$ is clearly unrealizable. □

It is worth noting that for $k = 2$, 3 and 7 there are unique k-regular graphs of order $k^2 + 1$ and diameter 2 (and so girth 5). In particular, for $k = 2$ it is a pentagon and for $k = 3$ it is the Petersen graph. However, it is not known whether or not $k = 57$ can be realized.

Sporadic simple groups are those simple groups that do not belong to one of the infinite sequences consisting of cyclic groups of prime order, alternating groups of degree at least 5 and simple groups of Lie type. Sporadic simple groups are often related to strongly regular graphs. For example, there is a strongly regular graph with parameters (162, 105, 81), and the McLaughlin group of order 898,128,000 is a subgroup of index 2 of the automorphism group of this graph. Similarly, there is a strongly regular graph with parameters (416, 100, 96) and the Suzuki group, which is a simple group of order 448,345,497,600, is a subgroup of index 2 of the automorphism group of this graph.

VIII.4 Enumeration and Pólya's Theorem

We cannot end this chapter without considering perhaps the most basic question about graphs, namely, how can we count graphs of various types? We may want to count graphs with a given set of vertices or we may be interested in the number of isomorphism classes of certain graphs. As we saw in Chapter VII, counting labelled graphs is relatively easy; for instance, there are $2^{\binom{n}{2}} = 2^N$ labelled graphs on n vertices, of which $\binom{N}{m}$ have m edges. Furthermore, by applying Corollary II. 13 to the complete graph, one can easily show that there are n^{n-2} labelled trees of order n. This result was first obtained by Cayley; we present it here with a proof due to Prüfer, which is independent of Corollary II. 13.

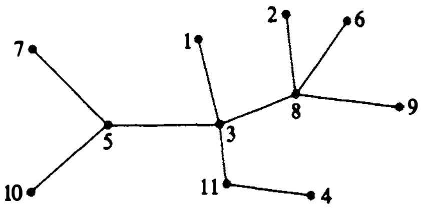


FIGURE VIII.12. The Prüfer code of this tree is (3, 8, 11, 8, 5, 8, 3, 5, 3).

Theorem 20 *There are n^{n-2} trees on n labelled vertices.*

Proof. As in Chapter VII, let $V = \{1, 2, \ldots, n\}$ be the set of vertices. Given a tree T, associate a code with T as follows. Remove the endvertex with the smallest label and write down the label of the adjacent vertex. Repeat the process until only two vertices remain. The code obtained is a sequence of length $n-2$ consisting of some numbers from $1, 2, \ldots, n$; of course, any number may occur several times in the code (see Fig. VIII.12). As the reader should check, each of the n^{n-2} possible codes corresponds to a unique tree. □

It is easily seen that the label of a vertex of degree d occurs exactly $d-1$ times in the Prüfer code of the tree. Thus the proof has the following consequence.

Corollary 21 *Let $d_1 \leq d_2 \leq \cdots \leq d_n$ be the degree sequence of a tree: $d_1 \geq 1$ and $\sum_{i=1}^{n} d_i = 2n-2$. Then the number of labelled trees of order n with degree sequence $(d_i)_1^n$ is given by the multinomial coefficient*

$$\binom{n-2}{d_1-1, d_2-1, \ldots, d_n-1}. \qquad \square$$

The difficulties we encounter change entirely if we wish to count certain classes of graphs up to isomorphism. Given graphs G_1 and G_2 with a common vertex set V, when are they isomorphic? They are isomorphic if there is a permutation π of V which maps G_1 onto G_2. Of course, strictly speaking π does not act on graphs, it only *induces* a permutation α of $X = V^{(2)}$, the *pairs* of vertices, and it is α that maps an *edge* of G_1 into an *edge* of G_2 and a *non-edge* of G_1 into a *non-edge* of G_2. Now, G_i is naturally identified with a subset of X or, equivalently, with a function $f_i : X \to \{0, 1\}$. Therefore G_1 is isomorphic to G_2 iff there is a permutation α of $X = V^{(2)}$ (coming from a permutation π of V) such that $\alpha^* f_1 = f_2$, where α^* is the permutation of the set of functions $\{0, 1\}^X$ induced by α. Thus counting graphs up to isomorphism is a special case of the following problem. Given sets X and Y, and a group Γ acting on X, let Γ act on the set of functions Y^X in the natural way. How many orbits are there in Y^X? The main aim of this section is to present a beautiful theorem of Pólya, proved in 1937, which answers this question.

Let Γ be a group of permutations acting on a (finite) set X. For $x, y \in X$ put $x \sim y$ if $y = \alpha x$ for some $\alpha \in \Gamma$. Then $\sim$ is an equivalence relation on X; if $x \sim y$ we say that x is *equivalent to* y *under* Γ. The equivalence class of x is called the Γ-*orbit* (or simply *orbit*) of x, and is denoted by $[x]$. For $x, y \in X$ put

$$\Gamma(x, y) = \{\alpha \in \Gamma : \alpha x = y\}.$$

Of course, $\Gamma(x, y)$ is non-empty iff $[x] = [y]$, that is, x and y belong to the same orbit. The set $\Gamma(x) = \Gamma(x, x)$ is the *stabilizer* of x; it is a subgroup of Γ. Note that if $y = \beta x$ then

$$\Gamma(x, y) = \{\alpha : \alpha x = y\} = \{\alpha : \alpha x = \beta x\} = \{\alpha : \beta^{-1}\alpha \in \Gamma(x)\} = \beta\Gamma(x),$$

so $\Gamma(x, y)$ is a coset of $\Gamma(x)$. We see that $|\Gamma(x)|$ depends on the equivalence class of x, so we may put $s([x]) = |\Gamma(x)|$. Clearly

$$\Gamma = \bigcup_{y \in [x]} \Gamma(x, y),$$

and this gives us

$$|\Gamma| = |[x]||\Gamma(x)| = |[x]|s([x]).$$

Pólya's enumeration theorem is based on a version of a lemma due to Cauchy and Frobenius, extensively used by Burnside, concerning the sum of the "weights" of orbits. (For many years, it was called Burnside's lemma.) Let $O_1, \dots, O_\ell$ be the Γ-orbits, let A be an arbitrary Abelian group (written additively) and let $w : X \to A$ be a function that is constant on orbits. We call w a *weight function* and define the weight of O_i by $w(O_i) = w(x)$, $x \in O_i$. For a permutation $\alpha \in \Gamma$ we denote by $F(\alpha)$ the set of elements fixed by α, that is, $F(\alpha) = \{x \in X : \alpha x = x\}$. Thus $x \in F(\alpha)$ iff $\alpha \in \Gamma(x)$. After this preparation, here is then the Cauchy–Frobenius lemma.

Lemma 22 $|\Gamma| \sum_{i=1}^{\ell} w(O_i) = \sum_{\alpha \in \Gamma} \sum_{x \in F(\alpha)} w(x)$.

Proof.

$$\begin{aligned}\sum_{\alpha \in \Gamma} \sum_{x \in F(\alpha)} w(x) &= \sum_{x \in X} \sum_{\alpha \in \Gamma(x)} w(x) = \sum_{i=1}^{\ell} \sum_{x \in O_i} \sum_{\alpha \in \Gamma(x)} w(x) \\ &= \sum_{i=1}^{\ell} w(O_i) \sum_{x \in O_i} \sum_{\alpha \in \Gamma(x)} 1 = \sum_{i=1}^{\ell} w(O_i)|O_i|s(O_i) \\ &= |\Gamma| \sum_{i=1}^{\ell} w(O_i).\end{aligned}$$

□

The original form of this lemma is obtained on choosing $A = \mathbb{Z}$ and $w \equiv 1$:

$$N(\Gamma) = \frac{1}{|\Gamma|} \sum_{\alpha \in \Gamma} \sum_{x \in F(\alpha)} 1 = \frac{1}{|\Gamma|} \sum_{\alpha \in \Gamma} |F(\alpha)|,$$

where $N(\Gamma)$ is the *number of orbits*.

We shall illustrate by three very simple examples that even the Cauchy–Frobenius lemma can be used to calculate the number of equivalence classes of certain objects.

EXAMPLE 1. Let $X = \{1, 2, 3, 4\}$ and $\Gamma = \{1, (12), (34), (12)(34)\}$. What is $N(\Gamma)$? Clearly, $F(1) = \{1, 2, 3, 4\}$, $F((12)) = \{3, 4\}$, $F((34)) = \{1, 2\}$ and $F((12)(34)) = \emptyset$. Thus $N(\Gamma) = \frac{1}{4}\{4 + 2 + 2 + 0\} = 2$.

EXAMPLE 2. Consider all bracelets made up of 5 beads. The beads can be red, blue and green, and two bracelets are considered to be identical if one can be obtained from the other by rotation. (Reflections are not allowed!) How many distinct bracelets are there?

In this case we choose X to be the set of all $3^5 = 243$ bracelets and let Γ be C_5, the cyclic group of order 5, acting on X. Then the question is: how many orbits does Γ have? For the identity $1 \in \Gamma$ clearly, $F(1) = X$. For every non-trivial rotation $\alpha \in \Gamma$ only the 3 monochromatic bracelets are invariant under α, so $N(\Gamma) = \frac{1}{5}\{243 + 3 + 3 + 3 + 3\} = 51$.

EXAMPLE 3. In how many essentially different ways can we colour the six faces of a cube with at most three colours, say red, white and green? Here two colourings are essentially the same if some rotation can take one into the other.

In this example Γ is the group of rotations of the cube, so $|\Gamma| = 24$. Let us catalogue the rotations and the numbers of colourings fixed by them. The identity rotation fixes all 3^6 colourings. There are 8 rotations through pairs of opposite vertices, each fixing 3^2 colourings. Each of the 6 rotations through mid-points of opposite edges fixes 3^3 colourings. There are 9 further rotations, through centres of opposite faces. Those through angle π fix 3^4 colourings each, and those through angle $\pi/2$ fix 3^3 colourings each. Hence there are

$$\frac{1}{24}\{3^6 + 8.3^2 + 6.3^3 + 3.3^4 + 6.3^3\} = 57$$

essentially different colourings of the faces of a cube with red, white and green.

The second and third examples resemble a little the problem we really want to tackle. Let Γ be a group of permutations of a (finite) set D. Let R be another (finite) set and let us consider the set R^D of all functions from D into R. Each $\alpha \in \Gamma$ can be made to act on R^D; namely define $\alpha^* : R^D \to R^D$ by

$$(\alpha^* f)(d) = f(\alpha d), \quad f \in R^D, \; d \in D.$$

Then

$$\Gamma^* = \{\alpha^* : \alpha \in \Gamma\}$$

is a group of permutations of R^D; as an abstract group, it is isomorphic to Γ and we distinguish it from Γ only to emphasize that it acts on R^D while Γ acts on D.

As is customary in connection with Pólya's enumeration theorem, we adopt an intuitive terminology. The set D is called the *domain* and its elements are *places*; R is the *range* and its elements are *figures*; the functions in R^D are called *configurations*; finally a *pattern* is an equivalence class of configurations under Γ^*, that is, a Γ^*-orbit. Our main aim is to calculate the number of distinct patterns.

The origin of this terminology is that a function $f \in R^D$ is an arrangement of some figures into the places in such a way that for each place there is exactly one figure in that place, but each figure can be put into as many places as we like. Two configurations mapped into each other by an element of Γ^* have the same pattern and are not distinguished. Thus in the second example the places are, say, 1, 2, 3, 4 and 5, the figures are r, b and g (for red, blue and green) and a configuration is a sequence of the type g, b, b, r, b, that is, a bracelet. The group Γ is generated by (12345) and distinct patterns correspond to distinguishable bracelets.

In addition to counting the number of distinct patterns, we may wish to count the number of patterns of a certain type. It turns out that all these problems can be solved at once, provided we learn enough about the cycle structure of permutations in Γ acting on D, and are willing to store a large amount of information about the patterns.

Each element $\alpha \in \Gamma$ is an essentially unique product of disjoint cycles (cyclic permutations) acting on D. If $\alpha = \xi_1\xi_2\cdots\xi_m$ is such a product, we say that $\xi_1, \dots, \xi_m$ are the *cycles* of α. In the product we include cycles of length 1 as well so that every $\alpha \in D$ appears in exactly one cycle; if $|\xi|$ denotes the number of elements in ξ then $\sum_{k=1}^{m} |\xi_k| = d$, where $d = |D|$ is the number of elements in D. Denote by $j_k(\alpha)$ the number of cycles of α having length k; by the previous equality $\sum_{k=1}^{m} kj_k(\alpha) = d$. Note that $|F(\alpha)|$, appearing in the Cauchy–Frobenius lemma, is exactly $j_1(\alpha)$, the number of elements of D fixed by α. We define the *cycle sum* of Γ to be

$$\widetilde{Z}(\Gamma; a_1, \dots, a_d) = \sum_{\alpha\in\Gamma} \prod_{k=1}^{d} a_k^{j_k(\alpha)}.$$

The reader should bear in mind that $\widetilde{Z}$ depends on the *action of* Γ *on* D, not only on the abstract group Γ. Note also that the cycle sum is a *polynomial* in $a_1, a_2, \dots, a_d$ with integer coefficients; it tells us the distribution of cycles in the elements of Γ. When writing down a cycle sum, it is useful to remember that $\sum_{k=1}^{d} kj_k(\alpha) = d$ for every α. The customary *cycle index* of Γ is $Z(\Gamma; a_1, \dots, a_d) = (1/|\Gamma|)\widetilde{Z}(\Gamma; a_1, \dots, a_d)$. As we shall consider general rings instead of the more usual polynomial ring with rational coefficients, we have to use the cycle sum since we cannot divide by $|\Gamma|$.

Let A be an arbitrary *commutative ring* and let $w : R \to A$ be a function. We call $w(r)$ the *weight* of the figure f, and for $k = 1, 2, \dots$ define the k^{th} *figure sum* as

$$s_k = \sum_{r\in R} w(r)^k.$$

Furthermore, the *weight* of a configuration $f \in R^D$ is

$$w(f) = \prod_{a\in D} w(f(a)).$$

Clearly, any two configurations equivalent under Γ^* have the same weight, so we may define the weight of a pattern O_i by

$$w(O_i) = w(f), \qquad f \in O_i.$$

Our aim is to learn about the *pattern sum*

$$S = \sum_{i=1}^{\ell} w(O_i),$$

where $O_1, O_2, \ldots, O_\ell$ are the Γ^*-orbits, that is, the distinct patterns.

Note that $w(r)$, $w(f)$, s_k and S are all elements of our commutative ring A. If we have a way of determining the pattern sum S, it is up to us to choose A and the weight function $w : R \to A$ in such a way that S can be "decoded" to tell us all we want to know about various sets of patterns. In practice one always chooses A to be a polynomial ring ($\mathbb{Z}[x]$, $\mathbb{Q}[x, y]$, etc.), and usually $w(r)$ is a monic polynomial; the information we look for is then given by certain coefficients of the polynomial S. We shall give several examples after the proof of our main result, *Pólya's enumeration theorem.*

Theorem 23 *With the notation above,*

$$|\Gamma| S = \widetilde{Z}(\Gamma; s_1, s_2, \ldots, s_d).$$

Proof. By Lemma 22,

$$|\Gamma| S = |\Gamma| \sum_{i=1}^{\ell} w(O_i) = \sum_{\alpha \in \Gamma} \sum_{f \in F(\alpha^*)} w(f).$$

Now, clearly $F(\alpha^*) = \{f \in R^D : f$ is constant on cycles of $\alpha\}$, so if $\xi_1, \xi_2, \ldots, \xi_m$ are the cycles of α, and $a \in \xi_i$ means that a is an element of the cycle ξ_i, then

$$F(\alpha^*) = \{f \in R^D : r_i \in R \text{ and } f(a) = r_i \text{ if } a \in \xi_i,\ i = 1, 2, \ldots, m\}.$$

Hence

$$\sum_{f \in F(\alpha^*)} w(f) = \sum_{(r_i) \subset R} \prod_{i=1}^{m} w(r_i)^{|\xi_i|} = \prod_{k=1}^{d} \left(\sum_{r \in R} w(r)^k \right)^{j_k(\alpha)} = \prod_{k=1}^{d} s_k^{j_k(\alpha)},$$

giving

$$|\Gamma| S = \sum_{\alpha \in \Gamma} \prod_{k=1}^{d} s_k^{j_k(\alpha)} = \widetilde{Z}(\Gamma; s_1, s_2, \ldots, s_d). \qquad \square$$

If $|\Gamma|$ has an inverse in the ring A, say if A is a polynomial ring over the rationals, then Theorem 23 can also be written in its more usual form:

$$S = Z(\Gamma; s_1, s_2, \ldots, s_d).$$

Let us illustrate now how the theorem can be applied.

EXAMPLE 4. Let us consider again the bracelets made up of five beads, which can be red, blue and green. Then $D = \{1, 2, 3, 4, 5\}$ is the set of places of the beads, $R = \{r, b, g\}$ is the set of colours (figures) and Γ is C_5, the cyclic group of order 5 generated by the permutations (12345). The cycle sum is $\widetilde{Z} = a_1^5 + 4a_5$.

On choosing $A = \mathbb{Z}$ and $w(r) = w(b) = w(g) = 1$, we find that $s_k = 3$ for every k, so $5S = 3^5 + 4.3$. Since each pattern (bracelet) has weight 1, there are $\frac{1}{5}\{3^5 + 12\} = 51$ distinct patterns (bracelets).

On choosing $A = \mathbb{Z}[x, y]$ and $w(r) = 1$, $w(b) = x$, $w(g) = y$, we find that $S = \frac{1}{5}\{(1+x+y)^5 + 4(1+x^5+y^5)\}$. Now it is easy to extract information from this form of S. For example, a bracelet has weight xy^2 iff it has 2 red, 1 blue and 2 green beads. Thus the number of such bracelets is the coefficient of xy^2 in the polynomial S; that is, $(1/5)(5!/(2!2!)) = 6$.

EXAMPLE 5. What happens if in the previous example we allow reflections? Then Γ is the dihedral group D_5, the group of symmetries of the regular pentagon, whose cycle sum is $a_1^5 + 4a_5 + 5a_1a_2^2$. Thus if we take, as before, $A = \mathbb{Z}[x, y]$, $w(r) = 1$, $w(b) = x$ and $w(g) = y$, we find that the number of bracelets containing 2 red, 1 blue and 2 green beads is the coefficient of xy^2 in $\frac{1}{10}\left\{(1+x+y)^5 + 4(1+x^5+y^5) + 5(1+x+y)(1+x^2+y^2)^2\right\}$, that is, $3 + 1 = 4$.

EXAMPLE 6. This is the example Pólya used to illustrate his theorem. Place 3 red, 2 blue and 1 yellow ball in the 6 vertices of an octahedron. In how many distinct ways can this be done? The group of rotations of the octahedron has order 24 and cycle sum $a_1^6 + 6a_1^2a_4 + 3a_1^2a_2^2 + 6a_2^3 + 8a_3^2$: a_1^6 comes from the identity, $6a_1^2a_4$ from the rotations through $\pi/2$ about axes through opposite vertices, $3a_1^2a_2^2$ from the rotations through π about axes through opposite vertices, $6a_2^3$ from the rotations through π about axes through midpoints of edges, and, finally, a_3^2 is the summand corresponding to a rotation through $2\pi/3$ about an axis going through the centre of a face. On taking $A = \mathbb{Z}[x, y]$, $w(r) = 1$, $w(b) = x$ and $w(y) = y$, we see that the required number is the coefficient of x^2y in

$$\frac{1}{24}\{(1+x+y)^6 + 6(1+x+y)^2(1+x^4+y^4)$$
$$+ 3(1+x+y)^2(1+x^2+y^2)^2 + 6(1+x^2+y^2)^3 + 8(1+x^3+y^3)^2\},$$

that is, 3.

It should be clear by now that the theorem loses nothing from its generality if instead of a general commutative ring A we take $\mathbb{Z}[x_r : r \in R]$, the polynomial ring over the integers in variables indexed by the elements of R, and we define the weight function as $w(r) = x_r$. Then the pattern sum S contains all the information the theorem can ever give us. In particular, if $w : R \to A$ is an arbitrary weight function then the corresponding pattern sum is obtained by replacing x_r by $w(r)$ in S. However, if R is large, the calculations my get out of hand if we do not choose a "smaller" ring than $\mathbb{Z}[x_r : r \in R]$, which is tailor-made for the problem at hand. The choice of a smaller ring is, of course, equivalent to a substitution into S.

EXAMPLE 7. Place red, blue, green and yellow balls into the vertices of an octahedron. Denote by P_i the set of patterns in which the total number of red and blue balls is congruent to i modulo 4. What is $|P_0| - |P_2|$?

The cycle sum of the rotation group of the octahedron was calculated in Example 6 and was found to be $a_1^6 + 6a_1^2a_4 + 3a_1^2a_2^2 + 6a_2^3 + 8a_3^2$.

Let $A = \mathbb{C}$, the field of complex numbers, and put $w(r) = w(b) = i$, $w(g) = w(y) = 1$. Then for a pattern f we have $\operatorname{Re} w(f) = 1$ if $f \in P_0$, $\operatorname{Re} w(f) = -1$ if $f \in P_2$ and $\operatorname{Re} w(f) = 0$ if $f \in P_1 \cup P_3$. Thus $|P_0| - |P_2|$ is exactly the real part of the pattern sum. As $s_1 = 2(1+i)$, $s_2 = 0$, $s_3 = 2(1-i)$ and $s_4 = 4$, we see immediately that after substitution the real part of each term is 0, so $|P_0| = |P_2|$.

We were first led to our study of the orbits of a permutation group by our desire to count the number of graphs up to isomorphism. We realized that this amounted to counting the orbits of the group Γ_n^* acting on $\{0, 1\}^X$, where $X = V^{(2)}$ and Γ_n is the permutation group acting on X that is induced by the symmetric group acting on V. So according to Pólya's theorem our problem is solved when we know the cycle sum of the permutation group Γ_n. It is now a routine matter to write down an explicit expression for this cycle sum, though we don't display it here since its form is not very inspiring. Furthermore, except for small values of n, this expression is too unwieldy for practical calculations, and it is much easier to use asymptotic formulae derived by random graph techniques (see Exercises 22 and 23 of Chapter VII).

We remark finally that an extension of Pólya's theorem covers the case when there is also a group acting on the range of the functions. For instance, if we let S_2 act on $\{0, 1\}$ in the example above, we do not distinguish between a graph and its complement, and may thereby compute the number of graphs that are isomorphic to their complements.

VIII.5 Exercises

1.$^-$ Draw the Cayley diagram of the quaternion group $\langle a, b \mid a^2 = b^2 = (ab)^2 \rangle$.

2. Find the orders of $\langle a, b \mid a^3 = b^3 = 1, ab = ba \rangle$, $\langle a, b \mid a^4 = b^3 = 1, ab = ba \rangle$ and $\langle a, b \mid a^3 = b^4 = (ab)^2 = 1 \rangle$.

3. Use Euler's formula and information about the Cayley diagram to deduce that in the group $\langle a, b, c \mid a^5 = b^3 = c^2 = (abc)^{-1} \rangle$ we have $a^{610} = 1$.

4. Verify that the Cayley diagram associated with the trefoil knot is the diagram described in Fig. VIII.5.

5. Write down presentations of the knots shown in Fig. VIII.13. Show that the group of the quinquefoil is isomorphic to $\langle f, g \mid f^5 = g^2 \rangle$ and the group of the tweeny is $\langle a, b \mid ababa^{-1}b^{-1}a^{-1}babab^{-1}a^{-1}b^{-1} \rangle$.

6. Prove that no two of the groups of the knots shown in Figs VIII.4 and VIII.13 are isomorphic.

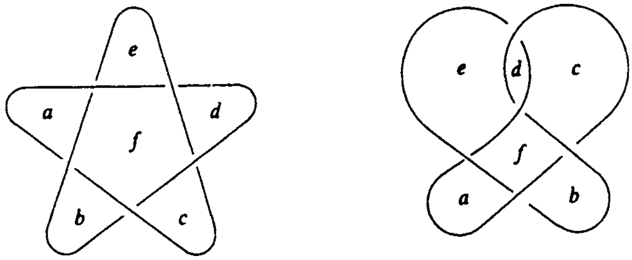


FIGURE VIII.13. The quinquefoil and the tweeny: the two knots whose minimal diagrams have five crossings.

7. A closed orientable surface of genus 2 is obtained by identifying pairs of non-adjacent sides of an octagon, say as in Fig. VIII.14. The fundamental group has a presentation

$$\langle a_1, a_2, a_3, a_4 \mid a_1 a_2^{-1} a_4 a_1^{-1} a_3 a_2 a_4^{-1} a_3^{-1} \rangle.$$

Show that the Cayley diagram is a tessellation of the hyperbolic plane. Deduce that any non-empty reduced word W equal to 1 must contain a subword of length at least 5 that is part of the cyclically written relator or its inverse. (A reduced word is one in which no generator occurs next to its inverse.) [*Hint.* Consider the part of the walk W furthest from 1.]

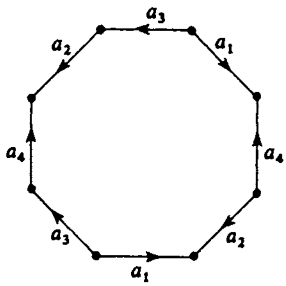


FIGURE VIII.14. An orientable surface of genus 2.

8. Show that the *dihedral group* D_n, the group of symmetries of a regular n-gon, has a presentation of the form $\langle a, b \mid a^n = b^2 = (ab)^2 = 1\rangle$. What is its Cayley diagram?

9. Give a group whose Cayley diagram is the truncated cube having 8 triangular and 6 octagonal faces.

10. Draw the Cayley diagram of
    (i) $\langle a, b \mid ab = ba\rangle$ in the Euclidean plane,
    (ii) $\langle a, b \mid b^n, ab = ba\rangle$ on an infinite cylinder,
    (iii) $\langle a, b \mid a^m = b^n = 1, ab = ba\rangle$ on a torus,
    (iv) $\langle a, b, c \mid a^4 = b^4 = c^4 = abc = 1\rangle$ in the hyperbolic plane.

11. Check the examples of Cayley diagrams illustrated in Fig. VIII.3 and prove Theorem 1.

12. Let $A = \langle a, b \mid a^3 = (ba)^2 = 1 \rangle$ and $B = \langle b \rangle$. What is the Schreier diagram of $A \bmod B$?

13. A group A is generated by a, b and c; the Schreier diagram of A modulo a subgroup B is shown in Fig. VIII.15. Read off a set of generators of B.

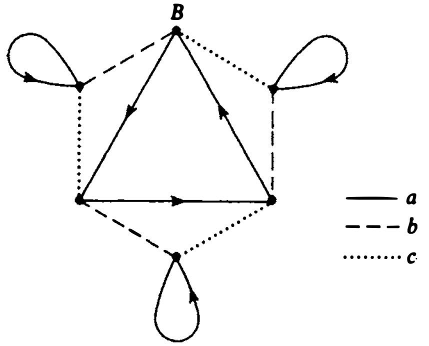


FIGURE VIII.15. The Schreier diagram of $A \bmod B$.

14. Let A be the free group on a, b and c, and let B be the subgroup consisting of all squares. What is the Schreier diagram of $A \bmod B$? Find a set of free generators for B.

15. How many subgroups of index 2 are there in a free group of rank k?

16. How many subgroups of index n are there in a free group on 2 generators?

17.$^+$ Show that a subgroup B of a finitely generated free group F is of finite index in F iff B is finitely generated and there is a natural number n for which $W^n \in B$ for every word W.

18. What is the automorphism group of the Petersen graph (shown in Fig. VIII.11)? Find the automorphism group of the Kneser graph $K_s^{(r)}$, where $s \geq 2r + 1$ (see Exercise V.10). Deduce that the automorphism group of $K_{2r+1}^{(r)}$ is *3-arc-transitive*, that is, any path of length 3 can be mapped into any other path of length 3 by an automorphism.

19. The *Tutte* 8-*cage* has vertices n_1, n_3, n_5 for $n = 1, 2, \ldots, 10$, with edges joining n_5 to n_1 and n_3, and n_i to m_i iff $|n - m| \equiv \pm i \pmod{10}$. Show that the automorphism group of the Tutte 8-cage is 5-arc-transitive.

20. Find the automorphism group of the Grötzsch graph (see Fig. V.11).

21. A vertex $x \in V(G)$ is a *centre* of a connected graph G if $\max\{d(x, y) : y \in V(G)\} = \min_u \max\{d(u, v) : v \in V(G)\}$. Show that every tree has either one or two centres, and in the latter case, they are adjacent.

22. Show that every automorphism of a tree either fixes some vertex or swaps some two adjacent vertices.

23. Does every tree have an Abelian automorphism group?

24. Construct a non-trivial tree with trivial automorphism group.

25.$^+$ Let $f(n)$ be the minimal size of a graph of order n with trivial automorphism group. Prove that $f(n) < n$ for every $n \geq 7$ and $f(n)/n \to 1$ as $n \to \infty$.

26. Show that if $\pi \in \operatorname{Aut} G$ has k odd cycles and l even cycles then G has at most $k + 2\ell$ simple eigenvalues.

27. Show that a connected graph G of odd order whose automorphism group is vertex transitive has exactly one simple eigenvalue.

28.$^-$ Let A be the adjacency matrix of a graph G. Show that the ij entry of A^ℓ is the number of v_i–v_j walks of length ℓ.

29. Given $k \geq 2$, let $p_0(x), p_1(x), \ldots$ be polynomials defined by $p_0(x) = 1$, $p_1(x) = x$, $p_2(x) = x^2 - k$ and

$$p_\ell(x) = xp_{\ell-1}(x) - (k-1)p_{\ell-2}(x), \qquad \ell \geq 3.$$

Show that if A is the adjacency matrix of a k-regular graph then $(p_\ell(A))_{ij}$ is the number of $v_i - v_j$ walks of length ℓ in which any two consecutive edges are distinct.

30. Complete the details of the second proof of Theorem 18, as suggested there.

31. Check that for $n \geq 2$ the adjacency matrix of the complete graph K_n has two distinct eigenvalues: $n-1$ and -1, with $m(n-1) = 1$ and $m(-1) = n-1$. Also, the n by n matrix J_n, with all entries 1, has eigenvalues n and 0, with $m(n) = 1$ and $m(0) = n - 1$.

32. Check that, for $n_1, n_2 \geq 1, n_1 + n_2 \geq 3$, the complete bipartite graph K_{n_1,n_2} has three distinct eigenvalues: $\sqrt{n_1 n_2}$, $-\sqrt{n_1 n_2}$ and 0, with $m(\sqrt{n_1 n_2}) = m(-\sqrt{n_1 n_2}) = 1$ and $m(0) = n - 2$.

33. Check that for $r \geq 3$ and $n \geq 1$ the complete r-partite graph $K_r(n)$ has three distinct eigenvalues: $(r-1)n$, $-n$ and 0, with $m((r-1)n) = 1, m(-n) = r-1$ and $m(0) = r(n-1)$.

34. Let G be a regular graph of order n, and let G^* be obtained from G by substituting r independent vertices for each vertex of G, as in Theorem 19. (Thus G^* has rn vertices, and each edge of G corresponds to a $K_{r,r}$.) Show that the eigenvalues of the adjacency matrix of G^* are precisely the eigenvalues of G, together with 0, which has (additional) multiplicity $(r-1)n$.

35.$^+$ Show that the eigenvalues of C_n are 2, $2\cos 2\pi/n$, $2\cos 4\pi/n, \ldots,$ $2\cos 2(n-1)\pi/n$. Thus if n is odd, 2 has multiplicity one and each other eigenvalue has multiplicity two; if n is even, each of 2 and -2 has multiplicity one, and every other eigenvalue has multiplicity two. [*Hint.* Assuming, as we may, that $V(C_n) = \mathbb{Z}_n$ and $E(C_n) = \{ij : i - j = \pm 1\}$, note that $(1, \omega, \omega^2, \dots, \omega^{n-1})$ is an eigenvector for each nth root of unity ω.]

36.$^+$ Let G be the rth power, C_n^r, of an n-cycle: $V(G) = \mathbb{Z}_n^r$, say, and $\mathbf{xy}$ is an edge of G if for some $1 \le j \le r$, we have $x_i = y_i$ for all $i \ne j$, and $x_j - y_j = \pm 1$. Show that the eigenvalues of G are $2\sum_{j=1}^r \cos \ell_j 2\pi/n$, $0 \le \ell_j < n, j = 1, \dots, r$, with appropriate multiplicities. [*Hint.* If $\omega_1, \dots, \omega_r$ are nth roots of unity, then $f : V(G) \to \mathbb{C}, (k_1, \dots, k_r) \mapsto \omega_1^{k_1} \cdots \omega_r^{k_r}$, is an eigenvector.]

37. Let A be the adjacency matrix of a regular graph G of order n, and let $\overline{A}$ be the adjacency matrix of $\overline{G}$. Note that $A + \overline{A} = J_n - I_n$. Deduce that if $\mu_1 \ge \cdots \ge \mu_n$ are the eigenvalues of A, enumerated with multiplicities, then $n - 1 - \mu_1, -1 - \mu_2, \dots, -1 - \mu_n$ are the eigenvalues of $\overline{A}$.

38. For a graph G with vertex set $\{v_1, v_2, \dots, v_n\}$, set

$$g_G(\mathbf{x}) = \sum_{i=1}^n x_i^2 + \sum_{v_i \sim v_j} x_i x_j = \sum_{i=1}^n x_i^2 + 2 \sum_{v_i v_j \in E(G)} x_i x_j.$$

Prove that $\min_{\mathbf{x} \in S} g_G(\mathbf{x}) = 1/\ell$, where S is the simplex

$$\left\{\mathbf{x} \in \mathbb{R}^n : x_i \ge 0, \sum x_i = 1\right\}$$

and $\ell = \beta(G)$.

39. Show that if the Lagrangian $f_G(x)$ attains its maximum in the interior of S then G is a complete k-partite graph, where $k = \omega(G)$.

40. Let Q^n be the graph of the n-dimensional cube. Thus Q^n has the vertex set $\{0, 1\}^n$, and two sequences $(a_i)_1^n, (b_i)_1^n \in \{0, 1\}^n$ are adjacent if they differ in precisely one term. For $1 \le d \le n$, let $Q^{n,d}$ be obtained from Q^n by joining vertices at distance d. Let A_n be the adjacency matrix of Q^n and $B_{n,d}$ the adjacency matrix of $Q^{n,d}$. Show that A_n and $B_{n,d}$ commute and $B_{n,d}$ is a polynomial of A_n.

41. (Exercise 40 contd.) Prove that the eigenvalues of A_n are $n - 2k$, $k = 0, 1, \dots, n$, with $n - 2k$ having multiplicity $\binom{n}{k}$. [*Hint.* Note that

$$A_n = \begin{pmatrix} A_{n-1} & I \\ I & A_{n-1} \end{pmatrix}.$$

Also for $\varepsilon = (\varepsilon_i)_1^n$, $\varepsilon_i = \pm 1$, define $\mathbf{v}_\varepsilon = (v_h)_0^{2^n - 1}$ as follows: if $h = \sum_{i=0}^{n-1} a_i 2^i$, with $a_i = 0, 1$, then $v_h = \prod \varepsilon_i^{a_i - 1}$. Check that if k of the ε_i are -1 then $\mathbf{v}_\varepsilon$ is an eigenvector of A_n with eigenvalue $n - 2k$.]

42.+ Let $G_1, \dots, G_\ell$ be (not necessarily distinct) complete subgraphs of K_n, each of order at most $n-1$, such that every edge of K_n belongs to the same number $\mu \geq 1$ of G_is. Prove *Fischer's inequality* that $\ell \geq n$, provided $n \geq 2$. [*Hint.* Set $V(K_n) = \{v_1, \dots, v_n\}$, and let B be the n by ℓ incidence matrix of the cover $K_n = \cup_{i=1}^{\ell} G_i$. Thus $(B)_{ij}$ is 1 if $v_i \in G_j$ and 0 otherwise. By considering its quadratic form, show that BB^t has rank n.]

43. Let G be a connected graph with $V(G) = \{v_1, \dots, v_n\}$, $n \geq 2$, adjacency matrix A and degree sequence $d(v_1), \dots, d(v_n)$. Let $D = \mathrm{Diag}(d(v_1), \dots, d(v_n))$. Let $L = D - A$ be the combinatorial Laplacian, and define the *analytic Laplacian* $\mathcal{L}$ of G as $\mathcal{L} = D^{-1/2}LD^{-1/2}$. For $\lambda \in \mathbb{C}$, let $E_{D^{-1}L}(\lambda) = (D^{-1}L - \lambda I)^{-1}(0)$ and $E_{\mathcal{L}}(\lambda) = (\mathcal{L} - \lambda I)^{-1}(0)$, so that $m_{D^{-1}L}(\lambda) = \dim E_{D^{-1}L}(\lambda)$ and $m_{\mathcal{L}}(\lambda) = \dim E_{\mathcal{L}}(\lambda)$ are the (geometric) multiplicities of an eigenvalue λ. Show that all eigenvalues of $D^{-1}L$ and $\mathcal{L}$ are real.
Show also that $D^{1/2}$ maps $E_{D^{-1}L}(\lambda)$ onto $E_{\mathcal{L}}(\lambda)$ for every λ and so $m_{D^{-1}L}(\lambda) = m_{\mathcal{L}}(\lambda)$.

44. (Exercise 43 contd.) Show that every eigenvalue of the analytic Laplacian $\mathcal{L}$ is nonnegative and $m_{\mathcal{L}}(0) = 1$. What are the eigenvectors of $\mathcal{L}$ with eigenvalue 0?

45. (Exercise 44 contd.) Prove that the following four properties are equivalent:
(a) G is bipartite,
(b) $m_{\mathcal{L}}(\lambda) = m_{\mathcal{L}}(2-\lambda)$ for every λ,
(c) $m_{\mathcal{L}}(2) = 1$,
(d) $m_{\mathcal{L}}(2) \geq 1$.

46. Prove that the matrix J (all of whose entries are 1) is a polynomial in the adjacency matrix A if G is regular and connected.

47. Let T be a hermitian operator on a complex inner product space V, with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, and associated quadratic form $q(\mathbf{x}) = \langle T\mathbf{x}, \mathbf{x}\rangle$. Let S be the unit sphere of V:

$$S = \{\mathbf{x} \in V : \|\mathbf{x}\| = 1\}.$$

Define vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in S$ as follows. Let $\mathbf{x} \in S$ be such that $q(\mathbf{x}_1) = \min\{q(\mathbf{x}) : \mathbf{x} \in S\}$. Suppose $1 \leq k < n$ and we have defined $\mathbf{x}_1, \dots, \mathbf{x}_k$. Let $\mathbf{x}_{k+1} \in S \cap \langle \mathbf{x}_1, \dots, \mathbf{x}_k\rangle^{\perp}$ be a vector at which q attains its minimum on $S \cap \langle \mathbf{x}_1, \dots, \mathbf{x}_k\rangle^{\perp}$:

$$q(\mathbf{x}_{k+1}) = \min\{q(\mathbf{x}) : \mathbf{x} \in S \text{ and } \langle \mathbf{x}, \mathbf{x}_i\rangle = 0 \text{ for } i = 1, \dots, k\}.$$

Show that for each i, $1 \leq i \leq n$, $\mathbf{x}_i$ is an eigenvector of T with eigenvalue λ_i.

48. (Exercise 47 contd.) Show that

$$\lambda_{k+1} = \max\{\min\{q(\mathbf{x}) : \mathbf{x} \in S \cap U_{n-k}\} : U_{n-k} \subset V,\ \dim U_{n-k} = n-k\}.$$

49. Show that for every $k \geq 0$ there are infinitely many k-connected graphs G with vertex connectivity $\kappa(G) = k = \lambda_2(G)$. [*Hint.* Consider $(K_r \cup K_r) + \overline{K}_k$ for $r > k$.]

50.$^+$ Let G be a graph of order n and maximal degree Δ, and set $c = 2\lambda_2(G)/(\Delta + 2\lambda_2(G))$. Show that for every set U of at most $n/2$ vertices there are at least $c|U|$ vertices not in U that are joined to vertices in U. (Such a graph is said to be an (n, Δ, c)-*expander*.)

51. Let $P_G(x) = \sum_{k=0}^{n} c_k x^{n-k}$ be the characteristic polynomial of (the adjacency matrix of) a graph G. Show that $c_0 = 1$, $c_1 = 0$, $c_2 = -e(G)$ and $-c_3$ is twice the number of triangles in G.

52.$^+$ Let $P_G(x)$ be the characteristic polynomial of G, as in Exercise 51. Show that if $e = xy$ is a bridge of G then

$$P_G(x) = P_{G-e}(x) - P_{G-\{x,y\}}(x).$$

Now, let F be a forest with $2n$ vertices and denote by d_k the number of k-element sets of independent edges. [Thus d_n is the number of 1-factors.] Prove that

$$P_G(x) = x^{2n} - d_1 x^{2n-2} + d_2 x^{2n-4} - \cdots + (-1)^n d_n.$$

What are the possible values for d_n?

53. Let G be a connected k-regular graph containing an odd cycle. At time 0 put a counter on a vertex. For each counter that is, on a vertex x at time t, place a counter on every vertex adjacent to x at time $t+1$ and remove the counters from t. Show that $n_t(x)/k^t$ tends to a limit as $t \to \infty$, where $n_t(x)$ is the number of counters on x at time t. What is the corresponding assertion if the counters are not removed from the vertices? [*Hint.* There is an orthonormal basis consisting of the eigenvectors of the adjacency matrix.]

54. A k-regular graph G of order n is such that any two non-adjacent vertices can be mapped into any other such two by some automorphism of G. Show that G is strongly regular. What are the eigenvalues of G? What is the condition that k and n have to satisfy if there is such a graph?

55. Let C be the collapsed adjacency matrix of a highly regular graph. Show that $(C^\ell)_{11}$ is the number of walks of length ℓ from a vertex to itself. Interpret the other entries of C^ℓ.

56. Why must the collapsed adjacency matrix of the Petersen graph, as shown in Fig. VIII.11, have rational eigenvalues?

57.$^+$ In the graph G every two adjacent vertices have exactly one common neighbour and every two non-adjacent vertices have exactly two common neighbours. Show that G is regular of degree $2k$ and has order $2k^2 + 1$, for some k in $\{1, 2, 7, 11, 56, 497\}$.

58.$^+$ Let G be a strongly regular graph of order 100 with parameters $(k, 0, b)$. What are the possible values of k? Find the eigenvalues when $k = 22$.

59.$^+$ Make use of the result of Exercise 19 to calculate the eigenvalues of the Tutte 8-cage.

60. Give detailed proofs of Theorem 10 and Corollary 11.

61. Prove that the number of trees with $n - 1 \geq 2$ labelled *edges* is n^{n-3}.

62. Show that a given vertex has degree 1 in about $1/e$ of all labelled trees, where $e = 2.71828\ldots$ (Cf. Exercise VII.15).

63. Denote by $T(n, k)$ the number of trees with n labelled vertices of which exactly k have degree 1. Prove that

$$\frac{k}{n}T(n,k) = (n-k)T(n-1,k-1) + kT(n-1,k).$$

64. Prove that there are

$$\frac{1}{m!}\sum_{j=0}^{m}\left(-\frac{1}{2}\right)^j \binom{m}{j}\binom{n-1}{m+j-1} n^{n-m-j}(m+j)!$$

acyclic graphs on n labelled vertices having $n - m$ edges. Deduce that there are $\frac{1}{2}(n-1)(n+6)n^{n-4}$ forests with two components.

65.$^+$ Show that the cycle sum of the symmetric group S_n acting on the usual n letters is

$$\sum \frac{n!}{\prod_{k=1}^{n} k^{j_k} j_k!} a_1^{j_1} a_2^{j_2} \cdots a_n^{j_n},$$

where the summation is over all partitions $j_1 + 2j_2 + \cdots + nj_n = n$.
What is the cycle sum of S_5 acting on the unordered pairs of 5 elements? How many non-isomorphic graphs are there on 5 vertices? How many of them have 5 edges?

66.$^+$ Let $Z(\Gamma; a_1, a_2, \ldots, a_d)$ be the cycle index of a permutation group Γ acting on a set D. Consider the action of Γ on the set of all k-subsets of D. How many orbits are there?

67. Determine the cycle index of the rotations of the cube acting on (i) the vertices, (ii) the edges, (iii) the faces, (iv) the faces and vertices. In how many distinct ways can you colour the vertices using some of n colours? The edges? The faces? The vertices and faces?

68. Show that the cycle sum of C_n, the cyclic group of order n, is

$$\tilde{Z}(C_n; a_1, \ldots, a_n) = \sum_{k|n} \phi(k) a_k^{n/k}$$

where $\phi(k)$ is the Euler function.

69.$^+$ Prove that the cycle index of the dihedral group D_n (cf. Exercise 8) is

$$Z(D_n; a_1, \ldots, a_{2n}) = \frac{1}{2} Z(C_n; a_1, \ldots, a_n) + f(a_1 a_2),$$

$$f(a_1, a_2) + \begin{cases} \frac{1}{2} a_1 a_2^{(n-1)/2} & \text{if } n \text{ is odd,} \\ \frac{1}{4}\left(a_2^{n/2} + a_1^2 a_2^{(n-2)/2}\right) & \text{if } n \text{ is even.} \end{cases}$$

How many bracelets are there with 20 beads coloured red, blue and green?

70. How many distinct ways are there of colouring the faces of a dodecahedron with red, blue and green, using each colour at least once?

71.$^+$ Let S_n be the symmetric group of all permutations of $[n] = \{1, 2, \ldots, n\}$. A *transposition basis* is a minimal set $B \subset S_n$ of transpositions generating S_n. Prove that there are precisely n^{n-2} transposition bases.

72. A graph is *vertex-transitive* if any two vertices can be mapped into each other by an automorphism, and it is *edge-transitive* if any two edges (*unordered* pairs of adjacent vertices) can be mapped into each other by an automorphism. Also, a graph is 1-*transitive* if any two *ordered* pairs of adjacent vertices can be mapped into each other. Find a graph which is edge-transitive but not vertex-transitive, and a graph that is edge-transitive but not 1-transitive.

73.$^+$ Construct a graph that is vertex-transitive and edge-transitive but not 1-transitive. [*Hint.* First find an infinite graph, and then 'project' it on a finite graph.]

In Exercises 74-76, G is a graph of order n, size m and maximal degree Δ, *with eigenvalues* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

74. Show that $\lambda_1 \geq \sqrt{\Delta}$

75. Show that $2m/n \leq \lambda_1 \leq \sqrt{2m(n-1)/n}$.

76. Show that if G is k-regular and has diameter D then $k - \lambda_2 = \lambda_1 - \lambda_2 > 1/D$.

77.$^+$ Show that if G is a bipartite graph without 0 as an eigenvalue then G has a complete matching.

Notes

There is a vast literature concerned with group presentations, including the use of Cayley and Schreier diagrams. The basic book is perhaps W. Magnus, A. Karrass and D. Solitar, *Combinatorial Group Theory*, 2nd ed., Dover, New York, 1976; the connections with geometry are emphasized in H.S.M. Coxeter, *Regular Complex Polytopes*, Cambridge University Press, New York, 1974. Numerous articles deal with the computational aspect, in particular J.A. Todd and H.S.M. Coxeter, A practical method for enumerating cosets of a finite abstract group, *Proc. Edinburgh Math. Soc.* (2) **5** (1936) 26–34, which was the first paper in this line and J. Leech, Computer proof of relations in groups, in *Topics in Group Theory and Computation* (M.P.J. Curran, ed.), Academic Press, New York, 1977, in which some more recent developments are described.

Max Dehn posed the word problem in Über unendliche diskontinuierliche Gruppen, *Math. Ann.* **71** (1911) 116–144, and gave the above discussed presentation of the group of the trefoil in Über die Topologie des dreidimensionalen Raumes, *Math. Ann.* **69** (1910) 137–168. The word problem was shown to be intrinsically connected to logic by P.S. Novikov, On the algorithmic unsolvability of the word problem, *Amer. Math. Soc. Transl.* (2) **9** (1958) 1–22 and G. Higman, Subgroups of finitely presented groups, *Proc. Royal Soc.* A, **262** (1961) 455–475. For the properties of knots and their groups the reader is referred to R.H. Crowell and R.H. Fox, *Introduction to Knot Theory*, Graduate Texts in Mathematics, Vol. 57, Springer-Verlag, New York, 1977, and G. Burde and H. Zieschang, *Knots*, de Gruyter Series in Mathematics 5, Walter de Gruyter, Berlin-New York, 1985.

An exposition of general matrix methods in graph theory can be found in N. Biggs, *Algebraic Graph Theory*, Cambridge University Press, 2nd ed., Cambridge, 1993. For a grounding in basic functional analysis, including the use of numerical ranges, see B. Bollobás, *Linear Analysis*, Cambridge Univ. Press, 1990. The first striking result obtained by matrix methods, Theorem 19, is due to A.J. Hoffman and R.R. Singleton, On Moore graphs with diameters 2 and 3, *IBM J. Res. Dev.* **4** (1960) 497–504. Theorem 10 and its corollary are from T.S. Motzkin and E.G. Straus, Maxima for graphs and a new proof of a theorem of Turán, *Canadian Journal of Mathematics* 17 (1965) 535–540. Alon and Milman were the first to make good use of the Laplacian in graph theory: for the first results, see N. Alon and V. Milman, λ_1, isoperimetric inequalities for graphs, and superconcentrators, *J. Combin. Theory* (B) **38** (1985) 73–88, and N. Alon, Eigenvalues and expanders, *Combinatorica* **6** (1986) 83–96. We should remark here that our use of λ_2 for the second smallest eigenvalue of the Laplacian is somewhat non-standard: it is more usual to write $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$ for the eigenvalues. In particular, the λ_1 in the title of the Alon–Milman paper above is our λ_2. For a detailed study of the spectrum of the Laplacian see F.R.K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, vol. 92, Amer. Math. Soc., Providence, 1997.

The Cauchy–Frobenius lemma used to be called Burnside's lemma, as it appeared without attribution in the book of W. Burnside, *Theory of Groups of Finite Order*, 2nd ed., Cambridge University Press, Cambridge, 1911. For a fascinating account of this lemma, see P. Neumann, A lemma that is not Burnside's, *Math. Scientist* **4** (1979) 133–141.

The fundamental enumeration theorem of G. Pólya appeared in Kombinatorische Anzahlbestimmungen für Gruppen und chemische Verbindungen, *Acta Math.* **68** (1937) 145–254. Many enumeration techniques were anticipated by J.H. Redfield, The theory of group-reduced distributions, *Amer. J. Math.* **49** (1927)

433–455. The standard reference book for Pólya-type enumeration is F. Harary and E.M. Palmer, *Graphical Enumeration*, Academic Press, New York, 1973.

The recent two-volume treatise *Handbook of Combinatorics* (R.L. Graham, M. Grötschel and L. Lovász, eds), North-Holland, Amsterdam, 1995, contains several articles going much deeper into various aspects of algebraic combinatorics than we could in this chapter: see, in particular, the review articles by N. Alon (Tools from higher algebra), L. Babai (Automorphism groups, isomorphism, reconstruction), P.J. Cameron (Permutation groups), I.M. Gessel and R.P. Stanley (Algebraic enumeration), and C.D. Godsil (Tools from linear algebra).

IX

Random Walks on Graphs

Random walks on graphs and Markov chains with a finite number of states have been investigated for over 90 years, but their study really took off only in the last two decades or so. The main reasons for this heightened activity are the systematic exploitation of the surprising and extremely useful connection with electrical networks, the emergence of intricate combinatorial arguments, the use of the spectral properties of relevant matrices, and applications of harmonic analysis. In this chapter we shall dip into the theory of random walks on graphs, emphasizing combinatorial arguments, the connection with electrical networks, and eigenvalues.

A *random walk* on a graph is precisely what its name says: a walk $X_0X_1\cdots$ obtained in a certain random fashion. In its simplest form, it depends only on the graph and nothing else. Starting a simple walk at X_0, its next vertex, X_1, is chosen at random from among the neighbours of X_0, then X_2 is a random neighbour of X_1, and so on. In fact, this *simple random walk* on a graph is only a little less general than a reversible finite Markov chain: attaching weights to the edges and allowing loops, every reversible finite Markov chain can be obtained in this way. Following the usual notation for Markov chains, instead of $X_0X_1\cdots$, we write $X_0, X_1, \ldots$ for a random walk. At the first sight, random walks on graphs seem to be rather special finite Markov chains, but this is not the case: finite Markov chains are just random walks on weighted directed graphs, with loops allowed. In view of this, it is not surprising that random walks on graphs are of great importance.

In Chapter II we introduced electrical networks and studied their basic properties, culminating in Kirchhoff's theorem. In Section 1 we shall go a little deeper into their theory: rather than taking a static view whereby currents are solutions of systems of linear equations or ratios of quantities described in terms of graphs, we describe currents as variables minimizing certain quadratic energy functions. This implies that the current distribution is rather stable: small changes in the resistances of the wires do not lead to a radically different current distribution. More importantly, it also implies that cutting a wire does not decrease the total resistance and shorting vertices does not increase the resistance.

The connection between random walks and electrical networks is established in Section 2. This intimate connection greatly benefits both areas: we can use random walks to prove results about electrical networks, and conversely, we can use our theory of electrical networks to prove beautiful results about random walks. A highlight of Section 2 is a stunning proof of Pólya's classical theorem on random walks on lattices, based on the connection with electrical networks.

In Section 3 we shall study the standard parameters of random walks such as hitting times, commute times and return times. In addition to combinatorial arguments, we shall continue exploiting the connection with electrical networks.

The last section concerns a central question of random walks: how fast is the convergence to the stationary distribution? As we shall see, the speed of convergence is governed by the expansion properties of the graph.

IX.1 Electrical Networks Revisited

Let us recapitulate briefly the concepts encountered in Chapter II. An *electrical network* $N = (V, E, r) = (G, r)$ is a multigraph $G = (V, E)$, together with a function $r : E \to \mathbb{R}^+$, where $r_e = r(e)$ is the *resistance* of the edge e. It is frequently convenient to give our network in the form $N = (G, c)$, where c is the conductance function, so that $c_e = 1/r_e$ is the *conductance* of the edge e.

If there is a *potential difference* $p_e = p_{ab}$ in an edge e from a to b, then an *electrical current* w_e will flow in e from a to b according to *Ohm's law* (OL): $w_e = p_e/r_e$. The distribution of currents is governed by Kirchhoff's laws.

Kirchhoff's potential law (KPL) postulates that the sum of potential differences around any cycle is 0, and *Kirchhoff's current law* (KCL) states that the total current into a vertex is 0. In calculating the total current into a vertex, we have to take into account the amount of current both entering and leaving the network at that vertex.

Kirchhoff's theorem (Theorems II.1) gives a combinatorial interpretation of the currents in the edges of an electrical network resulting in a current of size 1 from a source s to a sink t. This easily implies the corresponding result for several sources and sinks: if $s_1, \dots, s_k$ are vertices of a (connected) electrical network $N = (V, E, r)$ and the real numbers $w_1, \dots, w_k$ sum to 0, then there is a unique distribution of currents and potential differences in the edges such that, for $i = 1, \dots, k$, a current of size w_i enters ($-w_i$ leaves) the network at s_i, and at no other vertex does any current enter (or leave) the network.

Somewhat surprisingly, from Kirchhoff's theorem it is not easy to show that *if a wire is cut then the total resistance of the network between two vertices does* *not decrease.* Of course, if you believe that the three laws of electric currents describe a physical system with some properties we consider "natural", then this *monotonicity principle* is self-evident. However, we know that, having postulated the three laws, we have had our say: there is a *unique* distribution of currents and there is a *well defined total resistance,* so we cannot appeal to our physical intuition. This difficulty in proving the above monotonicity principle demonstrates the shortcomings of the static approach based on using at once the full force of Kirchhoff's laws and Ohm's law: although Kirchhoff's theorem tells us that there is a solution, this solution seems to be an unpredictable and unstable function of the equations.

As we shall see in this section, it is much better to use only *some* of the conditions given by our laws and define certain functions that attain their minima at places satisfying the remaining equations. Unlike the solutions of the full system of equations, these optimization problems behave in an easily predictable fashion, enabling us to get a much better insight into the distribution of currents and potentials. In particular, we shall give several explicit optimization methods for constructing the currents and potentials. As a result of these methods, we can give several proofs of the monotonicity principle.

If we do not wish to take into account all three laws at once, then there are two natural ways open to us in our search for the proper electric currents and potentials. We may consider currents satisfying Kirchhoff's current law, use Ohm's law to define potential differences and then use a function to select currents that satisfy Kirchhoff's potential law, or we may consider a distribution of potential differences satisfying KPL, define currents by OL and then use a function to select the currents that satisfy KCL. As we shall see, we do not have to *try* hard: it will actually be very easy.

Note that our aim is to prove the *existence* of a proper distribution of currents: as we know from Section II.1, *uniqueness* is immediate from the superposition principle.

Let us recall first that KPL is equivalent to the possibility of defining an *absolute potential* V_x for every vertex x such that $p_{xy} = V_x - V_y$ for each edge xy. Indeed, if (p_{xy}) is a distribution of potential differences satisfying KPL and $ux_1x_2 \cdots x_k v$ and $uy_1y_2 \cdots y_l v$ are u–v paths then

$$p_{ux_1} + p_{x_1x_2} + \cdots + p_{x_kv} = p_{uy_1} + p_{y_1y_2} + \cdots + p_{y_lv}. \tag{1}$$

To define absolute potentials, pick a vertex v, and set $V_v = 0$, say. For each vertex u, pick a u–v path $ux_1x_2 \cdots x_k v$ and set

$$V_u = p_{ux_1} + p_{x_1x_2} + \cdots + p_{x_{k-1}x_k} + p_{x_kv}.$$

By (1), V_u is well-defined, i.e., independent of the u–v path $ux_1x_2 \cdots x_k v$. It is immediate that $p_{xy} = V_x - V_y$ for every edge xy. The converse is even more trivial: if (V_x) is an assignment of absolute potentials to the vertices then $p_{xy} = V_x - V_y$ gives a distribution of potential differences satisfying KPL.

Let us write out explicitly the two natural ways of getting proper electric currents in a network, with no current leaving or entering the network at vertices other than $s_1, \dots, s_k$.

The approach assuming KCL and OL. Consider a flow (u_{xy}) with outlets (sources and sinks) $s_1, \dots, s_k$, i.e., with $\sum_{y \in \Gamma(x)} u_{xy} = 0$ for every $x \neq s_1, \dots, s_k$. In order to turn the flow into a proper electric current with outlets (sources and sinks) $s_1, \dots, s_k$, all we have to make sure is that KPL holds, i.e., that

$$\sum_{i=1}^{k} u_{x_i x_{i+1}} r_{x_i x_{i+1}} = 0 \tag{2}$$

for every cycle $x_1 x_2 \cdots x_k$, with $x_{k+1} \equiv x_1$.

The approach assuming KPL and OL. Consider a distribution (V_x) of absolute potentials on the vertices. This distribution gives a proper electric current with outlets $s_1, \dots, s_k$ iff KCL holds, i.e.,

$$\sum_{y \in \Gamma(x)} \frac{V_x - V_y}{r_{xy}} = 0 \tag{3}$$

for every vertex $x \neq s_1, \dots, s_k$. (Note that every assignment (V_x) of absolute potentials gives a distribution of currents, but there may be some current leaving or entering the network at vertices other than $s_1, \dots, s_k$.) Let us rewrite (3) in terms of the conductances $c_{xy} = 1/r_{xy}$, with $C_x = \sum_{y \in \Gamma(x)} c_{xy}$. A distribution (V_x) of absolute potentials results in a distribution of electric currents with outlets $s_1, \dots, s_k$ if, and only if, for every vertex $x \neq s_1, \dots, s_k$ we have

$$C_x V_x = \sum_{y \in \Gamma(x)} c_{xy} V_y. \tag{4}$$

In both methods above, we shall use the same function to find the currents. Given an edge xy with resistance r_{xy}, potential difference $p_{xy} = V_x - V_y$, and so a current of size $w_{xy} = p_{xy}/r_{xy} = (V_x - V_y)/r_{xy}$, the *energy* in xy is defined to be

$$w_{xy}^2 r_{xy} = \frac{(V_x - V_y)^2}{r_{xy}} = (V_x - V_y) w_{xy}.$$

The *total energy* in a network $N = (G, r) = (V(G), E(G), r)$ is

$$\sum_{xy \in E(G)} w_{xy}^2 r_{xy} = \sum_{xy \in E(G)} \frac{(V_x - V_y)^2}{r_{xy}} = \sum_{xy \in E(G)} (V_x - V_y) w_{xy}. \tag{5}$$

In the formulae above and in subsequent summations the edges are taken to be oriented in an arbitrary fashion. This is simply to avoid double summation; if the edges are not taken to be oriented, then the total energy is defined to be $\frac{1}{2} \sum_{x,y} w_{xy}^2 r_{xy}$, with the convention $0 \cdot \infty = 0$, in case xy is not an edge, so $r_{xy} = \infty$ and $w_{xy} = 0$. In particular, the last part of formula (5) is clearly ill-defined if the edges are not oriented. Furthermore, the formulae above are, strictly speaking, incorrect even with oriented edges: as our network may have several edges from x to y, in (3), (4) and (5) the sums stand for summations over *all* edges from x to y, and r_{xy}, c_{xy} and w_{xy} are functions of the particular *edge* from x to y, rather than of the pair (x, y). However, it is unlikely that this shorthand will lead to confusion.

Let us pause for a moment to remark that we lose nothing from the generality if we restrict our attention to *simple* electrical networks, that is, to networks in which each edge has resistance 1, since every electrical network can be approximated by an electrical network (with more edges) in which all edges have the same resistance. In many calculations it is convenient to have general resistances, while occasionally, as in (5), the concepts are clearer for simple networks. Indeed, for a simple network $N = (G, 1)$, the total energy in N is the value of the quadratic form given by the Laplacian $L = D - A$ on the vector (V_x) of absolute potentials. Of course, in the general case we are hardly worse off: all we have to take is the Laplacian of the weighted graph with the conductance $c_e = 1/r_e$ for the weight of an edge e.

Let us return to our task of showing the existence of currents satisfying all three laws. *Thomson's principle* says that currents and potentials are distributed in such a way as to minimize the total energy in the network. There are two forms of this result: in Theorem 1 we choose potentials and in Theorem 2 currents. Theorem 1 is also called *Dirichlet's principle.*

Theorem 1 *Let $N = (G, r)$ be an electrical network, $s_1, \dots, s_k \in V(G)$, and $V_{s_1}, \dots, V_{s_k} \in \mathbb{R}$. Then there are absolute potentials V_x, $x \in V(G) \setminus \{s_1, \dots, s_k\}$ such that*

$$E = E\big((V_x)\big) = \sum_{xy \in E(G)} \frac{(V_x - V_y)^2}{r_{xy}}$$

is minimal. This distribution (V_x) of absolute potentials gives a proper electric current with no outlet other than $s_1, \dots, s_k$. The minimum of E is precisely the total energy of the electric current.

Proof. Since the energy function E is a continuous function of the absolute potentials $(V_x) \in \mathbb{R}^{V(G)}$, and $E \to \infty$ as $\max|V_x| \to \infty$, the infimum of E is indeed attained at some (V_x). Furthermore, at this point $(V_x) \in \mathbb{R}^{V(G)}$ we have

$$\frac{\partial E}{\partial V_x} = 0$$

for every $x \neq s_1, \dots, s_k$, so

$$\sum_{y \in \Gamma(x)} \frac{2(V_x - V_y)}{r_{xy}} = 2 \sum_{y \in \Gamma(x)} w_{xy} = 0.$$

Hence the absolute potentials do define a distribution of currents (via Ohm's Law) satisfying KCL. $\square$

Theorem 2 *Let $N = (G, r)$ be an electrical network, $s_1 \dots, s_k \in V(G)$, and let $u_{s_1}, \dots, u_{s_k} \in \mathbb{R}$, with $\sum_{i=1}^{k} u_{s_i} = 0$. Consider the energy function*

$$E = E(u) = \sum_{xy \in E(G)} u_{xy}^2 r_{xy}$$

for flows $u = (u_{xy})$ in which a current of size u_{s_i} enters the network at s_i (i.e., a current of size $-u_{s_i}$ leaves the network at s_i), $i = 1, \dots, k$, and at no other vertex does any current enter or leave the network. There is such a flow minimizing $E(u)$, and this flow satisfies KPL, so it is a proper electric current. The minimum of $E(u)$ is precisely the total energy in the current.

Proof. Once again, compactness implies that the infimum of $E(u)$ is attained at some flow $u = (u_{xy})$. Given a cycle $x_1 x_2 \cdots x_\ell$, $x_{\ell+1} \equiv x_1$, let $u(\varepsilon)$ be the flow obtained from u by increasing each $u_{x_i x_{i+1}}$ by ε for $i = 1, \dots, \ell$. Then

$$\frac{dE(u(\varepsilon))}{d\varepsilon} = 0$$

at $\varepsilon = 0$, so

$$2 \sum_{i=1}^{\ell} u_{x_i x_{i+1}} r_{x_i x_{i+1}} = 0.$$

Thus KPL holds, as claimed. □

The *effective conductance* $C_{\text{eff}} = C_{\text{eff}}(s, t)$ of an electrical network from s to t is the value of the current from s to t if s and t are set at potential difference 1. The *effective resistance* is $R_{\text{eff}} = R_{\text{eff}}(s, t) = 1/C_{\text{eff}}(s, t)$, the potential difference between s and t ensuring a current of size 1 from s to t.

The next result, *Rayleigh's principle* or the *conservation of energy principle,*, implies that if we replace a network with a source s and a sink t with a single wire whose resistance is the effective resistance of the network, then the total energy in the system does not change.

Theorem 3 *Let $u = (u_{xy})$ be a flow from s to t with value*

$$u_s = \sum_{y \in \Gamma(s)} u_{sy} = - \sum_{z \in \Gamma(t)} u_{tz} = -u_t,$$

i.e., let u be a flow satisfying KCL at each vertex other than s and t, and let (V_x) be any function on the vertices. Then

$$(V_s - V_t) u_s = \sum_{xy \in E(G)} (V_x - V_y) u_{xy}.$$

Proof. The right-hand side is

$$\sum_{x \in V(G)} V_x \Big(\sum_{y \in \Gamma^+(x)} u_{xy} - \sum_{z \in \Gamma^-(x)} u_{zx} \Big) = V_s u_s + V_t u_t = (V_s - V_t) u_s. \quad \square$$

Corollary 4 *The total energy in an electric current from s to t is $(V_s - V_t)w_s$, where $w_s = \sum_{x\in\Gamma(s)} w_{sx}$ is the value of the current. If $V_s - V_t = 1$ then the total energy is equal to the size of the current; i.e., the total energy, the total current and the effective conductance are the same. If $w_s = 1$ then the total energy is the potential difference between s and t; i.e., the total energy, the potential difference and the effective resistance are the same.*

Proof. This is immediate from Theorem 3. □

Theorem 1 and Corollary 4 imply an expression for the effective conductance.

Corollary 5 *The effective conductance $C_{\text{eff}}(s, t)$ of a network between s and t is*

$$C_{\text{eff}}(s, t) = \inf\left\{ \sum_{xy\in E(G)} \frac{(V_x - V_y)^2}{r_{xy}} : V_s = 1, V_t = 0 \right\}. \tag{6}$$

Similarly, Theorem 2 and Corollary 4 give a rather useful expression for the effective resistance $R_{\text{eff}}(s, t)$.

Corollary 6 *The effective resistance $R_{\text{eff}}(s, t)$ of a network between s and t is*

$$R_{\text{eff}}(s, t) = \inf\left\{ \sum_{xy\in E(G)} u_{xy}^2 r_{xy} : (u_{xy}) \text{ is an } s\text{–}t \text{ flow of size } 1 \right\}. \tag{7}$$

Either of Corollaries 5 and 6 implies the Holy Grail of this section, the *monotonicity principle.*

Corollary 7 *If the resistance of a wire is increased then the effective resistance (between two vertices) does not decrease. In particular, if a wire is cut, the effective resistance does not decrease, and if two vertices are shorted, the effective resistance does not increase.*

Proof. If $r_{x_0y_0}$ is increased then the expression for $C_{\text{eff}}(s, t)$ in Corollary 5 does not increase. Equivalently, the expression for $R_{\text{eff}}(s, t)$ in Corollary 6 does not decrease. □

Our next aim is to establish a connection between random walks and electrical networks; the results above will then be very useful in attacking questions on random walks.

IX.2 Electrical Networks and Random Walks

Given a pair (G, S), where G is a simple graph and $S \subset V(G)$, a function $f : V(G) \to \mathbb{R}$ is said to be *harmonic on G with boundary S* if

$$f(x) = \frac{1}{d(x)} \sum_{y\in\Gamma(x)} f(y) \tag{8}$$

whenever $x \in V(G) \setminus S$ and $d(x) \geq 1$. Harmonic functions are of central importance in the theory of electrical networks, and we did encounter them in the previous section, without calling them by their name. Indeed, given a graph G, turn it into a simple electrical network by giving each edge xy conductance $c_{xy} = 1$, and take $S = \{s_1, \ldots, s_k\}$. Then (4) states precisely that if no current leaves or enters G at vertices other than the s_i then the function of absolute potentials, V_x, is harmonic on G with boundary S:

$$V_x = \frac{1}{d(x)} \sum_{y \in \Gamma(x)} V_y, \tag{9}$$

whenever $x \in V(G) \setminus S$ and $d(x) \geq 1$.

Another natural source of harmonic functions on a graph is a random walk on the graph. Given a connected graph G with a non-empty set $S \subset V(G)$, let g be a real-valued function on S. For each $x \in V(G)$, play the following game. Starting at x, move about in G at random, stopping as soon as you reach a vertex s of S: if this happens, x wins $g(s)$. To be precise, set $X_0 = x$. Having defined $X_t = y$, if $y \in S$, stop the sequence; otherwise, pick a neighbour z of y at random, and set $X_{t+1} = z$. If this random walk $X_0, X_1, \ldots$ terminates in $s \in S$, then x wins $g(s)$. Now, let E_x be the expected gain of x. Thus, if $x \in S$ then $E_x = g(x)$; otherwise, the expected gain of x is the average of the expectations after one step i.e.,

$$E_x = \frac{1}{d(x)} \sum_{y \in \Gamma(x)} E_y. \tag{10}$$

This shows that E_x is a harmonic function on G with boundary S. Also, if $S = \{s_1, \ldots, s_k\}$ and $g(s_i) = V_{s_i}$, then equations (9) and (10) imply that (E_x) is precisely the distribution of absolute potentials if each s_i is set at V_{s_i} and no current leaves or enters the network at vertices other than the s_i.

The importance of harmonic functions in the study of both electrical currents and random walks on graphs establishes an intimate connection between the two areas. This alone would suffice to make the study of random walks on graphs worthwhile, but there is another, even more compelling reason: random walks on weighted graphs, to be introduced next, are precisely the reversible finite Markov chains. For an easy justification of this statement see Exercise 11.

The aim of this section is to introduce random walks on graphs and to present the intimate connection between random walks and electrical networks. We shall show that this connection greatly benefits both theories by giving two more proofs of the monotonicity principle, and by making use of electrical networks in the study of random walks.

A (discrete-time) *Markov chain* on a finite or countable set V of *states* is a sequence of random variables $X_0, X_1, \ldots$ taking values in V such that for all $x_0, \ldots, x_{t+1} \in V$, the probability of $X_{t+1} = x_{t+1}$, conditional on $X_0 = x_0, \ldots,$ $X_t = x_t$, depends only on x_t and x_{t+1}. As most of our Markov chains will be defined by graphs, we shall tend to call them *random walks*.

Let G be a graph of order n without multiple edges but with a loop allowed at each vertex. To each edge and loop xy, we assign a positive *weight* [avoirdupois weight?] $a_{xy} > 0$. In particular, a_{xx} is the weight of the loop xx at x. Writing a for the function $xy \mapsto a_{xy}$, we have obtained the *weighted graph* (G, a).

Now, given a weighted graph (G, a), for every vertex $x \in V(G)$, let $A_x = \sum_{y \in \Gamma(x)} a_{xy}$, and for $x, y \in V(G)$, define

$$P_{xy} = \begin{cases} a_{xy}/A_x & \text{if } x \text{ is joined to } y \text{ by an edge or loop,} \\ 0 & \text{otherwise.} \end{cases}$$

Thus $P = (P_{xy})$ is an $n \times n$ matrix with non-negative entries in which each row-sum is 1.

A random walk defined by a weighted graph is a Markov chain on $V = V(G)$ with *transition probability matrix* $(P_{xy})_{x,y \in V}$, so that P_{xy} is the probability of going from x to y. By a *random walk* (RW) *on a weighted graph* we shall mean a random walk with this particular transition matrix. Thus an RW is a sequence of random variables $X_0, X_1, \ldots$, each taking values in the set V of vertices, such that

$$\mathbb{P}(X_{t+1} = x_{t+1} | X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t) = P_{x_t x_{t+1}}$$

for every walk (!) $x_0 x_1 \cdots x_t$ in the graph G. If G does not contain the walk $x_0 x_1 \cdots x_t$ then $\mathbb{P}(X_0 = x_0, \ldots, X_t = x_t) = 0$. If $X_t = y$ then we say that at time t the walk is at y.

Strictly speaking, we tend to consider the entire class of RWs with the same transition probability matrix $P = (P_{xy})$, so that $(X_t)_0^\infty$ stands for *any* of the RWs with this transition matrix. To select one of these RWs, we usually fix the *initial distribution*, i.e., the distribution of X_0. In fact, much of the time we start our random walk $(X_i)_0^\infty$ at a given vertex, so that $X_0 = x_0$ for some $x_0 \in V(G)$. It will be convenient to identify ourselves with the random walk; thus we may say that "starting at x, we get to y before we get to z".

It is only slightly less natural to define a random walk on a weighted *multigraph*, with multiple edges and loops allowed. Let (G, a) be a multigraph with weight function $a : E(G) \to \mathbb{R}$, $e \mapsto a_e > 0$. For $x \in V(G)$, let $\overline{A}_x$ be the sum of the weights a_e of all the edges and loops e incident with x, and for $x, y \in V(G)$ let $\overline{a}_{xy}$ be the sum of the weights of all the edges or loops joining x to y. Then $P = (P_{xy})$, given by $P_{xy} = \overline{a}_{xy}/\overline{A}_x$, is the transition probability matrix of an RW on the weighted graph (G, a).

Intuitively, in an RW on a weighted multigraph, if we are at a vertex x, then we choose at random one of the *edges* and *loops* incident with x (rather than one of the neighbouring vertices) according to their weights and traverse that edge to the other endvertex.

Now, the weighted multigraphs we are especially interested in are the electrical networks (G, c), with conductance function c: the weight of an edge e is taken to be its conductance c_e. An RW on an electrical network will always be taken to be an RW with this weight function.

To simplify the notation, we shall assume that our electrical network does not have multiple edges and loops, so that we can write c_{xy} for the conductance of *the* edge from x to y. In fact, it will be convenient to have no loops, although the existence of loops would not affect our formulae: the only difference is that if we have no loops then our RW is not allowed to linger at a vertex. Thus, from now on an electrical network $N = (G, c)$ will be assumed to be on a simple graph G, so that an RW on N will have transition probability matrix $P = (P_{xy})$, given by

$$P_{xy} = \begin{cases} c_{xy}/C_x & \text{if } xy \in E(G), \\ 0 & \text{otherwise,} \end{cases}$$

where $C_x = \sum_{y \in \Gamma(x)} c_{xy}$.

After all this preamble, let us get down to a little mathematics. First we return to the example of absolute potentials given by random walks, but state the result in a slightly different way. We shall again consider only networks with one source and one sink, but we shall prove a little more than in our earlier remarks. To describe the currents, we introduce the *probability* $P_{\text{esc}} = P_{\text{esc}}(s \to t)$ *of escaping from* s *to* t, or simply the *escape probability*: the probability that, starting at s, we get to t before we return to s.

Theorem 8 *Let* $N = (G, c)$ *be a connected electrical network, and let* $s, t \in V(G)$, $s \neq t$. *For* $x \in V(G)$ *define*

$$V_x = \mathbb{P}(\text{starting at } x, \text{ we get to } s \text{ before we get to } t),$$

so that $V_s = 1$ *and* $V_t = 0$. *Then* $(V_x)_{x \in V(G)}$ *is the distribution of absolute potentials when* s *is set at* 1 *and* t *at* 0. *The total current from* s *to* t *is*

$$C_{\text{eff}}(s, t) = C_s P_{\text{esc}}(s \to t). \tag{11}$$

Also,

$$\frac{P_{\text{esc}}(s \to t)}{P_{\text{esc}}(t \to s)} = \frac{C_t}{C_s}. \tag{12}$$

Proof. By considering the very first step of the RW started at $x \neq s, t$, we see that

$$V_x = \sum_y P_{xy} V_y = \sum_{y \in \Gamma(x)} \frac{c_{xy}}{C_x} V_y,$$

so (4) follows:

$$C_x V_x = \sum_{y \in \Gamma(x)} c_{xy} V_y.$$

Hence $(V_x)_{x \in V(G)}$ is indeed the claimed distribution of absolute potentials.

Note that

$$P_{\text{esc}}(s \to t) = 1 - \sum_{y \in \Gamma(s)} P_{sy} V_y,$$

since our first step takes us, with probability P_{sy}, to a neighbour y of s, and from there with probability V_y we get to s before we get to t. Hence the total current is

$$\begin{aligned} C_{\text{eff}}(s,t) &= \sum_{y \in \Gamma(s)} (V_s - V_y) c_{sy} = \sum_{y \in \Gamma(s)} (V_s - V_y) \frac{c_{sy} C_s}{C_s} \\ &= C_s \sum_{y \in \Gamma(s)} \left(\frac{c_{sy}}{C_s} - V_y \frac{c_{sy}}{C_s} \right) = C_s (1 - \sum_{y \in \Gamma(s)} P_{sy} V_y) \\ &= C_s P_{\text{esc}}(s \to t), \end{aligned}$$

giving us (11).

Finally, (12) follows easily:

$$\frac{P_{\text{esc}}(s \to t)}{P_{\text{esc}}(t \to s)} = \frac{C_{\text{eff}}(s,t)/C_s}{C_{\text{eff}}(t,s)/C_t} = \frac{C_t}{C_s},$$

since $C_{\text{eff}}(s,t) = C_{\text{eff}}(t,s)$. □

At the risk of being too formal, let us express V_x and $P_{\text{esc}}(s \to t)$ in terms of hitting times. For a set S of states, we define two *hitting times*:

$$\tau_S = \min\{t \geq 0 : X_t \in S\} \text{ and } \tau_S^+ = \min\{t \geq 1 : X_t \in S\}.$$

As we frequently start our RW at a state x in S, it is important to distinguish between the two hitting times. The same definitions can be used for general Markov chains, so that the hitting time defined for random graph processes is precisely τ_S. Also, for $x \in V$, let us write $\mathbb{P}_x$ and $\mathbb{E}_x$ for the probability and expectation conditional on our RW starting from x; if we start from a distribution $\mathbf{p}$, then we write $\mathbb{P}_{\mathbf{p}}$ and $\mathbb{E}_{\mathbf{p}}$. With this notation,

$$V_x = \mathbb{P}_x(X_{\tau_{\{s,t\}}} = s)$$

and

$$P_{\text{esc}}(s \to t) = \mathbb{P}_s(X_{\tau^+_{\{s,t\}}} = t).$$

Analogously to Theorem 8, there is a simple description of absolute potentials and currents in the edges, when a total current of size 1 flows from s to t. The description is in terms of our RW started at s and stopped when we first get to t. For a vertex x of our network, let $S_x = S_x(s \to t)$ be the expected *sojourn time* at x: the expected number of times we are at x before we reach t, if we start at s. In terms of hitting times and conditional expectations,

$$S_x(s \to t) = \mathbb{E}_s(|\{i < \tau_{\{t\}} : X_i = x\}|).$$

Thus if our network is the simplest nontrivial network with two vertices, s and t, joined by an edge of resistance 1, then $S_s(s \to t) = 1$, since if we start at s then $\tau_{\{t\}} = 1$ and $X_0 = s$. Also, $S_t(s \to t) = 0$ for every network.

Theorem 9 *Let $N = (G, c)$ be a connected electrical network with $s, t \in V(G)$, $s \neq t$. For $x \in V(G)$, set $V_x = S_x(s \to t)/C_x$. Furthermore, for $xy \in E(G)$, denote by E_{xy} the expected difference between the number of times we traverse the edge xy from x to y and the number of times we traverse it from y to x, if we start at s and stop when we get to t.*

Then, setting s at absolute potential $R_{\text{eff}}(s, t)$ and t at absolute potential 0, so that there is a current of size 1 from s to t through N, the distribution of absolute potentials is precisely (V_x). In particular,

$$R_{\text{eff}}(s, t) = \frac{S_s(s \to t)}{C_s}. \tag{13}$$

Furthermore, the current in an edge xy is E_{xy}.

Proof. We know that $S_t = 0$, so $V_t = 0$. Let us check that (V_x) satisfies (4) for every $x \neq s, t$. Indeed, we get to x from one of its neighbours, so

$$S_x = \sum_{y \in \Gamma(x)} S_y P_{yx} = \sum_{y \in \Gamma(x)} S_y \frac{c_{xy}}{C_y},$$

which is nothing else but (4):

$$C_x V_x = \sum_{y \in \Gamma(x)} c_{xy} V_y.$$

Hence the distribution (V_x) of absolute potentials does satisfy KCL at every vertex other than s and t. Therefore, with this distribution of absolute potentials, no current enters or leaves the network anywhere other than s and t.

All that remains to check is that we have the claimed current in each edge and that the size of the total current from s to t is 1.

What is the current w_{xy} in the edge xy induced by the potentials (V_x)? By Ohm's law it is

$$\begin{aligned} w_{xy} = (V_x - V_y)c_{xy} &= \left(\frac{S_x}{C_x} - \frac{S_y}{C_y}\right)c_{xy} \\ &= \frac{S_x c_{xy}}{C_x} - \frac{S_y c_{yx}}{C_y} = S_x P_{xy} - S_y P_{yx}, \end{aligned}$$

and the last quantity is precisely E_{xy}.

Finally, the total current through the network from s to t is indeed 1:

$$w_s = \sum_{y \in \Gamma(s)} w_{sy} = \sum_{y \in \Gamma(s)} E_{sy} = 1,$$

since *every* walk from s to t takes 1 more step *from* s (through an edge leaving s) than *to* s (through an edge into s). But since t is at absolute potential 0 and the total current from s to t is 1, the vertex s is at absolute potential $R_{\text{eff}}(s \to t)$, so $V_s = R_{\text{eff}}(s \to t)$, as claimed by (13). ☐

Theorems 8 and 9 give two alternative expressions for $R_{\text{eff}}(s, y)$. Equating them, we find that the escape probability is the reciprocal of the expected sojourn time at s:

$$P_{\text{esc}}(s \to t) = \frac{1}{S_s(s \to t)}. \tag{14}$$

This identity is easily proved directly (see Exercise 18).

Now let us turn to connected and locally finite *infinite* networks. Thus let $N = (G, c)$, where $G = (V(G), E(G))$ is a connected infinite graph in which every vertex has *finite* degree, and $c : E(G) \to \mathbb{R}^+ = (0, \infty)$ is the conductance function. As before, we define an RW on $V(G)$ by defining the transition probability P_{xy} to be c_{xy}/C_x if xy is an edge and 0 otherwise.

Pick a vertex $s \in V(G)$, and let $P_{\text{esc}}^{(\infty)} = P_{\text{esc}}(s, \infty)$ be the probability that, when starting at s, we never return to s. Our RW is said to be *transient* if $P_{\text{esc}}^{(\infty)} > 0$, and it is *recurrent* if $P_{\text{esc}}^{(\infty)} = 0$. It is easily seen that this definition is independent of our choice of s. Analogously to Theorem 8, we have the following result.

Theorem 10 *The RW on a connected, locally finite, infinite electrical network is transient iff the effective resistance between a vertex s and ∞ is finite, and it is recurrent iff the effective resistance is infinite.* □

Although it is intuitively clear what Theorem 10 means and how it follows from Theorem 8, let us be a little more pedantic.

Let us fix a vertex s and, for $l \in \mathbb{N}$, let N_l be the network obtained from N by shorting all the vertices at distance at least l from s to form a new vertex t_l. Let $R_{\text{eff}}^{(l)}$ be the effective resistance of the network N_l between s and t_l, and let $C_{\text{eff}}^{(l)} = 1/R_{\text{eff}}^{(l)}$ be its effective conductance. We know from the monotonicity principle that the sequence $(R_{\text{eff}}^{(l)})$ is increasing and the sequence $(C_{\text{eff}}^{(l)})$ is decreasing, so we may define the effective resistance of N between s and ∞ as $R_{\text{eff}}^{(\infty)} = \lim_{l\to\infty} R_{\text{eff}}^{(l)}$, and the effective conductance as $C_{\text{eff}}^{(\infty)} = \lim_{l\to\infty} C_{\text{eff}}^{(l)}$ (see also Exercise 5).

Let $P_{\text{esc}}^{(l)}$ be the probability that, starting at s, we get to at least distance l from s, before we return to s. It is easily seen that $P_{\text{esc}}^{(\infty)} = \lim_{l\to\infty} P_{\text{esc}}^{(l)}$ (see Exercise 6).

It is immediate that $P_{\text{esc}}^{(l)}$ is also the probability of escaping to t_l in N_l, when starting at s in N_l. By Theorem 8, $P_{\text{esc}}^{(l)} = C_{\text{eff}}^{(l)}/C_s$. Hence $P_{\text{esc}} > 0$ iff $C_{\text{eff}}^{(l)}$ is bounded away from 0, i.e., iff $R_{\text{eff}}^{(l)} = 1/C_{\text{eff}}^{(l)}$ is at most some real r for every l. But this holds iff $R_{\text{eff}}^{(\infty)} \le r$, proving the result.

In view of Theorem 10, we are interested in 'practical' ways of showing that $R_{\text{eff}}^{(\infty)}$ and $C_{\text{eff}}^{(\infty)}$ are bounded by certain quantities. We start with $R_{\text{eff}}^{(\infty)}$.

Theorem 11 *The effective resistance $R_{\text{eff}}^{(\infty)}$ of N between s and infinity is at most r iff there is a current (u_{xy}) in the network N such that a flow of size* 1 *enters the network at s, at no other vertex does any current enter or leave the network, and the total energy in the system, $\sum_{xy\in E(G)} u_{xy}^2 r_{xy}$, is at most r.*

Proof. Suppose that $R_{\text{eff}}^{(l)} \le r$ for every l. Corollary 6 guarantees a flow $u^{(l)}$ of size 1 from s to t_l in N_l, with total energy at most r. By compactness, a subsequence of $(u^{(l)})$ converges to a flow u with the required properties. By Corollary 6, the converse implication is trivial. □

The analogous result for $C_{\text{eff}}^{(\infty)}$ is even easier; it follows at once from Corollary 5 and the definition $C_{\text{eff}}^{(\infty)} = \lim_{l\to\infty} C_{\text{eff}}^{(l)}$.

Theorem 12 *We have $C_{\text{eff}}^{(\infty)} \le C$ iff for every $C' > C$ there is a function (V_x) on the vertex set $V(G)$ such that $V_s = 1$, $V_x = 0$ for all but finitely many vertices x, and $\sum_{xy\in E(G)}(V_x - V_y)^2 c_{xy} < C'$.* □

Theorems 11 and 12 give us the following more explicit version of Theorem 10.

Theorem 13 *Consider the RW on a connected, locally finite infinite electrical network $N = (G, c) = (G, 1/r)$, where $c = 1/r$ is the conductance and r is the resistance. This RW is transient iff there is a flow (u_{xy}) of finite energy $\sum_{xy\in E(G)} u_{xy}^2 r_{xy}$ in which no current leaves at any vertex, but some positive current enters at some vertex. Also, this RW is recurrent iff for every $\varepsilon > 0$ there is a function (V_x) on the vertex set such that $V_s \ge 1$ for some vertex s, $V_x = 0$ for all but finitely many vertices x, and $\sum_{xy\in E(G)}(V_x - V_y)^2 c_{xy} < \varepsilon$.* □

Theorem 13 implies the random walk variant of the monotonicity principle: in proving transience, we may cut edges, in proving recurrence, we may short vertices.

As a striking application of Theorem 13, let us prove Pólya's beautiful theorem on random walks on the lattices $\mathbb{Z}^d$.

Theorem 14 *The simple random walk on the d-dimensional lattice $\mathbb{Z}^d$ is recurrent for $d = 1, 2$ and transient for $d \ge 3$.*

Proof. The simple random walk (SRW) in question is the RW on the electrical network with graph $\mathbb{Z}^d$, where each edge has resistance 1. By the monotonicity principle, all we have to show is that the SRW on $\mathbb{Z}^2$ is recurrent and on $\mathbb{Z}^3$ it is transient.

To prove the first, for $n \ge 1$ short all $8n$ vertices $x = (x_1, x_2)$ with $\|x\|_\infty = \max(|x_1|, |x_2|) = n$ to a new vertex a_n, and set $a_0 = \mathbf{0}$. The new network is a one-way infinite path $a_0a_1a_2\cdots$, with $r_{a_na_{n+1}} = \frac{1}{8n+4}$. Since $\sum_{n=0}^{\infty} \frac{1}{8n+4} = \infty$, the effective resistance between a_0 and ∞ is ∞, so the SRW is indeed recurrent.

Let us give another argument, this time based on Theorem 13. Set $S_m = \sum_{i=1}^{m} 1/i$ and for $x = (x_1, x_2) \in \mathbb{Z}^2$ define

$$V_x = \begin{cases} 1 - S_k/S_n & \text{if } \max(|x_1|, |x_2|) = k < n, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\sum_{xy\in E(\mathbb{Z}^2)} (V_x - V_y)^2 \le \sum_{k=1}^{n} 8k(kS_n)^{-2} = \frac{8}{S_n},$$

and the right-hand side tends to 0 as $n \to \infty$.

To see that the SRW on $\mathbb{Z}^3$ is transient, define a flow $u = (u_{xy})$ in the positive octant as follows. Given a vertex $x = (x_1, x_2, x_3)$ with $x_i \ge 0$ and $x_1+x_2+x_3 = n$, send a current of size $2(x_i + 1)/(n+1)(n+2)(n+3)$ to the vertex $x + e_i$, where (e_1, e_2, e_3) is the standard basis of $\mathbb{R}^3$. Then a current of size $\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$ enters the network at $\mathbf{0}$, and KCL is satisfied at every other vertex, since the total current entering at a vertex $x = (x_1, x_2, x_3)$, $x_1, x_2, x_3 \ge 0$, $x_1 + x_2 + x_3 = n \ge 1$, is $2(x_1+x_2+x_3)/n(n+1)(n+2) = 2/(n+1)(n+2)$, and the total current leaving it is $2(x_1 + 1 + x_2 + 1 + x_3 + 1)/(n+1)(n+2)(n+3) = 2/(n+1)(n+2)$. The total energy of this current is

$$\sum_{n=0}^{\infty} \sum_{x_1,x_2,x_3\ge 0, x_1+x_2+x_3=n} \sum_{i=1}^{3} \frac{4(x_i+1)^2}{(n+1)^2(n+2)^2(n+3)^2}$$
$$\le \sum_{n=0}^{\infty} \binom{n+2}{2} \frac{4(n+1)(n+3)}{(n+1)^2(n+2)^2(n+3)^2},$$

since there are $\binom{n+2}{2}$ points (x_1, x_2, x_3) in $\mathbb{Z}^3$ with $x_1+x_2+x_3 = n$, and if $x_i \ge 0$ and $x_1 + x_2 + x_3 = n$ then

$$(x_1+1)^2 + (x_2+1)^2 + (x_3+1)^2 \le (n+1)^2 + 1 + 1 \le (n+1)(n+3).$$

Consequently, the total energy is at most

$$\sum_{n=0}^{\infty} \frac{2}{(n+2)(n+3)} = 1.$$

Hence, by Theorem 13, the SRW on $\mathbb{Z}^3$ is transient. □

Clearly, this proof of Pólya's theorem did not really test the power of Theorem 13, which can be used to prove the transience or recurrence of random walks on much more general infinite graphs than lattices.

IX.3 Hitting Times and Commute Times

The results in the previous section were obtained by appealing to only the most rudimentary facts concerning random walks. As we wish to keep our presentation essentially self-contained, we shall continue in this vein; nevertheless, we shall find it convenient to use some basic properties of random walks.

Our aim in this section is to study the important parameters of random walks on graphs, like the expected hitting times, commuting times and sojourn times. In addition to combinatorial arguments, we shall make use of the connection between random walks and electrical networks, to the benefit of both theories. Among other results, we shall prove Foster's theorem that the sum of effective resistances across edges in a simple graph depends only on the order of the graph.

An attractive feature of this theory is that there are a great many interconnections: the order of the results in our presentation is just one of many possibilities.

In order to emphasize the combinatorial nature of the results and to keep the notation simple, we shall consider the simple random walk on a fixed graph G with n vertices and m edges, so that the transition probability matrix is $P = (P_{xy})$, where $P_{xy} = 1/d(x)$ if $xy \in E(G)$. In fact, by doing this, we lose no generality: all the results can be translated instantly to the case of general conductances.

Given an initial probability distribution $\mathbf{p} = (p_x)_{x \in V(G)}$, the probability distribution after one step is $\mathbf{p}P = (\sum_x p_x P_{xy})_{y \in V(G)}$, since y gets a 'mass' or 'probability' $p_x P_{xy}$ from each vertex x. If G is a multigraph then $P_{xy} = m(xy)/d(x)$, where $m(xy)$ is the number of edges from x to y. In this case in one step each *edge* e from x to y carries $1/d(x)$ proportion of the probability p_x at x to y, and $1/d(y)$ proportion of the probability p_y at y to x. For notational simplicity, we shall restrict our attention to simple graphs without loops, although it is clear that all the results carry over to multigraphs and, a little more generally, to weighted multigraphs, as in §2.

Write $\boldsymbol{\pi} = (\pi_x)_{x \in V}$ for the probability distribution on $V = V(G)$ with $\pi_x = d(x)/2m$. If π is our initial probability distribution, then each edge xy transmits $1/d(x)$ of the probability $\pi_x = d(x)/2m$ at x to y, i.e., each edge transmits the same probability $1/2m$ in either direction. In particular, the matrix P and the vector $\boldsymbol{\pi}$ satisfy the *detailed balance equations*

$$\pi_x P_{xy} = \pi_y P_{yx}$$

for all $x, y \in V$. From this it is clear that $\boldsymbol{\pi}$ is a *stationary distribution* for P:

$$\pi P = \pi.$$

Indeed,

$$(\boldsymbol{\pi} P)_y = \sum_x \pi_x P_{xy} = \sum_x \pi_y P_{yx} = \pi_y \sum_x P_{yx} = \pi_y.$$

What happens if our SRW starts from a probability distribution $\mathbf{p} = (p_x)_{x \in V}$? If G is bipartite then $\mathbf{p}P^k$, the distribution after k steps, need not tend to $\boldsymbol{\pi}$ as $k \to \infty$, but if G is not bipartite then it is easily seen that $\mathbf{p}^k = \mathbf{p}P^k$ does tend to $\boldsymbol{\pi}$ (see Exercises 14–17). In particular, for every fixed i, x and y,

$$\mathbf{P}(X_j = x \mid X_i = y) \to \pi_x = \frac{d(x)}{2m} \tag{15}$$

as $j \to \infty$. This implies that if $\varepsilon > 0$ and $|j - i|$ is sufficiently large then

$$|\mathbb{P}(X_i = x,\ X_j = y) - \mathbb{P}(X_i = x)\mathbb{P}(X_j = y)| < \varepsilon. \tag{16}$$

In what follows, we shall start our SRW from a probability distribution $\mathbf{p} = (p_x)_{x \in V}$, so that $\mathbb{P}(X_0 = x) = p_x$, and write $\mathbf{p}^k = (p_x^{(k)})_{x \in V}$, so that

$$p_x^{(k)} = \mathbb{P}(X_k = x) = (\mathbf{p}P^k)_x.$$

Let $S_k(x)$ be the number of times we visit x during the first k steps, that is, the number of times x occurs in the sequence $X_1, \ldots, X_k$.

Theorem 15 *We have* $\lim_{k \to \infty} \mathbb{E}(S_k(x)/k) = d(x)/2m$, *and* $(S_k(x)/k)_{x \in V}$ *tends to* π *in probability as* $k \to \infty$.

Proof. Note first that

$$\mathbb{E}(S_k(x)) = \sum_{i=1}^{k} \mathbb{P}(X_i = x),$$

so

$$\lim_{k \to \infty} \mathbb{E}(S_k(x)/k) = \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} p_x^{(i)} = \frac{d(x)}{2m}. \tag{17}$$

In order to estimate the variance of $S_k(x)/k$, note that, very crudely, if (16) holds for $|j - i| \geq k_0$ then

$$\begin{aligned}
\sigma^2(S_k(x)) &= \mathbb{E}(S_k(x))^2 - (\mathbb{E}S_k(x))^2 \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} (\mathbb{P}(X_i = x, X_j = x) - \mathbb{P}(X_i = x)\mathbb{P}(X_j = x)) \\
&= \sum_{\substack{|i-j|<k_0 \\ i,j \leq k}} (\mathbb{P}(X_i = x, X_j = x) - \mathbb{P}(X_i = x)\mathbb{P}(X_j = x)) \\
&\quad + \sum_{\substack{|i-j| \geq k_0 \\ i,j \leq k}} (\mathbb{P}(X_i = x, X_j = x) - \mathbf{P}(X_i = x)\mathbb{P}(X_j = x)) \\
&\leq 2k_0 k + k^2 \varepsilon.
\end{aligned} \tag{18}$$

Hence if $k \geq 2k_0/\varepsilon$ then this gives

$$\sigma^2(S_k(x)/k) = \frac{\mathbb{E}(S_k(x))^2 - (\mathbb{E}S_k(x))^2}{k^2} \leq \frac{2k_0}{k} + \varepsilon \leq 2\varepsilon.$$

Therefore

$$\mathbb{P}\left(\left|\frac{S_k(x)}{k} - \frac{\mathbb{E}(S_k(x)}{k})\right| \geq \eta\right) \to 0$$

for every $\eta > 0$ so, by (17), $S_k(x)/k \to d(x)/2m$ in probability. $\square$

Denote by $H(x, y)$ the *mean hitting time* of y from x, namely the expected time it takes to go from x to y: $H(x, y) = \mathbb{E}_x(\tau_{\{y\}}^+)$. Clearly,

$$H(x, y) = \sum_{k=1}^{\infty} k\mathbb{P}(X_k = y, X_i \neq y \text{ for } 1 \leq i < k \mid X_0 = x).$$

Putting $x = y$ in this formula, we see that $H(x,x) = \mathbb{E}_x(\tau^+_{\{x\}})$ is the *mean return time* to x. Starting at x, with probability $P_{xy} = 1/d(x)$ the first step takes us to y, so

$$H(x,x) = 1 + \sum_{z\in V} P_{xz}H(z,x) = 1 + \frac{1}{d(x)} \sum_{z\in\Gamma(x)} H(z,x). \tag{19}$$

Occasionally, $H(x, y)$ is also called the *hitting time of* y *from* x or the *access time of* y *from* x.

The function $H'(x,y) = \mathbb{E}_x(\tau_{\{y\}})$ is almost as natural as $H(x,y)$. Clearly, $H'(x, y) = H(x, y)$ for $x \neq y$, but if $x = y$ then H' tells us nothing: $H'(x, x) = 0$. In fact, $H'(x, y)$ is a rather useful tool in calculating $H(x, y)$ since, arguing as in (19),

$$H(x, y) = 1 + \frac{1}{d(x)} \sum_{z\in\Gamma(x)} H'(z, y) \tag{20}$$

for all x, y.

There is no reason to expect $H(x, y)$ to be symmetric and, indeed, it is not. For example, if G is a path xyz then $H(x, y) = 1$ and $H(y, x) = 3$. However, the next result, about $H(x, x)$, holds no surprises.

Theorem 16 *The mean return time to a vertex x in a connected graph is $H(x,x) = 2m/d(x)$.*

Proof. Set $Y_0 = 0$ and let Y_ℓ be the time our random walk $(X_i)_0^\infty$ returns to x for the ℓth time when started at $X_0 = x$. Then $Y_1 = Y_1 - Y_0, Y_2 - Y_1, Y_3 - Y_2, \ldots$ are i.i.d. random variables, so $\mathbb{E}(Y_\ell) = \ell\mathbb{E}(Y_1) = \ell H(x,x)$. Also, $Y_\ell \le k$ if and only if $S_k(x) \ge \ell$. Hence, for $\alpha > 0$,

$$Y_\ell/\ell \le \alpha \quad \text{if and only if} \quad S_{\lfloor \ell\alpha \rfloor} \ge \ell.$$

In particular, $\mathbb{P}(S_{\lfloor \ell\alpha \rfloor}/\ell\alpha \ge 1/\alpha) = \mathbb{P}(Y_\ell/\ell \le \alpha)$ so, by Theorem 15,

$$\lim_{\ell\to\infty} \mathbb{P}\left(\frac{Y_\ell}{\ell} \le \alpha\right) = \lim_{\ell\to\infty} \mathbb{P}\left(\frac{S_{\lfloor \ell\alpha \rfloor}}{\ell\alpha} \ge \frac{1}{\alpha}\right) = \begin{cases} 1 & \text{if } \alpha > 2m/d(x), \\ 0 & \text{if } \alpha < 2m/d(x). \end{cases}$$

Hence Y_ℓ/ℓ tends to $2m/d(x)$ in probability, so $H(x, x) = 2m/d(x)$. □

In fact, more is true: not only is the mean return time to x exactly $2m/d(x)$, but we expect to return to x through *each edge* yx in $2m$ steps.

Theorem 17 *Let xy be a fixed edge of our graph G. The expected time it takes for the simple random walk on G, started at x, to return to x through yx is $2m$. Thus if $X_0, X_1, X_2, \ldots$ is our SRW, with $X_0 = x$, and $Z = \min\{k \ge 2 : X_{k-1} = y, X_k = x\}$, then $\mathbb{E}(Z) = 2m$.*

Proof. The probability that we pass through yx at time $k+1$ is

$$\mathbb{P}(X_k = y, X_{k+1} = x) = \frac{\mathbb{P}(X_k = y)}{d(y)}.$$

Therefore, writing $S_k(yx)$ for the number of times we pass through yx up to time $k+1$,

$$\frac{\mathbb{E}S_k(yx)}{k} = \frac{\mathbb{E}S_k(y)}{kd(y)} \to \frac{1}{2m}.$$

The proof can be completed as in Theorem 16: writing Z_ℓ for the time k our random walk $(X_i)_0^\infty$, started at $X_0 = x$, returns to x for the ℓth time, i.e. $X_{k-1} = y$ and $X_k = x$ for the ℓth time, $\mathbb{E}(Z_\ell) = \ell\mathbb{E}(Z_1)$, and $S_k(xy) \le \ell$ if and only if $Z_\ell \le k$. □

As a slight variant of Theorem 17, it is easily seen that, no matter where we start our SRW and what oriented edge uv we take, $S_k(uv)/k \to 1/2m$ in probability. Loosely speaking, this means that our SRW spends equal amounts of time in each edge, going in either direction. This is far from surprising: in the long run, $\pi_u = d(u)/2m$ of the time we are at u, no matter where we start, and from u with probability $1/d(u)$ we traverse the edge uv.

Since $d(x)/2m$ is just the x coordinate of the stationary distribution $\boldsymbol{\pi}$, by Theorem 16 we have

$$H(x,x) = \frac{1}{\pi_x}. \tag{21}$$

In fact, (21) holds in a considerably more general form, for any ergodic finite Markov chain $(X_k)_0^\infty$. Let $\boldsymbol{\pi} = (\pi_x)_{x\in V}$ be a stationary distribution, so that if X_0 has distribution $\boldsymbol{\pi}$ then each X_k has distribution $\boldsymbol{\pi}$. For $S \subset V$, define $\boldsymbol{\pi}_S$ by $\pi_S(x) = \pi_x/\pi(S)$; let π_S be essentially the conditional probability on S, so that $\pi_S(x) = \pi(x)/\pi(S)$ for $x \in S$; and for $y \in V \setminus S$ set $\pi_S(y) = 0$. In other words, $\boldsymbol{\pi}_S$ is $\boldsymbol{\pi}$ conditioned on the Markov chain being in S. Then we have *Kac's formula*:

$$\pi(S)\mathbb{E}_{\pi_S}(\tau_S^+) = 1, \tag{22}$$

where $\mathbb{E}_{\pi_S}$ denotes the expectation when our chain is started from the initial distribution $\boldsymbol{\pi}_S$. It is easy to check that the proof of Theorem 16 can be repeated to give this more general assertion. Here is another way of proving (22). Note that

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\pi}}(\tau_S^+ = k) &= \mathbb{P}_{\boldsymbol{\pi}}(X_1 \notin S, \dots, X_{k-1} \notin S, X_k \in S) \\
&= \mathbb{P}_{\boldsymbol{\pi}}(X_1 \notin S, \dots, X_{k-1} \notin S) - \mathbb{P}_{\boldsymbol{\pi}}(X_1 \notin S, \dots, X_k \notin S) \\
&= \mathbb{P}_{\boldsymbol{\pi}}(X_1 \notin S, \dots, X_{k-1} \notin S) - \mathbb{P}_{\boldsymbol{\pi}}(X_0 \notin S, \dots, X_{k-1} \notin S) \\
&= \mathbb{P}_{\boldsymbol{\pi}}(X_0 \in S, X_1 \notin S, \dots, X_{k-1} \notin S) \\
&= \pi(S)\mathbb{P}_{\pi_S}(\tau_S^+ \ge k).
\end{aligned}$$

Summing over k, we get

$$1 = \sum_{k=1}^{\infty} \mathbb{P}_\pi(\tau_S^+ = k) = \pi(S) \sum_{k=1}^{\infty} \mathbb{P}_{\boldsymbol{\pi}_S}(\tau_S^+ \ge k) = \pi(S)\mathbb{E}_{\boldsymbol{\pi}_S}(\tau_S^+),$$

as claimed by (22).

For a regular graph G of degree d, combining (21) and Theorem 16, we find that for a fixed vertex x the average of $H(x, y)$ over $y \in \Gamma(x)$ is precisely $n-1$:

$$\frac{1}{d}\sum_{y\in\Gamma(x)} H(y,x) = H(x,x) - 1 = \frac{2m}{d-1} = n-1.$$

A similar result holds without assuming regularity, but we have to take the average of the hitting times $H(x, y)$ over all $2m$ ordered pairs of adjacent vertices. Putting it another way, starting from the stationary distribution, we expect to return to the original position in n steps.

Theorem 18 *Let G be a connected graph of order n and size m. The mean hitting times $H(x, y)$ of the SRW on G satisfy*

$$\frac{1}{2m}\sum_{x\in V(G)}\sum_{y\in\Gamma(x)} H(x,y) = n-1. \tag{23}$$

Proof. Let $\pi = (\pi_x)$ be the stationary distribution for the transition matrix $P = (P_{xy})$, so that $\pi P = \pi$ and $\pi_x P_{xy} = 1/2m$ for $xy \in E(G)$. Then

$$\begin{aligned}\frac{1}{2m}\sum_{x}\sum_{y\in\Gamma(x)} H(x,y) &= \sum_{x,y}\pi_x P_{xy} H(y,x)\\ &= \sum_x \pi_x\left(\sum_y P_{xy}H(y,x)\right) = \sum_x \pi_x(H(x,x)-1)\\ &= \sum_x \pi_x\left(\frac{1}{\pi_x}-1\right) = n-1.\end{aligned}$$

□

Theorem 18 has an attractive reformulation in terms of another invariant of random walks, the mean commute time. For vertices $x \neq y$, the *mean commute time* between x and y, denoted by $C(x, y)$, is the expected number of steps in a *round-trip*, in a walk from x to y and then back to x. Thus

$$C(x,y) = H(x,y) + H(y,x).$$

Then (23) is equivalent to the following:

$$\frac{1}{2m}\sum_{xy\in E(G)} C(x,y) = n-1. \tag{24}$$

Thus the average of the commute times between the m pairs of adjacent vertices in a connected graph of order n is $2(n-1)$.

Theorem 19 *With the notation above,*

$$P_{\text{esc}}(s\to t) = \frac{C_{\text{eff}}(s,t)}{d(s)} = \frac{H(s,s)}{C(s,t)} = \frac{2m}{d(s)C(s,t)}. \tag{25}$$

Furthermore,

$$C(s,t) = 2mR_{\text{eff}}(s,t). \tag{26}$$

Proof. The first equality in (25) follows from relation (13) in Theorem 8. To see the other equalities in (25), let R be the *first* time the random walk returns to s, and let A be the *first* time it returns to s *after* having visited t. Then $\mathbb{E}(R) = H(s,s) = 2m/d(s)$ and, by definition, $\mathbb{E}(A) = C(s,t)$. We always have $R \leq A$ and

$$\mathbb{P}(R = A) = P_{\text{esc}}(s \to t) = q,$$

say. Also,

$$\mathbb{E}(A - R) = (1 - q)\mathbb{E}(A),$$

so

$$C(s,t) = \mathbb{E}(A) = \frac{\mathbb{E}(R)}{q} = \frac{2m}{d(s)q}.$$

Thus $P_{\text{esc}}(s \to t) = 2m/d(s)C(s,t)$, as claimed. As $H(s,s) = 2m/d(s)$, equality (25) is proved. Finally, (26) is immediate from (25). □

The results of Section 2 also have attractive formulations for our simple random walks. For example, the expected number of times we traverse a fixed edge sx from s to x if we start our random walk at s and stop it when we get to t is just $1/d(s)$ times the expected sojourn time at s, $S_s(s \to t)$. But by (14) and (11) (or (25)) this is exactly

$$\frac{S_s(s \to t)}{d(s)} = \frac{1}{d(s)P_{\text{esc}}(s \to t)} = R_{\text{eff}}(s,t). \tag{27}$$

Let us illustrate these results on the simple graph G_0 on three vertices shown in Fig. IX.1. First let us calculate the data for the simple electrical network G_0. The effective resistance of G_0 between s and t is $R_{\text{eff}}(s,t) = \frac{1}{k} + \frac{1}{\ell} = \frac{k+\ell}{kl}$, so setting s at absolute potential $V_s = 1$, and t at $V_t = 0$, we get a current of size $k\ell/(k+\ell)$. In an edge incident with s, there is a current of size $\ell/(k+\ell)$, and in an edge incident with t there is a current of size $k/(k+\ell)$. Also, u is at absolute potential $V_u = k/(k+\ell)$.

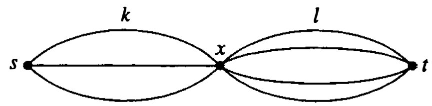


FIGURE IX.1. The graph G_0 has three vertices, s, t and x, and $k+\ell$ edges: k from s to x, and ℓ from x to t.

And now for the simple random walk on the graph G_0. The probability that starting at u we get to s before t is simply that the first step from u is to s: this has probability $k/(k+\ell)$, which is just V_u. The escape probability $P_{\text{esc}}(s \to t)$ is just the probability that after the first step, which takes us to u, we go to t rather than s: this has probability $\ell/(k+\ell)$. Hence $C_s P_{\text{esc}}(s \to t) = k\ell/(k+\ell)$, which is indeed the total current. Also, $1/(d(s)P_{\text{esc}}(s \to t)) = (k+\ell)/k\ell$, which is just the effective resistance between s and t, as claimed by (27).

In the SRW on G_0, the expected sojourn time at s on our way from s to t is

$$\frac{k\ell}{k+\ell}\sum_{j=1}^{\infty}\frac{jk^{j-1}}{(k+\ell)^{j-1}}=\frac{k+\ell}{\ell},$$

so the expected number of times we cross an edge incident with s is $(k+\ell)/k\ell$, which is precisely $R_{\mathrm{eff}}(s,t)$, as we know that it has to be.

What is the expected number of times our SRW crosses a fixed edge *from u to t*? If we start at s and stop at t then this expectation is $1/\ell$, as we know, and if we start at t and stop at s then it is

$$\frac{k}{\ell(k+\ell)}\sum_{j=1}^{\infty}\frac{j\ell^{j}}{(k+\ell)^{j}}=\frac{1}{k}.$$

Hence $R_{\mathrm{eff}}(s,t)$ is the sum of these two expectations. In fact, this holds in general, not only for our simple example G_0 in Fig. IX.1.

To show this, we first set the scene. Let s,t,x be vertices of a connected graph G, with $s\neq t$. Write $S_x(s\leftrightarrow t)$ for the *expected sojourn time* at x in a random round-trip from s to t. In other words, $S_x(s\leftrightarrow t)$ is the expected number of times the SRW on G is at x if we start at s, continue till we get to t, and then stop when we are next at s. Also, for an edge xy, let $S_{xy}(s\leftrightarrow t)$ be the expected number of times we traverse the edge xy from x to y during a round-trip from s to t. Clearly,

$$S_{xy}(s\leftrightarrow t)=\frac{1}{d(x)}S_x(s\leftrightarrow t).$$

Theorem 20 *For a connected graph G, vertices $s\neq t$, and edge $xy\in E(G)$ we have*

$$R_{\mathrm{eff}}(s,t)=S_{xy}(s\leftrightarrow t)=\frac{S_x(s\to t)}{d(x)}+\frac{S_x(t\to s)}{d(x)}.$$

Proof. With the notation in Theorem 9, $S_x(s\leftrightarrow t)$ is

$$S_x(s\to t)+S_x(t\to s)=V_x(s\to t)d(x)+V_x(t\to s)d(x).$$

But

$$V_z(s\to t)+V_z(t\to s)=R_{\mathrm{eff}}(s,t)=R_{\mathrm{eff}}(t,s)\tag{28}$$

for all z. Indeed, $V_z(s,t)$ is the potential of z if s is set at $R_{\mathrm{eff}}(s,t)$ and t at 0, and $V_z(t,s)$ is the potential of z if t is set at $R_{\mathrm{eff}}(s,t)$ and s is set at 0. Hence, (28) holds by the principle of superposition discussed in Section II.1. □

Theorem 20 can be used to give another proof of Theorem 19. Let us restate Theorem 19 as follows:

$$C(s,t)=2mR_{\mathrm{eff}}(s,t)=\frac{2m}{d(s)P_{\mathrm{esc}}(s\to t)}=\frac{H(s,s)}{P_{\mathrm{esc}}(s\to t)}.\tag{29}$$

Now, at a round-trip from s to t and back to s, in each step we traverse an edge until we stop the walk at t, so

$$C(s,t) = \sum_{x \in V(G)} \sum_{y \in \Gamma(x)} S_{xy}(s \leftrightarrow t).$$

By Theorem 10, each of the $2m$ summands on the right-hand side is $R_{\text{eff}}(s,t)$, so the first equality of (29) follows. The second and third equalities are immediate from (11) and (25).

There is another variant of the escape probability that is interest to us. Given vertices s and u, denote by $P_{\text{esc}}(s \to t < u)$ the probability that if we start at s then we get to t before either s or u. The following result is a mild extension of Theorem 8. To prove it, add a vertex t' to the graph and join t' to t and u by many edges. Apply Theorems 8 and 9 to measure the current from s to t'.

Theorem 21 *Let s, t and u be distinct vertices of a graph. Set s at potential* 1, *and t and u at potential* 0. *Then a current of size*

$$d(s)P_{\text{esc}}(s \to t < u)$$

leaves G at t.

The following reciprocity law is obvious if we believe in physical intuition. As we shall see, it also follows easily from the reversibility of the random walk on a graph.

Theorem 22 *Let s, t and u be distinct vertices of a graph G. Then*

$$d(s)P_{\text{esc}}(s \to t < u) = d(t)P_{\text{esc}}(t \to s < u).$$

Proof. Let $W_{s,t;u}$ be the set of walks $W = x_0x_1 \cdots x_\ell$ in $G - u$ such that $x_i = s$ iff $i = 0$ and $x_i = t$ iff $i = \ell$. Then, writing $(X_i)_0^\infty$ for our random walk,

$$P_{\text{esc}}(s \to t < u) = \sum_{W \in W_{s,t;u}} \mathbb{P}(X_i = x_i, 1 \le i \le \ell | X_0 = s)$$

and

$$P_{\text{esc}}(t \to s < u) = \sum_{W \in W_{s,t;u}} \mathbb{P}(X_i = x_{\ell-i},\ 1 \le i \le \ell | X_0 = t).$$

But for $W \in W_{s,t;u}$ we have

$$\mathbb{P}(X_i = x_i, 1 \le i \le \ell | X_0 = s) = \prod_{i=0}^{\ell-1} d(x_i)^{-1}$$

and

$$\mathbb{P}(X_i = x_{\ell-i}, 1 \le i \le \ell | X_0 = t) = \prod_{i=1}^{\ell} d(x_i)^{-1}.$$

The ratio of these two qualities is $d(t)/d(s)$, so the assertion follows. □

Theorem 22 is precisely the result needed to prove the fundamental result in the theory of electrical networks that every network with attachment set U is equivalent to a network with vertex set U (see Exercises 10–12 in Chapter II).

A slight variant of the proof of Theorem 22 gives an important property of mean hitting times: although $H(s,t)$ need not equal $H(t,s)$, 'taking them in triples', we do get equality.

Theorem 23 *Let s, t and u be vertices of a graph G. Then*

$$H(s,t) + H(t,u) + H(u,s) = H(s,u) + H(u,t) + H(t,s).$$

Proof. The left-hand side is the expected time it takes to go from s to t, then on to u and, finally, back to s, and the right-hand side is the expected length of a tour in the opposite direction. Thus, writing τ for the first time a walk starting at s completes a tour $s \to t \to u \to s$, and defining τ' analogously for $s \to u \to t \to s$, the theorem claims exactly that $\mathbb{E}_s(\tau) = \mathbb{E}_s(\tau')$. Consider a closed walk $W = x_0x_1 \cdots x_\ell$ starting and ending at s, so that $x_0 = x_\ell = s$. Clearly,

$$\begin{aligned}\mathbb{P}(X_i = x_i, 1 \le i \le \ell | X_0 = s) &= \mathbb{P}(X_i = x_{\ell-i}, 1 \le i \le \ell | X_0 = s)\\ &= \prod_{i=0}^{\ell-1} d(x_i)^{-1},\end{aligned}$$

that is, the probability of going round this walk one way is precisely the probability of tracing it the other way.

Fix N, and let $S = x_0x_1 \cdots$ be an infinite walk with $x_0 = s$. Set $\ell = \ell(S,N) = \max\{i \le N : x_i = s\}$, and let S' be the walk $x_\ell x_{\ell-1} \cdots x_0 x_{\ell+1} x_{\ell+2} \cdots$. By the observation above, the map $S \mapsto S'$ is a measure preserving transformation of the space of random walks started at s. Since $\tau(S) \le N$ iff $\tau'(S') \le N$, we have $\mathbb{P}_s(\tau \le N) = \mathbb{P}_s(\tau' \le N)$. Hence $\mathbb{E}_s(\tau) = \mathbb{E}_s(\tau')$, as required. □

Theorem 22 implies another form of the reciprocity law.

Theorem 24 *The expected sojourn times satisfy*

$$d(s)S_x(s \to t) = d(x)S_s(x \to t). \tag{30}$$

Proof. Let us define a random walk on the set $\{s,t,x\}$ with transition probabilities $p_{st} = P_{\text{esc}}(s \to t < x)$, $p_{sx} = P_{\text{esc}}(s \to x < t)$, $p_{ss} = 1 - p_{st} - p_{sx}$, and so on. Theorem 22 implies that this new RW is, in fact, also reversible; that is, it can be defined on the weighted triangle on $\{s,t,x\}$, with loops at the vertices. Hence it suffices to check (30) for this RW: we leave this as an exercise (Exercise 20). □

The last theorem we prove is a classical result in the theory of electrical networks: Foster's theorem. If our graph is a tree of order n then the effective resistance across every edge is 1, and the sum of effective resistances across edges is $n-1$. Also, we know from Kirchhoff's theorem in Section II.1 (see Exercise II.4) that if every edge of a connected graph with n vertices and m edges is in the same number of spanning trees then the effective resistance across any edge is precisely $(n-1)/m$, so the sum of effective resistances across edges is again $n-1$. The surprising result that this sum is $n-1$ for every connected graph of order n was proved by Foster in 1949. The two beautiful proofs below, found by Tetali in 1991 and 1994, illustrate the power of the intricate edifice of relations we have constructed.

Theorem 25 *Let G be a connected graph of order n. Then*

$$\sum_{st\in E(G)} R_{\text{eff}}(s,t) = n-1.$$

First Proof. By Theorem 24, for any two vertices t and x we have

$$\sum_{s\in\Gamma(t)} \frac{S_x(s\to t)}{d(x)} = \sum_{s\in\Gamma(t)} \frac{S_s(x\to t)}{d(s)},$$

since the two sums are equal term by term. Now, if $x\neq t$ then the right-hand side is 1, since it is precisely the expected number of times we reach t from one of its neighbours in a random walk from x to t. On the other hand, for $x=t$ the right-hand side is 0. Hence, summing over $V=V(G)$, we find that

$$\sum_{t\in V}\sum_{s\in\Gamma(t)} \frac{S_x(s\to t)}{d(x)} = n-1.$$

But the left-hand side is

$$\sum_{st\in E(G)} \left\{\frac{S_x(s\to t)}{d(x)} + \frac{S_x(t\to s)}{d(x)}\right\} = \sum_{st\in E(G)} R_{\text{eff}}(s,t),$$

with the equality following from Theorem 20. □

Second Proof. By Theorems 18 and 19 (or by relations (23) and (25)),

$$n-1 = \frac{1}{2m}\sum_{st\in E(G)} C(s,t) = \frac{1}{2m}\sum_{st\in E(G)} 2mR_{\text{eff}}(s,t) = \sum_{st\in E(G)} R_{\text{eff}}(s,t). \ \square$$

Numerous other results concerning random walks on graphs are given among the exercises.

IX.4 Conductance and Rapid Mixing

We know that a simple random walk on a non-bipartite connected (multi)-graph converges to the unique stationary distribution, no matter what our initial probability distribution is. Our aim in this section is to study the speed of this convergence: in particular, we shall connect the speed of convergence with a down-to-earth expansion property of the graph.

In order to avoid unnecessary clutter, we shall restrict our attention to regular graphs. We shall adopt the convention that is natural when considering random walks on graphs that a loop contributes *one* (rather than the usual two) to the degree of a vertex. Let us fix a *connected*, *non-bipartite*, *d-regular* (multi)graph G: we shall study random walks on this graph. For notational simplicity, we take $V(G) = [n] = \{1, \dots, n\}$, so that the transition probability matrix P is of the form $P = (P_{ij})_{i,j=1}^n$.

As in Section 3, for a simple random walk $\tilde{X} = (X_t)_0^\infty$ on G, set $p_i^{(t)} = \mathbb{P}(X_t = i)$. Thus $\tilde{X}$ is the SRW with initial distribution $\mathbf{p}_0 = (p_i^{(0)})_{i=1}^n$, and $\mathbf{p}_t = (p_i^{(t)})_{i=1}^n = \mathbf{p}_0 P^t$ is the distribution of X_t. Then $\mathbf{p}_t$ tends to the stationary distribution $\boldsymbol{\pi} = (\frac{1}{n}, \dots, \frac{1}{n})$. In measuring the speed of convergence $\mathbf{p}_t \to \boldsymbol{\pi}$, it does not matter much which norm we take on $C_0(G)$, although it is customary to work with the ℓ_2-norm $\|\mathbf{x}\|_2 = (\sum_{i=1}^n |x_i|^2)^{1/2}$ or the ℓ_1-norm $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$. In particular, the *mixing rate* of the random walks on G is

$$\mu = \sup_{\mathbf{p}_0} \limsup_{t\to\infty} \|\mathbf{p}_t - \pi\|_2^{1/t},$$

where the supremum is taken over all initial distributions $\mathbf{p}_0$. In fact, it is easily seen that the supremum is attained on a great many distributions, and in this definition we may take any norm on $C_0(G)$ instead of the ℓ_2-norm. As we shall see shortly, the mixing rate μ is easily described in terms of the eigenvalues of P.

By definition, $P = A/d$, where $A = (a_{ij})_{i,j=1}^n$ is the adjacency matrix of G, i.e., for $i \neq j$, a_{ij} is the number of edges from i to j, and a_{ii} is the number of loops at i. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of the hermitian matrix P, enumerated with multiplicities. We know from Theorem VIII.5 that $\lambda_1 = 1 > \lambda_2 \geq \cdots \geq \lambda_n > -1$. The space $C_0(G)$ has an orthogonal basis $\mathbf{w}_1 = \boldsymbol{\pi}$, $\mathbf{w}_2, \dots, \mathbf{w}_n$ consisting of eigenvectors of P, so that $\mathbf{w}_i P = \lambda_i \mathbf{w}_i$, $i = 1, \cdots, n$. In particular, the subspace $\boldsymbol{\pi}^\perp$ of $C_0(G)$, orthogonal to $\pi = (\frac{1}{n}, \dots, \frac{1}{n})$ is invariant under P, and the restriction of P to $\boldsymbol{\pi}^\perp$ is a hermitian operator with eigenvalues $\lambda_2 \geq \ldots \geq \lambda_n$ and norm $\lambda = \max\{\lambda_2, |\lambda_n|\}$.

Theorem 26 *The mixing rate μ is precisely* $\lambda = \max\{\lambda_2, |\lambda_n|\}$.

Proof. Given a distribution $\mathbf{p}_0$, set

$$\mathbf{p}_0 = \alpha\boldsymbol{\pi} + \mathbf{p}_0',$$

where

$$\langle \mathbf{p}_0', \pi\rangle = 0.$$

Then $1 = \langle \mathbf{p}_0, n\boldsymbol{\pi}\rangle = \alpha\langle \boldsymbol{\pi}, n\boldsymbol{\pi}\rangle = \alpha$, so

$$\mathbf{p}_0 = \pi + \mathbf{p}_0'.$$

Hence

$$\begin{aligned}\|\mathbf{p}_t - \boldsymbol{\pi}\|_2 &= \|\mathbf{p}_0 P^t - \pi P^t\|_2 = \|(\mathbf{p}_0 - \pi)P^t\|_2 \\ &= \|\mathbf{p}_0' P^t\|_2 \leq \lambda^t \|\mathbf{p}_0'\| \leq \lambda^t.\end{aligned}$$

Therefore, $\mu = \sup_{\mathbf{p}_0} \limsup_{t\to\infty} \|\mathbf{p}_t - \pi\|_2^{1/t} \leq \lambda$.

The converse inequality is just as simple. Assuming that $|\lambda_j| = \lambda$, pick a probability distribution $\mathbf{p}_0$ such that

$$\mathbf{p}_0 = \sum_{i=1}^{n} \xi_i \mathbf{w}_i,$$

with $\xi_j \neq 0$. In fact, we can find such a $\mathbf{p}_0 = (p_i^{(0)})$, even among the distributions such that $p_h^{(0)} = 1$ for some h and $p_i^{(0)} = 0$ for $i \neq h$. But if for our $\mathbf{p}_0$ we have $\xi_j \neq 0$ then

$$\|\mathbf{p}_t - \boldsymbol{\pi}\|_2 = \|(\mathbf{p}_0 - \pi)P^t\|_2 \geq \lambda^t \|\xi_j \mathbf{w}_j\|$$

implies that $\mu \geq \lambda$, as claimed. □

More often than not, it is not easy to determine or estimate $\lambda = \max\{\lambda_2, |\lambda_n|\}$. In fact, the crucial quantity here is λ_2, rather than $|\lambda_n|$. This can be made sure by "slowing down" our RW, as we shall do below. Let then $G = (V, E)$ be a fixed d-regular graph. The *lazy random walk* (LRW) on G with initial random variable X_0 is a Markov chain $\tilde{X} = (X_i)_0^\infty$ such that the random variables X_t take values in $V = V(G) = [n]$, and for $i, j \in V$ we have

$$P(X_{t+1} = j | X_t = i) = \begin{cases} \dfrac{1}{2} & \text{if } i = j, \\ \dfrac{1}{2d} & \text{if } i \sim j, \\ 0 & \text{otherwise.} \end{cases}$$

Putting it slightly differently, we attach d loops to each vertex and run the simple random walk on this multigraph.

Note that if P_S is the transition matrix of the SRW on G then the LRW has transition matrix $P_L = (P_S + I)/2$. In particular, if P_S has eigenvalues $\lambda_1 = 1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ then P_L has eigenvalues $\frac{1}{2}(\lambda_1 + 1) = 1 \geq \frac{1}{2}(\lambda_2 + 1) \geq \cdots \geq \frac{1}{2}(\lambda_n + 1) \geq 0$. Hence the mixing rate of the LRW is $\frac{1}{2}(\lambda_2 + 1)$, the second–largest eigenvalue of P_L. In particular, if λ_2 is close to 1 then the SRW converges at most about twice as fast as the LRW.

In fact, rather than giving an upper bound for $\frac{1}{2}(\lambda_2 + 1)$, we shall estimate the speed of convergence to the stationary distribution in terms of the *conductance* Φ_G of the graph G, defined as

$$\Phi_G = \min_{U \subset V} \frac{e(U, \overline{U})}{d \min\{|U|, |\overline{U}|\}},$$

where $\overline{U} = V \setminus U$ and $e(U, \overline{U})$ is the number of edges from U to $\overline{U}$. Note that if $|U| \leq n/2$, as we may assume, then $d|U| = \sum_{u \in U} d(u)$ is the maximal number of edges that may leave U, so $e(U, \overline{U})/d|U|$ is the proportion of "half-edges" leaving U that go to $\overline{U}$.

Clearly, $0 \leq \Phi_G \leq 1$, but if n is large then Φ_G can hardly be bigger than $1/2$ (see Exercises 26–28). Clearly, $\Phi_G = 0$ if and only if G is disconnected. More generally, when is Φ_G small? If for some set $U \subset V$ there are relatively few $U - \overline{U}$ edges; that is, if there is a "bottleneck" in the graph. It is precisely the existence of such a bottleneck that slows down the convergence $\mathbf{p}_t \to \pi$.

Given an RW $\tilde{X} = (X_t)_0^\infty$ on a graph $G = (V, E)$ with $V = [n]$ as before, we write $p_i^{(t)} = \mathbb{P}(X_t = i)$. Define the *excess probability at vertex i at time t* as $e_{i,t} = p_i^{(t)} - 1/n$ and set

$$d_2(t) = \|\mathbf{p}_t - \pi\|_2^2 = \sum_{i=1}^n e_{i,t}^2.$$

Note that the excess probabilities satisfy

$$\begin{aligned} e_{i,t+1} = p_i^{(t+1)} - \frac{1}{n} &= \left(\frac{1}{2}p_i^{(t)} + \frac{1}{2d}\sum_{j\in\Gamma(i)} p_j^{(t)}\right) - \frac{1}{n} \\ &= \frac{1}{2}\left(p_i^{(t)} - 1n\right) + \frac{1}{2d}\sum_{j\in\Gamma(i)}\left(p_j^{(t)} - \frac{1}{n}\right) \\ &= \frac{1}{2}e_{i,t} + \frac{1}{2d}\sum_{j\in\Gamma(i)} e_{j,t} = \frac{1}{2d}\sum_{j\in\Gamma(i)}\left(e_{i,t} + e_{j,t}\right). \end{aligned} \tag{31}$$

We are ready to state the main result of this section, establishing a close relationship between the conductance of a graph and the speed of convergence of lazy random walks.

Theorem 27 *Let G be a non-trivial regular graph with conductance Φ_G. Then every LRW on G is such that*

$$d_2(t+1) \le \left(1 - \frac{1}{4}\Phi_G^2\right) d_2(t).$$

We shall deduce Theorem 27 from two lemmas that are of interest in their own right.

Lemma 28 *Let G be a d-regular graph with $d \ge 1$. Then, with the notation as above,*

$$d_2(t+1) \le d_2(t) - \frac{1}{2d}\sum_{ij\in E}(e_{i,t} - e_{j,t})^2.$$

Proof. By relation (31),

$$d_2(t+1) = \frac{1}{4d^2}\sum_{i=1}^n \left\{\sum_{j\in\Gamma(i)}(e_{i,t} + e_{j,t})\right\}^2.$$

Since $|\Gamma(i)| = d$ for every i, applying the Cauchy–Schwarz inequality to the inner sum, we find that

$$\begin{aligned} d_2(t+1) &\le \frac{1}{4d^2}\sum_{i=1}^{n}\left\{\sum_{j\in\Gamma(i)}(e_{i,t}+e_{j,t})^2\right\}d \\ &= \frac{1}{2d}\sum_{ij\in E}(e_{i,t}+e_{j,t})^2 = \frac{1}{2d}\sum_{ij\in E}\left\{2\left(e_{i,t}^2+e_{j,t}^2\right)-(e_{i,t}-e_{j,t})^2\right\} \\ &= d_2(t) - \frac{1}{2d}\sum_{ij\in E}(e_{i,t}-e_{j,t})^2, \end{aligned}$$

as claimed. □

The second lemma needs a little more work.

Lemma 29 *Let $G = (V, E)$ be a d-regular graph with conductance Φ_G, and let $x : V \to \mathbb{R}$, $i \mapsto x_i$, be such that $\sum_{i=1}^n x_i = 0$. Then*

$$\sum_{ij\in E}(x_i - x_j)^2 \ge \frac{d}{2}\Phi_G^2\sum_{i=1}^n x_i^2. \tag{32}$$

Proof. Set $m = \lceil n/2\rceil$. We shall prove that if $y_1 \ge y_2 \ge \cdots \ge y_n$, with $y_m = 0$, then

$$\sum_{ij\in E}(y_i - y_j)^2 \ge \frac{d}{2}\Phi_G^2\sum_{i=1}^n y_i^2. \tag{33}$$

It is easily seen that this inequality is stronger than (32). Indeed, in (32) we may assume that $x_1 \ge x_2 \ge \cdots \ge x_n$. Setting $y_i = x_i - x_m$, inequality (33) gives

$$\begin{aligned} \sum_{ij\in E}(x_i - x_j)^2 = \sum_{ij\in E}(y_i - y_j)^2 &\ge \frac{d}{2}\Phi_G^2\sum_{i=1}^n (x_i - x_m)^2 \\ &\ge \frac{d}{2}\Phi_G^2\sum_{i=1}^n x_i^2 + \frac{nd}{2}\Phi_G^2 x_m^2, \end{aligned}$$

since $\sum_{i=1}^n x_i = 0$.

Now, in order to prove (33), set

$$u_i = \begin{cases} y_i & \text{if } i \le m, \\ 0 & \text{if } i > m, \end{cases}$$

$$v_i = \begin{cases} 0 & \text{if } i \le m, \\ y_i & \text{if } i > m. \end{cases}$$

Thus $y_i = u_i + v_i$ for every i. Also, if $u_i \ne 0$ then $u_i > 0$ and $i < m$, and if $v_i \ne 0$ then $v_i < 0$ and $i > m$. Since $(y_i - y_j)^2 = (u_i - u_j + v_i - v_j)^2 \ge$

$(u_i - u_j)^2 + (v_i - v_j)^2$ for every edge ij, it suffices to prove that

$$\sum_{ij \in E} (u_i - u_j)^2 \geq \frac{d}{2} \Phi_G^2 \sum_{i=1}^{m} u_i^2 \tag{34}$$

and

$$\sum_{ij \in E} (v_i - v_j)^2 \geq \frac{d}{2} \Phi_G^2 \sum_{i=m}^{n} v_i^2. \tag{35}$$

Furthermore, as $m \geq n - m$, it suffices to prove (34). In our proof of (34) we may assume that $u_1 > 0$. By the Cauchy–Schwarz inequality,

$$\begin{aligned}
\left\{ \sum_{ij \in E} \left(u_i^2 - u_j^2 \right) \right\}^2 &= \left\{ \sum_{ij \in E} (u_i - u_j)(u_i + u_j) \right\}^2 \\
&\leq \sum_{ij \in E} (u_i - u_j)^2 \sum_{k\ell \in E} (u_k + u_\ell)^2 \\
&\leq \sum_{ij \in E} (u_i - u_j)^2 \sum_{k\ell \in E} 2 \left(u_k^2 + u_\ell^2 \right) \\
&= 2d \sum_{k=1}^{n} u_k^2 \sum_{ij \in E} (u_i - u_j)^2.
\end{aligned} \tag{36}$$

In what follows, our summations are over all edges $ij \in E$ with $i < j$. Clearly,

$$\sum_{ij \in E} \left(u_i^2 - u_j^2 \right) = \sum_{ij \in E} \sum_{\ell=i}^{j-1} \left(u_\ell^2 - u_{\ell+1}^2 \right) = \sum_{\ell=1}^{n-1} \left(u_\ell^2 - u_{\ell+1}^2 \right) e(U_\ell, \overline{U}_\ell),$$

where $U_\ell = [\ell]$ and $\overline{U}_\ell = V - U_\ell = [n] - [\ell]$. Since $u_m = u_{m+1} = \cdots = u_n = 0$, this gives

$$\begin{aligned}
\sum_{ij \in E} \left(u_i^2 - u_j^2 \right) &= \sum_{\ell=1}^{m-1} \left(u_\ell^2 - u_{\ell+1}^2 \right) e(U_\ell, \overline{U}_\ell) \geq \sum_{\ell=1}^{m-1} \left(u_\ell^2 - u_{\ell+1}^2 \right) d\Phi_G \ell \\
&= d\Phi_G \sum_{\ell=1}^{m-1} u_\ell^2 = d\Phi_G \sum_{\ell=1}^{n} u_\ell^2.
\end{aligned} \tag{37}$$

Inequalities (36) and (37) give

$$\sum_{ij \in E} (u_i - u_j)^2 \geq \left\{ d\Phi_G \sum_{i=1}^{n} u_i^2 \right\}^2 \Big/ \left\{ 2d \sum_{i=1}^{n} u_i^2 \right\} = \frac{d}{2} \Phi_G^2 \sum_{i=1}^{n} u_i^2,$$

as desired. □

Armed with these two lemmas, it is easy to prove Theorem 27.

Proof of Theorem 27. By Lemma 28,

$$d_2(t) - d_2(t+1) \geq \frac{1}{2d} \sum_{ij \in E} (e_{i,t} - e_{j,t})^2.$$

Applying Lemma 29 with $x_i = e_{i,t}$, we find that

$$d_2(t) - d_2(t+1) \geq \frac{1}{4} \Phi_G^2 \sum_{i=1}^{n} e_{i,t}^2 = \frac{1}{4} \Phi_G^2 d_2(t),$$

completing the proof. □

The distance of the distribution $\mathbf{p}_t$ at time t from the stationary distribution π is usually measured by the ℓ_1-norm of the difference, which is twice the so-called *total variation distance*: $d_1(t) = 2d_{TV}(\mathbf{p}_t, \pi) = \|\mathbf{p}_t - \pi\|_1$. If we wish to emphasize the dependence of $d_1(t)$ on the RW $\tilde{X} = (X_t)_0^\infty$ then we write $d_1(\tilde{X}, t)$ instead of $d_1(t)$. By the Cauchy–Schwarz inequality,

$$d_1(t) = \sum_{i=1}^{n} |e_{i,t}| \leq \left(\sum_{i=1}^{n} e_{i,t}^2 \right)^{1/2} \left(\sum_{i=1}^{n} 1^2 \right)^{1/2} = (n d_2(t))^{1/2}.$$

Also, trivially, $d_2(0) \leq 2$ for every distribution. Hence Theorem 27 has the following important consequence.

Corollary 30 *Every LRW on a regular graph G of order n and conductance Φ_G is such that*

$$d_1(t) \leq (n d_2(t))^{1/2} \leq (2n)^{1/2} \left(1 - \frac{1}{4}\Phi_G^2\right)^{t/2} \leq (2n)^{1/2} \left(1 - \frac{1}{2}\Phi_G^2\right)^{t}.$$

Furthermore, the mixing rate μ satisfies

$$\mu \leq \left(1 - \frac{1}{4}\Phi_G^2\right)^{1/2} \leq 1 - \frac{1}{2}\Phi_G^2.$$ □

We know that if λ_2 is the second–largest eigenvalue of the SRW on a regular graph then the second–largest eigenvalue of the LRW is $\frac{1}{2}(\lambda_2 + 1)$. Hence Corollary 30 has the following consequence.

Corollary 31 *The second eigenvalue of the SRW on a regular graph with conductance Φ_G is at most $1 - \Phi_G^2$.*

Proof. With the notation as above, $\frac{1}{2}(\lambda_2 + 1) \leq 1 - \frac{1}{2}\Phi_G^2$, so $\lambda_2 \leq 1 - \Phi_G^2$. □

Corollaries 30 and 31 can frequently be used to prove that certain random walks converge very fast to the stationary distribution. Given a sequence $G_1, G_2, \ldots$ of graphs, with $|G_i| = n_i \to \infty$, we say that the lazy random walks on this sequence are *rapidly mixing* random walks if there is a polynomial f, depending only on the sequence (G_i), such that if $0 < \epsilon < 1$ and $t \geq f(\log n_i)\log(1/\epsilon)$ then $d_1(\tilde{X}_i, t) \leq \epsilon$ whenever $\tilde{X}_i$ is a lazy random walk on G_i. Thus, if our random walks are rapidly mixing then we are within ϵ of the stationary distribution after polynomially many steps in the *logarithm* of the order! This is fast indeed: it suffices to take far fewer steps than the order of the graph.

The larger the conductance, the faster convergence is guaranteed but, in fact, fairly small conductance suffices to ensure rapid mixing. Indeed, if G is connected and has order $n \geq 2$, so that $\Phi_G > 0$, and

$$t \geq 8\Phi_G^{-2}\left\{\log(1/\varepsilon) + \frac{1}{2}\log(2n)\right\}$$

for some $0 < \epsilon < 1$, then

$$d_1(t) \leq (2n)^{1/2}\left(1 - \frac{1}{4}\Phi_G^2\right)^{t/2} < \exp\left\{\frac{1}{2}\log(2n) - \frac{1}{8}\Phi_G^2 t\right\} \leq \varepsilon.$$

In particular, if $n \geq 3$ and $t \geq 8\Phi_G^{-2}\log n \log(1/\varepsilon)$ then $d_1(t) < \varepsilon$. This gives us the following sufficient condition for rapid mixing.

Theorem 32 *Let $(G_i)_1^\infty$ be a sequence of regular graphs with $|G_i| = n_i \to \infty$. If there is a $k \in \mathbb{N}$ such that*

$$\Phi_{G_i} \geq (\log n_i)^{-k} \tag{38}$$

for every sufficiently large i, then the lazy random walks on $(G_i)_1^\infty$ are rapidly mixing.

Proof. We have just seen that if $t \geq 8(\log n_i)^{2k+1}\log(1/\epsilon)$ then $d_1(t) < \epsilon$, provided n_i is large enough. □

There are many families of regular graphs for which we can give a good lower bound for the conductance. As a trivial example, take the complete graph K_n. It is immediate that $\Phi_{K_n} > \frac{1}{2}$ for $n \geq 2$ so the lazy random walks on (K_n) are rapidly mixing. Of course, this is very simple from first principles as well.

As a less trivial example, we take the hypercubes or simply cubes $Q^1, Q^2, \ldots$ Here $Q^d = \{0, 1\}^d$ is the d-dimensional cube: its vertex set is the set of all 2^d sequences $x = (x_i)_1^n$, $x_i = 0$ or 1, with two sequences joined by an edge if they differ in only one term. Clearly, Q^d is d-regular, and it is rather easy to prove that $\Phi_{Q^d} = 1/d$. The worst bottlenecks arise between the "top" and "bottom" of Q^n: for $U = \{(x_i) \in Q^d : x_1 = 1\}$ and $\overline{U} = \{(x_i) \in Q^d : x_1 = 0\}$, say. Clearly, $e(U, \overline{U}) = |U| = 2^{d-1}$, so that $\Phi_{Q^n}(U) = 1/d$.

As $\Phi_{Q^d} = 1/d = 1/\log n$, where $n = 2^d = |Q^d|$, the lazy random walks on $(Q^d)_1^\infty$ are rapidly mixing.

The cube Q^d is just $K_2^d = K_2 \times \cdots \times K_2$, that is, the product of d paths of lengths 1. Taking the product of d cycles, each of length ℓ, we get the *torus* T_ℓ^d. This graph has ℓ^d vertices, and it is $2d$-regular. Also, one can show that for $G = T_{2\ell}^d$ we have $\Phi_G = \frac{2}{\ell d}$. (Note that T_4^d is just the cube Q^{2d}.) Hence, for a fixed value of ℓ, the lazy random walks on $(T_{2\ell}^d)_{d=1}^\infty$ are rapidly mixing.

It is easy to extend Theorem 27 to reversible random walks or, what amounts to the same, to simple random walks on general multigraphs. Given a multigraph $G = (V, E)$ define the volume of a set $U \subset V(G)$ to be $\text{vol}U = \sum_{u \in U} d(u)$. As before, in the degree $d(u)$ of u we count 1 for each loop at u. Then the conductance of G is

$$\Phi_G = \min_{U \subset V} \frac{e(U, \overline{U})}{\min\{\mathrm{vol}\, U, \mathrm{vol}\, \overline{U}\}},$$

where again $\overline{U} = V \setminus U$. With this definition of the conductance, it is easy to prove the analogue of Theorem 27.

In conclusion, we remark that rapidly mixing random walks have numerous algorithmic applications. In particular, rapidly mixing random walks are frequently used to generate approximately random elements in large sets that are not easily described, such as the set of perfect matchings or spanning trees in a graph, or the set of lattice points in a convex body. The generation of approximately random elements enables one to enumerate the elements asymptotically. A striking application of rapidly mixing random walks concerns randomized polynomial–time algorithms giving precise estimates for the volume of a convex body in $\mathbb{R}^n$.

IX.5 Exercises

1. Let T_k be the rooted infinite tree with every vertex having k descendents. For what values of k is the SRW on T_k recurrent?
   Is there a subtree of $\mathbb{Z}^3$ on which the SRW is transient?

2. Show that the SRW is recurrent on the hexagonal lattice and on the triangular lattice.
   Show also that if G is a graph whose vertex set is contained in the plane $\mathbb{R}^2$, with any two vertices at distance at least 1 and no edge joining vertices at distance greater than 10^{10}, then the SRW on G is recurrent.

3. Given a real number $r > 0$, construct a locally finite infinite graph G containing a vertex s such that in the network obtained from G by giving each edge resistance 1, we have $R_{\mathrm{eff}}(s, \infty) = r$.

4. Let $N = (G, c)$ be an electrical network on a locally finite infinite graph G. Let $s \in V(G)$ and let $\mathcal{T}$ be the set of subsets T of $V(G)$ with $s \notin \mathcal{T}$ and $V(G) \setminus T$ finite. For $T \in \mathcal{T}$, let N/T be the network obtained from N by fusing all the vertices of T to a single vertex ∞. Write $R_{\mathrm{eff}}^{(T)}(s, \infty)$ for the resistance of N/T between s and ∞. Finally, let $T_1, T_2, \ldots \in \mathcal{T}$ be such that $d(s, T_\ell) = \min\{d_G(s, t) \colon t \in T_\ell\} \to \infty$. Show that $\lim_{\ell \to \infty} R_{\mathrm{eff}}^{(T_\ell)}(s, \infty) = \inf_{T \in \mathcal{T}} R_{\mathrm{eff}}^{(T)}(s, \infty)$.

5. Let $N = (G, c)$ be an electrical network on a locally finite infinite graph G, and let s, $N^{(\ell)}$ and t_ℓ be as after Theorem 10. Show that for every $\ell \geq 1$ there is an $\epsilon_\ell > 0$ such that if $d(s, x) \leq \ell$ then $\mathbb{P}_x(\mathcal{T}_{\{s\}} \leq \ell) \geq \epsilon_\ell$. Deduce from this that $P_{\mathrm{esc}}^{(\infty)} = \lim_{\ell \to \infty} P_{\mathrm{esc}}(s \to t_\ell) = \lim_{\ell \to \infty} P_{\mathrm{esc}}^{(\ell)}$.

6.$^+$ Consider the simple electrical network of the hypercube Q^n with vertex set $\{0, 1\}^n$. Show that $R_{\text{eff}}(s, t)$ is a monotone increasing function of the distance $d(s, t)$.

7. Formalize the remarks at the beginning of Section 2 as follows. Let (G, a) be a connected weighted graph, with weight function $xy \mapsto a_{xy}$, and let $S \subset V(G)$. A function $f : V(G) \to \mathbb{R}$ is said to be *harmonic* on (G, a), with boundary S, if

$$f(x) = \frac{1}{A_x} \sum_{y \sim x} a_{xy} f(y)$$

for every $x \in V(G) \setminus S$, where $A_x = \sum_{y \sim x} a_{xy}$.

(i) Prove the *maximum modulus principle* that the maximum of a non-constant harmonic function is attained on S; also, if $G - S$ is connected then the maximum is attained at some point of $V(G) \setminus S$ if, and only if, the function is constant.

(ii) Prove the *superposition principle* that if f_1 and f_2 are harmonic on (G, a) with boundary S then so is $c_1 f_1 + c_2 f_2$ for any $c_1, c_2 \in \mathbb{R}$.

(iii) Show that for every $g : S \to \mathbb{R}$ there is a unique function $f : V(G) \to \mathbb{R}$ which is harmonic on (G, a), with boundary S, such that $f(x) = g(x)$ for all $x \in S$. [*Hint.* Consider the RW $(X_t)_0^\infty$ on (G, a) and, for $x \in V(G)$, define $f(x) = \mathbb{E}_x(g(X_{\tau_S}))$, where τ_S is the hitting time of the set S of states.]

In the next four exercises, we consider a Markov chain or a random walk $\underset{\sim}{X} = (X_t)_0^\infty$ on a finite state space V with transition probability matrix $P = (P_{xy})_{x,y \in V}$. Thus P is a stochastic matrix: $\sum_y P_{xy} = 1$ for every x.

8. A Markov chain is *ergodic* if for any two states x and y, there is a positive probability of going from x to y in some number of steps. Show that a finite MC is ergodic iff for some $t \geq 1$, every entry of $\sum_{i=1}^t P^i$ is positive. Show also that a weighted graph defines an ergodic RW iff the graph is connected.

9. A probability distribution $\boldsymbol{\pi} = (\pi_x)_{x \in V}$ on V is *stationary* if $\boldsymbol{\pi} P = \boldsymbol{\pi}$, i.e. if $\underset{\sim}{X} = (X_t)_0^\infty$ is started with initial distribution $\boldsymbol{\pi}$ then each X_t has distribution $\boldsymbol{\pi}$. Write $S_x(s \to s)$ for the expected sojourn time in $x \in V$ during a tour from s to s:

$$S_x(s \to s) = \mathbb{E}_s(|\{i < \tau_{\{s\}}^+ : X_i = x\}|),$$

so that $S_s(s \to s) = 1$. Show that, for $y \neq x$,

$$S_y(s \to s) = \sum_{x \in V} P_{xy} S_x(s \to s),$$

and deduce that $(S_x(s \to s))_{x \in V}$ is a positive multiple of a stationary distribution. Show also that if $\underset{\sim}{X}$ is ergodic then the stationary distribution is unique.

10. An ergodic MC is *reversible* if there is a probability distribution $\boldsymbol{\pi} = (\pi_x)_{x\in V}$ on V such that the detailed balance equations are satisfied: $\pi_x P_{xy} = \pi_y P_{yx}$ for all $x, y \in V$. Show that this distribution $\boldsymbol{\pi}$ is stationary.

11. Check that an RW on a weighted connected graph (G, a) is reversible, with the stationary distribution given by $\pi_x = A_x / \sum_{y\in V} A_y$.
Show also that every reversible ergodic MC on V is an RW on an appropriate weighted graph (with loops). [*Hint*. Define G with $V(G) = V$ by joining x to y if $P_{xy} > 0$. For an edge (or loop) xy of G, set $a_{xy} = \pi_x P_{xy} = \pi_y P_{yx}$.]

12. Show that an ergodic MC is reversible if, and only if, $P_{xy} > 0$ implies $P_{yx} > 0$ and, for every sequence $x_1, \dots, x_n, x_{n+1} = x_1$ of states, $p_C = \prod_{i=1}^n P_{x_i x_{i+1}} > 0$ implies that $\prod_{i=1}^n P_{x_i} P_{x_{i-1}} = p_C$.

13. Let $P = (P_{xy})$ be the transition matrix of a finite Markov chain, and let t be a state such that, for every state $x \neq t$, $P_{xt}^{(k)} > 0$ for some $k \geq 1$, where $P^k = (P_{xy}^{(k)})$. Let Q be the matrix obtained from P by deleting the row and column corresponding to t. Prove that $I - Q$ is invertible, and $(I - Q)^{-1} = (N_{xy})$, where $N_{xy} = S_y(x \to t)$, i.e. N_{xy} is the expected sojourn time in y in a chain started in state x and ending in t.

14. Show that the distribution of an SRW on a bipartite graph need not converge.

15. Let G be a connected non-bipartite graph of order n. Show that for any two vertices $x, y \in V(G)$, there is a walk of length $2n - 4$ from x to y. Show also that for $\ell < 2n - 4$ there need not exist a walk of length ℓ from x to y.

16. Let P be the transition matrix of an SRW on a connected non-bipartite graph G of order n and size m. Show that *every* entry of P^{2n-4} is strictly positive. Deduce that for every probability distribution $\mathbf{p}$ on $V(G)$, $\mathbf{p}P^t$ tends to the stationary distribution $\boldsymbol{\pi} = (d(x)/2m)_{x\in V(G)}$.

17. Let G and P be as in the previous exercise, and let $\mathbf{q} = (q(x))_{x\in V(G)}$ be such that $\sum_{x\in V(G)} q(x) = 0$ and $\|\mathbf{q}\|_1 = \sum_{x\in V(G)} |q(x)| = 1$. Show that $\|\mathbf{q}P^k\|_1 < 1$ for $k = n - 2$. Show also that k need not exist with $k \leq n - 3$. Deduce again the result in Exercise 16, and give a bound for the speed of convergence.

18. Prove directly that $P_{\text{esc}}(s \to t) = 1/S_s(s \to t)$.

19. Let P be the transition probability matrix of the RW on a connected electrical network $N = (G, c)$ with n vertices. Show that the stationary distribution $\boldsymbol{\pi} = (\pi_x)$ and the expected hitting times $H(x, y)$ satisfy

$$\sum_{x,y\in V(G)} \pi_x P_{xy} H(y, x) = n - 1.$$

20.[+] Let G be a weighted triangle with vertex set $\{s, t, x\}$, having weighted loops at the vertices. Show that the RW on G satisfies $d(s)S_x(s \to t) = d(x)S_s(x \to t)$.

21.$^{++}$ In a game of *patience*, played on a finite set of counters on $\mathbb{Z}$, there are two legal moves:
(a) if there are two counters on the same integer k, we may move one to $k-1$ and the other to $k+1$,
(b) if there is a counter on k and another on $m > k$ then we may move them to $k+1$ and $m-1$.
Starting with n counters, all on 0, what is the maximal distance between two counters that can be achieved by a sequence of legal moves?

22. Let G be a graph with m edges, and let s, t be distinct vertices at distance $d(s,t)$. Show that $C(s,t) \leq 2md(s,t)$.

23. Let $N = (G, r)$ be a connected electrical network with n vertices and resistance function r_{xy}. Give two proofs of Foster's theorem that

$$\sum_{xy \in E(G)} \frac{R_{\text{eff}}(x,y)}{r_{xy}} = n - 1.$$

24. Deduce from Theorem 23 that for every graph G there is an order on $V(G)$ such that if s precedes t in the order then $H(s,t) \leq H(t,s)$.

25. Deduce from the result in the previous exercise that if the automorphism group of a graph G is vertex-transitive then $H(s,t) = H(t,s)$ for any two vertices s and t.

26.$^{-}$ Show that the only non-trivial graphs of conductance 1 are K_2 and K_3.

27.$^{-}$ Let G be a regular graph of order n. Show that

$$\Phi_G \leq \lfloor n^2/4 \rfloor \Big/ \binom{n}{2} \leq \frac{1}{2} + \frac{1}{2(n-1)}.$$

28.$^{-}$ Show that if G is an incomplete regular graph of order n then $\Phi_G \leq \frac{1}{2}$.

29. Let $\mathbf{p}$ be a probability distribution on $[n]$, and let $\pi = (1/n, \ldots, 1/n)$. Show that

$$\|\mathbf{p} - \pi\|_2^2 \leq 1 - 1/n.$$

30. Let G be a connected regular graph of order n, with stationary distribution $\pi = (1/n, \ldots, 1/n)$. Let $\widetilde{X} = (X_t)_0^\infty$ be the SRW on G with X_t having distribution $\mathbf{p}_t = (p_1^{(t)}, \ldots, p_n^{(t)})$. Show that

$$\limsup_{t\to\infty} \max_i |p_i^{(t)} - \frac{1}{n}|^{1/t}$$

is the modulus of an eigenvalue of the transition matrix. Which one?

31. Let G be a k-edge-connected d-regular of order n. Show that $\Phi_G \geq k/d\lfloor n/2 \rfloor$.

32.$^{+}$ The *lollipop graph* L_k is a clique of order $2k$ to which a path of length k has been attached (see Fig. IX.2). Show that if s and t are vertices of L_k such that

s is the endvertex of the path furthest from the clique and t is not on the path (and so t is one of $2k-1$ vertices of the clique), then

$$H(s,t) = 4k^3 + o(k^3).$$

[In fact, for any two vertices s, t of a connected graph of order n, $H(s,t) \leq 4n^3/27 + o(n^3)$, so the graph above is essentially worst possible for hitting times.]

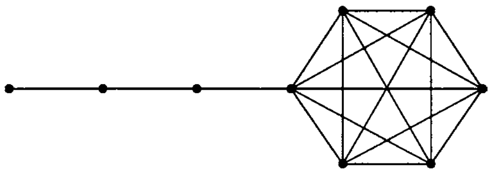


FIGURE IX.2. The lollipop graph L_3.

33.$^+$ Let s be a vertex of a connected graph G with n vertices and m edges. The *mean cover time*, or simply *cover time*, $C(s)$ is the expected number of steps taken by the SRW on G started at s to visit *all* vertices of G. Let T be a spanning tree of G, and set

$$R(T) = \sum_{xy \in E(T)} R_{\text{eff}}(x,y).$$

Show that $C(s) \leq 2mR(T)$. Deduce that $C(s) \leq n(n-1)^2$. [*Hint.* Show that there is an enumeration $x_1, x_2, \dots, x_n$ of the vertices, such that $x_1 = s$ and $d_T(x_i, x_{i+1}) \leq 2$, for $i = 1, \dots, n$, where d_T denotes the distance on T and x_{n+1} is taken to be x_1.]

34. Show that the mixing rate of the SRW on a regular graph of order $n \geq 3$ is at least $1/(n-1)$, with equality only for the complete graph K_n.

35. Show that for $n \geq 2$ the mixing rate of the lazy random walk on the complete graph K_n is $1/2$.

36. Compute the eigenvalues of the adjacency matrix of the $(r-1)n$-regular graph $K_r(n)$, and deduce that for $r \geq 2$ the mixing rate of the SRW on G is $1/(r-1)$.

37. Let G be obtained from K_{2n} by deleting a 1-factor, with $n \geq 2$. Compute the eigenvalues of the adjacency matrix of G and note that the mixing rate of the SRW on G is $1/(n-1)$, over twice the mixing rate of K_{2n}. [*Hint.* Let $V(G) = [2n]$, and for $\mathbf{x} = (x_i)_1^n$ set $\mathbf{x}' = (x_1, \dots, x_n, x_1, \dots, x_n)$, $\mathbf{x}'' = (x_1, \dots, x_n, -x_1, \dots, -x_n) \in C_0(G)$. Note that if $\sum_1^n x_i = 0$ then $\mathbf{x}'$ is an eigenvector with eigenvalue -2 and $\mathbf{x}''$ is an eigenvector with eigenvalue 0.]

38.$^+$ Let G be a connected regular graph of order $n \geq 5$ without loops. Show that $\Phi_G \geq \frac{4}{n(n-4)}$. Show also that if equality is attained then $n = 4k+2$ for some

$k \geq 2$; also, for every $n = 2k + 2$ there is a unique graph for which equality holds (cf. Fig. IX.3).

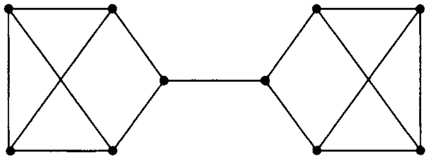


FIGURE IX.3. A cubic graph of order 10 and conductance $1/15$.

39. Consider the SRW on the path on length n, with vertex set $\{0, 1, \ldots, n\}$.
 (i) Note that $H(0, 0) = H(n, n) = 2n$.
 (ii) Deduce that $H(s, s + 1) = 2s + 1$.
 (iii) Show that for $0 \leq s < t \leq n$ we have $H(s, t) = t^2 - s^2$.

40.$^+$ Given a graph G and a vertex $s \in V(G)$, the *mean cover time starting from* s, $C(s)$, is the expected number of steps to visit every vertex if our SRW on G starts at s. Let G be the path of length n on $\{0, 1, \ldots, n\}$, as in the previous exercise, and let $C(s)$ be the mean cover time from s, as in Exercise 33.
 (i) Show that $C(0) = C(n) = n^2$.
 (ii) Check that $C^*(s) = C(s) - s(n - s)$ is harmonic on G and deduce that $C(s) = n^2 + sn - s^2 = 5n^2/4 - (n/2 - s)^2$.

41. Let G be the cycle of length $2n$, and write $H(d)$ for the mean hitting time $H(s, t)$, where s and t are vertices at distance d. Show that $H(n) = H(n - 1) + 1$, $H(n-1) = H(n-2) + 3$, and so on, and deduce that $H(d) = d(2n-d)$ for $0 \leq d \leq n$. Show also that if G is a cycle of length $2n + 1$ and $H(d)$ is as before, then $H(n) = H(n - 1) + 2$, $H(n - 1) = H(n - 2) + 4$, and so on, implying $H(d) = d(2n + 1 - d)$.

42.$^+$ Prove that the mean cover time of a cycle of length n (starting from any vertex) is $n(n - 1)/2$. [*Hint.* The expected number of steps to reach all but one of the vertices is precisely the mean cover time of the cycle of length $n - 1$. To cover the rest, we just have to hit a neighbouring vertex. Make use of the result in the previous exercise.]

43. Let s and t be distinct vertices of a complete graph of order n. Determine $H(s, t)$. [*Hint.* Note that $H(s, t) = 1 + \frac{n-2}{n-1} H(s, t)$.]

44. Prove that the mean cover time of a complete graph of order n (from any vertex) is $(n - 1) \sum_{k=1}^{n-1} 1/k$. [*Hint.* This is just the classical coupon collector's problem. Having covered k vertices, what is the expected number of steps it takes to get to a new vertex?]

45. Let $P = (P_{ij})_{i,j=1}^n$ be a stochastic matrix, i.e., the transition probability matrix of a MC on $[n]$. Set $\epsilon(P) = \min_{i,j} P_{ij}$. Show that if $\mathbf{q} = (q_i)_1^n$ is such that $\sum_1^n q_i = 0$ then $\|\mathbf{q}P\|_1 \leq (1 - n\epsilon)\|\mathbf{q}\|_1$.

46.+ Let T be a finite set of states of an RW, and let s be a state not in T. For $0 \le \alpha < 1$, let $\tau_{\{T;\alpha\}}$ be the first time we have visited more than $\alpha|T|$ of the states in T, when starting from s, and let $H(s;\, T, \alpha)$ be the mean hitting time: $H(s;\, T, \alpha) = \mathbb{E}_s(\tau_{\{T;\alpha\}})$. Prove that

$$H(s;\, T, \alpha) \le \frac{1}{1-\alpha} \max_{t \in T} H(s, t).$$

[*Hint.* Note that, at every point of our probability space, there are at least $(1-\alpha)|T|$ states $t \in T$ such that $\tau_{\{t\}} \ge \tau_{\{T;\alpha\}}$. Deduce that $\mathbb{E}_s(\tau_{\{T;\alpha\}}) \le \frac{1}{(1-\alpha)|T|} \sum_{t \in T} H(s, t)$.]

47. Deduce from the result in Exercise 46 that the main cover time of an RW with n states, started from any state is at most $2 \log_2 n \max_{s,t} H(s, t)$.

48. Let G be the n-dimensional cube Q^n, as in Exercises 40 and 41 in Chapter VIII. By induction on n, show that if $U \subset V(Q^n)$, $|U| \le 2^{n-1}$, then $e(U, \overline{U}) \ge |U|$. Deduce that $\Phi_{Q^n} = 1/n$.

49. Recall from Exercise 41 of Chapter VIII that the second largest eigenvalue of the adjacency matrix of Q^d is $1 - 2/d$. Deduce that the mixing rate of the LRW on Q^d is $1 - 1/d$, considerably better than the rate given by Corollary 30.

50.+ Redefine the LRW on the cube Q^d as follows: in each step chose a coordinate (direction) at random, then change that coordinate or stay still, each with probability $1/2$. By considering the event that all directions have been picked, show that $d_{TV}(\mathbf{p}_t, \pi) \le e^{-c}$ for $t \ge d(\log d + c)$ for every LRW on Q^d, no matter what the initial distribution is.

IX.6 Notes

The connection between random walks and electrical networks was recognized over fifty years ago by S. Kakutani, *Markov processes and the Dirichlet problem*, Proc. Jap. Acad. **21** (1945), 227–233; it seems that C.St.J.A. Nash-Williams was the first to use it with great success, in *Random walks and electric currents in networks*, Proc. Cambridge Phil. Soc. **55** (1959), 181–194. Nevertheless, the subject really took off only in the last two decades, especially after the publication of a beautiful little book by P.G. Doyle and J.L. Snell, Random Walks and Electrical Networks, Carus Math. Monogr., vol. **22**, Mathematical Assoc. of America, Washington, 1984, xiii+159 pp.

For the dawn of the theory of electrical networks, see J.W.S. Rayleigh, On the theory of resonance, in *Collected Scientific Papers*, vol. 1, Cambridge, 1899, pp. 33–75. Pólya's theorem on random walks on lattices is from G. Pólya, Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz, Mathematische Annalen **84** (1921), 149–160, and Foster's theorem is from R.M. Foster, The average impedance of an electric network, in *Contributions to Applied Mechanics* (Reisser Anniv. Vol.), Edwards Bros., Ann Arbor, pp. 333–340. The two beautiful proofs of Foster's theorem are from P. Tetali, Random walks and effective resistance of networks, J. Theoretical Probab. **4** (1991), 101–109, and An extension of Foster's network theorem, Combinatorics, Probability and Computing **3** (1994), 421–427.

The reader is encouraged to consult some of the beautiful papers on random walks on graphs, including R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász, and C. Rackoff, Random walks, universal traversal sequences, and the complexity of maze problems, in *20th Annual Symposium on Foundations of Computer Science*, San Juan, Puerto Rico, October 1979, pp. 218–223, A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky and P. Tiwari, The electrical resistance of a graph captures its commute and cover times, in *Proceedings of the 21st Annual ACM symposium on Theory of Computing*, Seattle, WA, May 1989, pp. 574–586, G. Brightwell and P. Winkler, Maximum hitting times for random walks on graphs, *Random Structures and Algorithms* **1** (1990), 263–276, and D. Coppersmith, P. Tetali, and P. Winkler, Collisions among random walks on a graph, *SIAM J. Disc. Math.* **6** (1993), 363–374. For numerous results presented as exercises, see L. Lovász, Combinatorial Problems and Exercises, Elsevier Science, 1993, 635 pp.

Many results connecting the expansion properties of a graph to the speed of convergence of random walks can be found in N. Alon and V.D. Milman, λ_1, isoperimetric inequalities for graphs and superconcentrators, *J. Combinatorial Theory, Series B* **38** (1985), 73–88, N. Alon, Eigenvalues and expanders, *Combinatorica* (2)**6** (1986), 86–96, D. Aldous, On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing, *Prob. in Eng. and Inf. Sci.* **1** (1987), 33–46, M.R. Jerrum and A.J. Sinclair, Conductance and the rapidly mixing property for Markov chains: The approximation of the permanent resolved, *Proceedings of the 20th Annual Symposium on the Theory of Computing*, 1988, pp. 235–244, Approximate counting, uniform generation and rapidly mixing Markov chains, *Information and Computation* **82** (1989), 93–133, Approximating the permanent, *SIAM J. Computing* **18** (1989), 1149–1178, and M. Mihail, Conductance and convergence of Markov chains—A combinatorial treatment of expanders, *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, 1989. In particular, the notion of conductance was introduced by Jerrum and Sinclair. The presentation in Section 4 is based on B. Bollobás, Volume estimates and rapid mixing, in *Flavors of Geometry* (S. Levy, ed.), Cambridge University Press, 1997, pp. 151-180. For the problem of estimating the volume of a convex body in $\mathbb{R}^n$, see M.E. Dyer, A.M. Frieze and R. Kannan, A random polynomial-time algorithm for approximating the volume of convex bodies, *J. Assoc. Comput. Mach.* **38** (1991), 1–17, M.E. Dyer and A.M. Frieze, On the complexity of computing the volume of a polyhedron, *SIAM J. Computing* **17** (1988), 967–974, and Computing the volume of convex bodies: A case where randomness provably helps, in *Probabilistic Combinatorics and Its Applications*, (B. Bollobás, ed.), Proc. Symp. Applied Math. **44** (1991), pp. 123–169, and L. Lovász and M. Simonovits, Mixing rate of Markov chains, an isoperimetric inequality, and computing the volume, *Proc. 31st Annual Symp. on Found. of Computer Science*, IEEE Computer Soc., 1990, pp. 346–355.

X

The Tutte Polynomial

So far we have encountered several polynomials associated with a graph, including the chromatic polynomial, the characteristic polynomial and the minimal polynomial. Our aim in this chapter is to study a polynomial that gives us much more information about our graph than any of these.

This polynomial, a considerable generalization of the chromatic polynomial, was constructed by Tutte in 1954, building on his work seven years earlier. Although Tutte called this two-variable polynomial $T_G(x, y) = T(G; x, y)$ the *dichromate* of the graph G, by now it has come to be called the *Tutte polynomial* of G.

Similarly to the chromatic polynomial, the Tutte polynomial can be defined recursively by the *cut* and *fuse* operations introduced in Section V.1. The main virtue of the Tutte polynomial is that during the process much less information is lost about the graph than in the case of the chromatic polynomial.

The first section is devoted to the introduction and simplest properties of the Tutte polynomial, and in the second section we shall show that a certain universal polynomial can easily be obtained from this polynomial. In order to illustrate the ubiquitous nature of the Tutte polynomial, in Section 3 we shall introduce several models of disordered systems used in statistical mechanics, and show that their so-called partition functions are easy transforms of the Tutte polynomial. In Section 4 we shall show that the values of the Tutte polynomial at various places enumerate certain natural structures associated with our graph. A fundamental property of the Tutte polynomial is that it has a spanning tree expansion: we present this in Section 5.

It would take very little additional effort to define the Tutte polynomial on more general structures, namely on *matroids*, but we do not wish to burden the reader with more definitions. Anybody even only vaguely familiar with matroids will find the extension to matroids child's play.

The last section of the chapter concerns polynomials associated with (equivalence classes of) knots, especially the *Jones* and *Kauffman* polynomials. These polynomials, defined in terms of so-called *diagrams* of knots, were discovered only in the mid-1980s. This is rather surprising since these polynomials greatly resemble the Tutte polynomial, which had been constructed over thirty years earlier: in particular, they are naturally defined in terms of the obvious analogues of the cut and fuse operations. In fact, the resemblance is not only skin deep: on some classes of knots they are simple functions of the Tutte polynomial.

The knot polynomials related to the Tutte polynomial made it possible to prove several deep results about knots: we shall sketch a proof of one of these theorems. Needless to say, our brief excursion into the theory of knot polynomials hardly scratches the surface.

The natural setting for the polynomials to be studied in this chapter is the class of finite *multigraphs*; accordingly, all graphs occurring in this chapter are *multigraphs with loops*.

X.1 Basic Properties of the Tutte Polynomial

The Tutte polynomial is the best known member of a small family of equivalent polynomials. There are several natural ways of introducing these polynomials: here we choose to start with the *rank-generating polynomial.*

To prepare the ground, let $\mathcal{G}$ be the class of all finite multigraphs with loops. Strictly speaking, $\mathcal{G}$ is the set of all *isomorphism classes* of finite multigraphs, but it will be more convenient to consider the elements of $\mathcal{G}$ to be graphs rather than isomorphism classes of graphs. Also, we shall frequently refer to the elements of $\mathcal{G}$ as *graphs.*

For simplicity, we write $G = (V, E)$ for a multigraph with vertex set V, in which E is the set of multiple edges and loops. With a slight abuse of notation, we call $E = E(G)$ the *set of edges* of G, although in fact an edge of G is an element of E that is not a loop.

Let us define the cut and fuse operations, also called deletion and contraction, for the class $\mathcal{G}$. The *cut* operation is as before: given $G = (V, E)$ and $e \in E$, let $G - e = (V, E - \{e\})$. Thus $G - e$ is obtained from G by *cutting* (*deleting*) the edge e. Also, let G/e be the multigraph obtained from G by *fusing* (*contracting*) the edge e. Thus if $e \in E$ is incident with u and v (with $u = v$ if e is a loop) then in G/e the vertices u and v are replaced by a single vertex $w = (uv)$, and *each* element $f \in E - \{e\}$ that is incident with either u or v is replaced by an edge or loop incident with w (see Fig. X.1).

Note that the edge e itself corresponds to no edge of G/e; in particular, if ℓ is a loop then $G/\ell = G - \ell$. It is important to observe that both $G - e$ and G/e

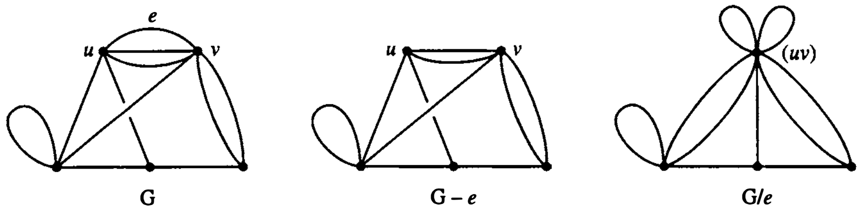


FIGURE X.1. A graph G together with $G - e$ and G/e for an edge $e = uv$.

have precisely one fewer multiple edge and loop than G: *every* element of $E - \{e\}$ corresponds to a unique element of $E(G/e)$.

Given a multigraph $G = (V, E)$, we write $k(G)$ for the number of components of G. The *rank* $r(G)$ and *nullity* $n(G)$ of G are defined as for graphs: $r(G) = |V| - k(G) = |G| - k(G)$ and $n(G) = |E| - |V| + k(G)$.

We shall work with spanning subgraphs of G. These subgraphs are naturally identified with their edge sets: for $F \subset E$ we write $\langle F \rangle$ for the graph (V, F), and $r\langle F \rangle$, $n\langle F \rangle$, $k\langle F \rangle$ for the rank, nullity and number of components of this graph. In particular, $r\langle E \rangle = r(G)$, $n\langle E \rangle = n(G)$ and $k\langle E \rangle = k(G)$.

We are now ready to define the *rank-generating polynomial* $S(G; x, y)$ of a graph $G = (V, E)$:

$$S(G; x, y) = \sum_{F \subset E(G)} x^{r\langle E \rangle - r\langle F \rangle} y^{n\langle F \rangle} = \sum_{F \subset E(G)} x^{k\langle F \rangle - k\langle E \rangle} y^{n\langle F \rangle}.$$

Our convention is that $S(G) = S(G; x, y)$ is a polynomial in x and y and this polynomial is a function of G. As for each G we have integer coefficients, $S(G; x, y) \in \mathbb{Z}[x, y]$ for every $G \in \mathcal{G}$. We shall use the same convention for other polynomials depending on a graph G, although we frequently prefer to write G as a subscript rather than an argument of the function.

The basic properties of the rank-generating polynomial are given in our first result. Although the proof is rather pedestrian, being only a sequence of formal manipulations, because of the importance of the result we spell it out in great detail.

Theorem 1 *Let $G = (V, E)$ be a graph with $e \in E$. Then*

$$S(G; x, y) = \begin{cases} (x+1)S(G-e; x, y) & \text{if } e \text{ is a bridge,} \\ (y+1)S(G-e; x, y) & \text{if } e \text{ is a loop,} \\ S(G-e; x, y) + S(G/e; x, y) & \text{if } e \text{ is neither a bridge} \\ & \text{nor a loop.} \end{cases}$$

Furthermore, $S(E_n; x, y) = 1$ for the empty n-graph E_n, $n \geq 1$.

Proof. Set $G' = G - e$, $G'' = G/e$, and write r' and n' for the rank and nullity functions in G', and r'' and n'' for those in G''.

Let us collect the simple properties of these functions we shall use below. If $e \in E$ and $F \subset E - e$ then $r\langle F \rangle = r'\langle F \rangle$, $n\langle F \rangle = n'\langle F \rangle$, $r\langle E \rangle - r\langle F \cup e \rangle =$

$$r''\langle E-e\rangle - r''\langle F\rangle = r(G'') - r''\langle F\rangle,$$

$$r\langle E\rangle = \begin{cases} r'\langle E-e\rangle + 1 & \text{if } e \text{ is a bridge,} \\ r'\langle E-e\rangle & \text{otherwise,} \end{cases}$$

and

$$n\langle F\cup e\rangle = \begin{cases} n''\langle F\rangle + 1 & \text{if } e \text{ is a loop,} \\ n''\langle F\rangle & \text{otherwise.} \end{cases}$$

Here, as in the future, $F \cup e$ and $E - e$ stand for $F \cup \{e\}$ and $E - \{e\}$.

Let us split $S(G; x, y)$ as follows:

$$S(G; x, y) = S_0(G; x, y) + S_1(G; x, y),$$

where

$$S_0(G; x, y) = \sum_{F\subset E, e\notin F} x^{r\langle E\rangle - r\langle F\rangle} y^{n\langle F\rangle}$$

and

$$S_1(G; x, y) = \sum_{F\subset E, e\in F} x^{r\langle E\rangle - r\langle F\rangle} y^{n\langle F\rangle}.$$

Recall that the sets $E - e$, $E(G') = E(G - e)$ and $E(G'') = E(G/e)$ are naturally identified. Hence, by the formulae above,

$$\begin{aligned} S_0(G; x, y) &= \sum_{F\subset E-e} x^{r\langle E\rangle - r\langle F\rangle} y^{n\langle F\rangle} \\ &= \begin{cases} \displaystyle\sum_{F\subset E(G')} x^{r'\langle E-e\rangle + 1 - r'\langle F\rangle} y^{n'\langle F\rangle} & \text{if } e \text{ is a bridge,} \\ \displaystyle\sum_{F\subset E(G')} x^{r'\langle E-e\rangle - r'\langle F\rangle} y^{n'\langle F\rangle} & \text{otherwise,} \end{cases} \\ &= \begin{cases} xS(G-e; x, y) & \text{if } e \text{ is a bridge,} \\ S(G-e; x, y) & \text{otherwise,} \end{cases} \end{aligned}$$

and

$$\begin{aligned} S_1(G; x, y) &= \sum_{F\subset E-e} x^{r\langle E\rangle - r\langle F\cup e\rangle} y^{n\langle F\cup e\rangle} \\ &= \begin{cases} \displaystyle\sum_{F\subset E(G'')} x^{r(G'') - r''\langle F\rangle} y^{n''\langle F\rangle + 1} & \text{if } e \text{ is a loop,} \\ \displaystyle\sum_{F\subset E(G'')} x^{r(G'') - r''\langle F\rangle} y^{n''\langle F\rangle} & \text{otherwise,} \end{cases} \\ &= \begin{cases} yS(G/e; x, y) & \text{if } e \text{ is a loop,} \\ S(G/e; x, y) & \text{otherwise.} \end{cases} \end{aligned}$$

The result follows by adding together the expressions for S_0 and S_1, and noting that if e is a bridge or a loop then $S(G/e; x, y) = S(G - e; x, y)$. This is obvious if e is a loop since then $G/e \cong G - e$; if e is a bridge, the assertion holds since $r''\langle E - e\rangle - r''\langle F\rangle = r'\langle E - e\rangle - r'\langle F\rangle$ and $n''\langle F\rangle = n'\langle F\rangle$ for all $F \subset E - e$.

Finally, the last part of the theorem is immediate from the definition of S. $\square$

The *Tutte polynomial* $T_G = T(G) = T(G; x, y) = T_G(x, y)$ is a simple function of the rank-generating polynomial:

$$T_G(x, y) = S(G; x - 1, y - 1) = \sum_{F \subset E} (x - 1)^{r\langle E\rangle - r\langle F\rangle}(y - 1)^{n\langle F\rangle}. \quad (1)$$

Most of the time we use T_G for the Tutte polynomial and write $T_G(x, y)$ if we wish to draw attention to the arguments; however, the notation $T(G)$ better expresses the fact that T is a map from the set of (equivalence classes of) finite multigraphs into $\mathbb{Z}[x, y]$. Tutte himself used $T(G; x, y)$ for the 'dichromate'. Note that $T_{E_n}(x, y) = 1$ and

$$T_G = \begin{cases} xT_{G-e} & \text{if } e \text{ is a bridge,} \\ yT_{G-e} & \text{if } e \text{ is a loop,} \\ T_{G-e} + T_{G/e} & \text{if } e \text{ is neither a bridge nor a loop.} \end{cases}$$

The recursion above will be used over and over again to show that a good many functions of graphs can be obtained by evaluating the Tutte polynomial at certain places.

When applying the reduction formulae above, it is worth noting that if $e \in E(G)$ is a bridge or a loop then $T_{G-e} = T_{G/e}$, since the analogous relation holds for S.

The functions S and T are *multiplicative* in the sense that if $G = G_1 \cup G_2$ with the graphs G_1 and G_2 sharing at most one vertex, then the value on G is the product of the values on G_1 and G_2 (see Exercise 4). In fact, these functions are determined by this multiplicativity condition, together with the recursion for an edge e that is neither a loop nor a bridge, and the values of the functions on K_2 and L, where L is the 'loop graph', that is, the graph with one vertex and one loop. By multiplicativity, $S(K_1) = T(K_1) = 1$; also, $S(K_2) = x + 1$, $T(K_2) = x$, $S(L) = y + 1$ and $T(L) = y$.

There are numerous other variants of the last observation. For example, T is also the unique function on graphs such that:

(i) if G has b bridges, l loops and no other edges then $T_G = x^b y^l$,

(ii) if G is obtained from a graph H by adding b bridges and l loops then $T_G = x^b y^l T_H$,

(iii) if G has no bridges or loops then the third recursion formula holds for *some* edge e, that is, there is an edge $e \in E(G)$ such that

$$T_G = T_{G-e} + T_{G/e}.$$

The reader is encouraged to check this simple assertion.

X.2 The Universal Form of the Tutte Polynomial

Our next aim is to show that the Tutte polynomial is easily lifted to a seemingly more general polynomial. In carrying out this task, we shall need that the maximum degrees of x and y in $T_G(x, y)$ can be read out of the definition of the Tutte polynomial: as

$$T_G(x, y) = \sum_{F \subset E} (x-1)^{r\langle E\rangle - r\langle F\rangle}(y-1)^{n\langle F\rangle},$$

we have

$$\deg_x T_G(x, y) = \max\{r(G) - r\langle F\rangle : F \subset E(G)\} = r(G),$$

and

$$\deg_y T_G(x, y) = \max\{n\langle F\rangle : F \subset E(G)\} = n(G).$$

The following simple assertion is a slight extension of a result of Oxley and Welsh.

Theorem 2 *There is a unique map $U : \mathcal{G} \to \mathbb{Z}[x, y, \alpha, \sigma, \tau]$ such that*

$$U(E_n) = U(E_n; x, y, \alpha, \sigma, \tau) = \alpha^n$$

for every $n \geq 1$, and for every $e \in E(G)$ we have

$$U(G) = \begin{cases} xU(G-e) & \text{if } e \text{ is a bridge,} \\ yU(G-e) & \text{if } e \text{ is a loop,} \\ \sigma U(G-e) + \tau U(G/e) & \text{if } e \text{ is neither a bridge nor a loop.} \end{cases}$$

Furthermore,

$$U(G) = \alpha^{k(G)}\sigma^{n(G)}\tau^{r(G)}T_G(\alpha x/\tau, y/\sigma). \tag{2}$$

Proof. The uniqueness is immediate since if $e \in E(G)$ then $U(G)$ is determined by $U(G-e)$ and $U(G/e)$. Hence all we have to prove is that the function U given by (2) has the required properties: i.e., $U(G)$ is a polynomial for every graph G, these polynomials satisfy the reduction formulae, and $U(E_n) = \alpha^n$ for every $n \geq 1$.

The fact that $U(G) \in \mathbb{Z}[x, y, \alpha, \sigma, \tau]$ follows from $\deg_x T_G(x, y) = r(G)$ and $\deg_y T_G(x, y) = n(G)$. Also, as $k(E_n) = n$ and $r(E_n) = n(E_n) = 0$, we have

$$U(E_n) = \alpha^n T_{E_n}(\alpha x/\tau, y/\sigma) = \alpha^n.$$

Most importantly, U satisfies the reduction formulae since the Tutte polynomial satisfies them with $\sigma = \tau = 1$. To spell it out, if $e \in E(G)$ is a bridge then $k(G-e) = k(G) + 1$, $n(G-e) = n(G)$ and $r(G-e) = r(G) - 1$, so

$$\begin{aligned} U(G) &= \alpha^{k(G)}\sigma^{n(G)}\tau^{r(G)}T_G(\alpha x/\tau, y/\sigma) \\ &= \alpha^{k(G-e)-1}\sigma^{n(G-e)}\tau^{r(G-e)+1}(\alpha x/\tau)T_{G-e}(\alpha x/\tau, y/\sigma) = xU(G-e). \end{aligned}$$

If e is a loop of G then $k(G-e) = k(G)$, $n(G-e) = n(G) - 1$ and $r(G-e) = r(G)$, so

$$\begin{aligned} U(G) &= \alpha^{k(G)}\sigma^{n(G)}\tau^{r(G)}T_G(\alpha x/\tau, y/\sigma) \\ &= \alpha^{k(G-e)}\sigma^{n(G-e)+1}\tau^{r(G-e)}(y/\sigma)T_{G-e}(\alpha x/\tau, y/\sigma) = yU(G-e). \end{aligned}$$

Finally, if $e \in E(G)$ is neither a bridge nor a loop of G then $k(G-e) = k(G/e) = k(G)$, $n(G-e) = n(G) - 1$, $n(G/e) = n(G)$, $r(G-e) = r(G)$ and $r(G/e) = r(G) - 1$, so

$$\begin{aligned} U(G) &= \alpha^{k(G)}\sigma^{n(G)}\tau^{r(G)}T_G(\alpha x/\tau, y/\sigma) \\ &= \alpha^{k(G)}\sigma^{n(G)}\tau^{r(G)}\{T_{G-e}(\alpha x/\tau, y/\sigma) + T_{G/e}(\alpha x/\tau, y/\sigma)\} \\ &= \alpha^{k(G-e)}\sigma^{n(G-e)+1}\tau^{r(G-e)}T_{G-e}(\alpha x/\tau, y/\sigma) \\ &\quad + \alpha^{k(G/e)}\sigma^{n(G/e)}\tau^{r(G/e)+1}T_{G/e}(\alpha x/\tau, y/\sigma) \\ &= \sigma U(G-e) + \tau U(G/e), \end{aligned}$$

completing the proof. □

We call the polynomial U in Theorem 2 the *universal polynomial* of graphs. Theorem 2 implies that if R is a commutative ring and $x, y, \alpha, \sigma, \tau \in R$ then there is a unique map $\mathcal{G} \to R$ satisfying the conditions of the theorem: that map is obtained by evaluating the polynomial U at the required place. In particular, there is a map for $R = \mathbb{Z}[x, y]$ and every choice of $\alpha, \sigma, \tau \in \mathbb{Z}$. For example, the Tutte polynomial itself is just U evaluated at $\alpha = \sigma = \tau = 1$.

The polynomial U is also multiplicative, but in a somewhat weaker sense than S and T: if G_1 and G_2 are vertex disjoint graphs then

$$U(G_1 \cup G_2) = U(G_1)U(G_2),$$

and if G_1 and G_2 share one vertex then

$$U(G_1 \cup G_2) = \frac{U(G_1)U(G_2)}{\alpha}.$$

(In particular, for $\alpha = 1$ the function U is multiplicative in the stronger sense of S and T.) It is easily seen that U is determined by this multiplicativity property, together with the conditions $U(L) = \alpha y$ and $U(K_2) = \alpha^2 x$. Once again, $U(K_1) = \alpha$ follows from multiplicativity.

A word of caution: unless σ and τ are both non-zero, we have to evaluate the *expanded* polynomial in (2) rather than the factors one by one. In fact, a little work shows that if $\sigma = 0$ or $\tau = 0$ then U takes a particularly pleasant form. Writing $\ell(G)$ for the number of loops of G, $b(G)$ for the number of bridges and, as always, $|G|$ for the number of vertices,

$$U(G; x, y, \alpha, \sigma, 0) = \alpha^{|G|}\sigma^{n(G)-\ell(G)}x^{r(G)}y^{\ell(G)} \tag{3}$$

and

$$U(G; x, y, \alpha, 0, \tau) = \alpha^{k(G)+b(G)}\tau^{r(G)-b(G)}x^{b(G)}y^{n(G)}. \tag{4}$$

Also,

$$U(G; x, y, \alpha, 0, 0) = \begin{cases} \alpha^{|G|} x^{b(G)} y^{\ell(G)} & \text{if } E(G) \text{ consists of loops and bridges,} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

In §4 we shall show that a good many values of the Tutte polynomial have considerably more interesting interpretations.

Let us mention two more members of the family of polynomials related to the Tutte polynomial. The first, $Q_G = Q_G(t, z) \in \mathbb{Z}[t, z]$, we shall call the *Whitney-Tutte polynomial.* It is defined as

$$Q_G(t, z) = \sum_{F \subset E(G)} t^{k\langle F \rangle} z^{n\langle F \rangle},$$

so that

$$Q_G(t, z) = t^{k(G)} S(G; t, z)$$

and

$$T_G(x, y) = (x-1)^{-k(G)} Q_G(x-1, y-1).$$

We shall call the second, $Z_G = Z_G(q, v) \in \mathbb{Z}[q, v]$, the *dichromatic polynomial.* It is the unique polynomial such that $Z_{E_n} = q^n$ for every $n \geq 1$, and

$$Z_G = Z_{G-e} + v Z_{G/e} \tag{6}$$

for *every* edge $e \in E(G)$, whether e is a bridge, a loop or neither. It is easily checked (see Exercise 6) that Z is just U evaluated at $\alpha = q$, $\sigma = 1$, $\tau = v$, $x = (q+v)/q$ and $y = 1 + v$, that is,

$$Z_G(q, v) = q^{k(G)} v^{r(G)} T_G((q+v)/v,\ 1+v).$$

As we shall see in the next section, it is precisely the incarnation Z_G of the Tutte polynomial that appears in statistical physics.

X.3 The Tutte Polynomial in Statistical Mechanics

In statistical mechanics we wish to study random disordered systems, especially in the neighbourhood of their phase transitions. In many instances, even before we start our investigations, we have to overcome the somewhat unexpected difficulty that although it is easy to give a *measure* proportional to the probability measure we wish to study, it is not easy to normalize it so that it becomes a probability measure. The total measure of the space (and so our normalizing factor) is the *partition function.* As we shall see now, in several important cases the partition functions are simple variants of the Tutte polynomial.

Let us start with the q-state Potts model, where $q \geq 1$ is an integer. To introduce this model, let $G = (V, E)$ be a multigraph. We call a function $\omega : V \to [q]$, $a \mapsto \omega_a$, a *state* of the *q-state Potts ferromagnetic model on G*. The value ω_a is the *state* of a or the *spin* at a. To define a measure on the set $\Omega = [q]^V$ of all states, we need the *Hamiltonian* $H(\omega)$ of a state ω:

$$H(\omega) = \sum_{ab \in E} (1 - \delta(\omega_a, \omega_b)).$$

Here and in what follows, we consider the set E of edges to be a multiset, so that in a sum $\sum_{xy \in E} f(x, y)$ the multiplicity of $f(x, y)$ is precisely the number of edges or loops with endvertices x and y. The *Potts measure* on $\Omega = [q]^V$ is defined by

$$\mu_G^{q,\beta}(\omega) = e^{-H(\omega)/k_B T} = e^{-\beta H(\omega)},$$

where k_B is the Boltzmann constant (1.38×10^{-23} joules/Kelvin), T is the *temperature* of the system, and β is the *inverse temperature*. The *partition function* of the *q-state Potts model* on G is

$$P_G(q, \beta) = \mu_G^{q,\beta}(\Omega) = \sum_{\omega \in \Omega} \mu_G^{q,\beta}(\omega),$$

and the probability of a state ω is $\mu_G^{q,\beta}(\omega)/P_G(q, \beta)$. Note that at high temperatures all states have about the same probability, while at low temperatures a small change in the Hamiltonian changes the probability a great deal. Much of our interest in the system is due to the fact that the structure of a 'typical' system changes suddenly as the temperature passes through a certain critical value. This *phase transition* is reminiscent of the phenomenon in the evolution of random graphs we discussed in Section VII.5.

Theorem 3 *Let $G = (V, E)$ be a multigraph, $q \geq 1$ an integer and $\beta \in \mathbb{R}$. Then the partition function of the q-state Potts model on G, with inverse temperature β, is*

$$P_G(q, \beta) = e^{-\beta|E|} Z_G(q, v),$$

where Z_G is the dichromatic polynomial and $v = e^\beta - 1$.

Proof. Set

$$\tilde{P}_G(q, \beta) = e^{\beta|E|} P_G(q, \beta),$$

so that we have to show that $\tilde{P}_G(q, \beta) = Z_G(q, v)$. If $G = E_n$ then $H(\omega) = 0$ for every state ω, so $\tilde{P}_{E_n} = P_{E_n} = q^n$. In order to prove that $\tilde{P}_G(q, \beta) = Z_G(q, v)$, all we have to check is that $\tilde{P}_G(q, \beta)$ satisfies the reduction formula (6). Note that

$$\begin{aligned} \tilde{P}_G(q, \beta) &= e^{\beta|E|} \sum_{\omega \in \Omega} e^{\beta \sum_{ab \in E}(\delta(\omega_a, \omega_b) - 1)} \\ &= \sum_{\omega \in \Omega} \prod_{ab \in E} e^{\beta\delta(\omega_a, \omega_b)} \\ &= \sum_{\omega \in \Omega} \prod_{ab \in E} (1 + v\delta(\omega_a, \omega_b)). \end{aligned}$$

To prove the reduction formula, let e be an edge from c to d. Let us split the sum above: first let us sum over the states ω with $\omega_c \neq \omega_d$, and then over the states with $\omega_c = \omega_d$. Thus

$$\begin{aligned}\tilde{P}_G(q,\beta) &= \sum_{\omega_c \neq \omega_d} \prod_{ab \in E-e} (1 + v\delta(\omega_a, \omega_b)) \\ &\quad + (1+v) \sum_{\omega_c = \omega_d} \prod_{ab \in E-e} (1 + v\delta(\omega_a, \omega_b)) \\ &= \sum_{\omega \in \Omega} \prod_{ab \in E-e} (1 + v\delta(\omega_a, \omega_b)) + v \sum_{\omega_c = \omega_d} \prod_{ab \in E-e} (1 + v\delta(\omega_a, \omega_b)) \\ &= \tilde{P}_{G-e}(q,\beta) + v\tilde{P}_{G/e}(q,\beta).\end{aligned}$$

Hence (6) is satisfied, and we are done. □

The Potts model is a generalization of the *Ising model*, which is just the case $q = 2$. In fact, Fortuin and Kasteleyn constructed an extension of the Potts model itself: our next aim is to introduce this extension, the so-called random cluster model. Let $G = (V, E)$ be a multigraph, and let $0 < p < 1$ and $q > 0$ be fixed. Most importantly, we do not take q to be an integer. The *random cluster model* on G, with parameters q and p, is a probability space on all spanning subgraphs of G. As before, such a subgraph will be identified with its edge set $F \subset E$. The measure of a subgraph $\langle F \rangle$ is

$$\nu_G^{q,p}(F) = p^{|F|}(1-p)^{|E|-|F|} q^{k\langle F \rangle},$$

and so the *partition function* of the random cluster model is

$$R_G(q,p) = \sum_{F \subset E} p^{|F|}(1-p)^{|E|-|F|} q^{k\langle F \rangle}.$$

To turn $\nu_G^{q,p}$ into a probability measure, we have to divide it by $R_G(q,p)$.

The random cluster model is not too far from the standard random graph model $\mathcal{G}(G,p)$, in which we obtain a random subgraph of G by selecting the edges of G (and only those!) with probability p, independently of each other. (Thus, $\mathcal{G}(n,p) = \mathcal{G}(K_n,p)$.) To get a random graph in a random cluster model, we bias the standard probability of a graph with k components by a factor of q^k. So, if $q = 1$, we get precisely $\mathcal{G}(G,p)$, but if q is large then we heavily favour graphs with many components. Nevertheless, the similarity to the standard model $\mathcal{G}(G,p)$ occasionally allows one to use the methods and results of the theory of random graphs to study random cluster models.

Theorem 4 *The partition function of the random cluster model is*

$$R_G(q,p) = (1-p)^{|E|} Z_G(q,v),$$

where $v = p/(1-p)$.

Proof. Set $\tilde{R}_G(q,p) = (1-p)^{-|E|} R_G(q,p)$, so that

$$\tilde{R}_G(q,p) = \sum_{F \subset E} v^{|F|} q^{k\langle F \rangle}, \tag{7}$$

and we have to show that $\tilde{R}_G(q, p) = Z_G(q, v)$. Clearly, $\tilde{R}_{E_n} = q^n$, so all we have to check is that the reduction formula holds. To this end, let $e \in E$. Let us partition the subsets of E into pairs,

$$\{F : F \subset E\} = \bigcup_{F \subset E-e} \{F, F \cup \{e\}\},$$

and let us split (7) accordingly:

$$\begin{aligned}\tilde{R}_G(q, p) &= \sum_{F \subset E-e} \{v^{|F|} q^{k\langle F\rangle} + v^{|F|+1} q^{k\langle F \cup e\rangle}\} \\ &= \sum_{F \subset E-e} v^{|F|} q^{k\langle F\rangle} + v \sum_{F \subset E-e} v^{|F|} q^{k\langle F \cup e\rangle}.\end{aligned}$$

The first sum is precisely $\tilde{R}_{G-e}(q, p)$. As $\langle F \cup \{e\}\rangle$ has precisely as many components in G as $\langle F\rangle$ has in G/e, the second sum is $\tilde{R}_{G/e}(q, p)$, and we are finished. □

Needless to say, these partition functions are easily expressed in terms of the Tutte polynomial (see Exercise 42), but the expressions are not too edifying. However, formula (7) gives yet another very simple way of defining the dichromatic polynomial Z_G, and so the Tutte polynomial (see Exercise 44).

X.4 Special Values of the Tutte Polynomial

As we remarked earlier, the Tutte polynomial of a graph carries much more information about the graph than the chromatic polynomial. In particular, as we shall soon see, the chromatic polynomial is simply the Tutte polynomial with $y = 0$, suitably normalized. But first let us note that if $x, y \in \{1, 2\}$ then $T_G(x, y)$ is the number of certain simple subgraphs of G.

Theorem 5 *Let G be a connected graph. Then $T_G(1, 1)$ is the number of spanning trees of G, $T_G(2, 1)$ is the number of (edge sets forming) forests in G, $T_G(1, 2)$ is the number of connected spanning subgraphs, and $T_G(2, 2)$ is the number of spanning subgraphs.*

Proof. Each of these assertions is immediate from the definition (1) of T. Thus,

$$\begin{aligned}T_G(1, 1) &= \sum_{F \subset E(G)} 0^{r(G)-r\langle F\rangle} 0^{n\langle F\rangle} \\ &= |\{F : \quad F \subset E(G), r\langle F\rangle = r(G) \text{ and } n\langle F\rangle = 0\}|,\end{aligned}$$

and $F \subset E(G)$ is the edge set of a spanning tree iff $r\langle F\rangle = r(G)$ and $n\langle F\rangle = 0$.

Similarly,

$$T_G(2, 1) = \sum_{F \subset E(G)} 1^{r(G)-r\langle F\rangle} 0^{n\langle F\rangle} = |\{F : F \subset E(G) \quad \text{and} \quad n\langle F\rangle = 0\}|$$

is the number of edge sets F forming forests, and

$$T_G(1,2) = \sum_{F \subset E(G)} 0^{r(G)-r\langle F\rangle} 1^{n\langle F\rangle} = |\{F : F \subset E(G) \quad \text{and} \quad r\langle F\rangle = r(G)\}|$$

is the number of connected spanning subgraphs of G.

Finally,

$$T_G(2,2) = \sum_{F \subset E(G)} 1^{r(G)-r\langle F\rangle} 1^{n\langle F\rangle} = |\{F : F \subset E(G)\}| = 2^{|E(G)|},$$

as claimed. □

Let us turn to families of values of the Tutte polynomial. As our first general result, we shall show that the chromatic polynomial is just the Tutte polynomial with $y = 0$, suitably normalized.

In keeping with our earlier notation, given a multigraph G with loops and a positive integer x, let us write $p_G(x)$ for the number of proper vertex-colourings of G, that is for the number of maps $c : V(G) \to \{1, 2, \ldots, x\}$ such that if u and v are adjacent vertices then $c(u) \neq c(v)$. Clearly $p_{E_n}(x) = x^n$ and as in Chapter V, for every edge $e \in E(G)$ we have

$$p_G(x) = p_{G-e}(x) - p_{G/e}(x).$$

In particular, $p_G(x)$ is a polynomial in x, the chromatic polynomial of G. Let us note two more simple properties of the chromatic polynomial. First, if there is a loop at a vertex then the vertex is adjacent to itself, so $p_G(x) = 0$ if G contains a loop. Clearly, if H is the graph obtained from G by replacing the multiple edges by simple edges then $p_G(x) = p_H(x)$. Secondly, recall from Exercise 48 of Chapter V that if e is a bridge of G then

$$p_G(x) = \frac{x-1}{x} p_{G-e}(x).$$

Theorem 6 *The chromatic polynomial $p_G(x)$ of a graph G is*

$$p_G(x) = (-1)^{r(G)} x^{k(G)} T_G(1-x, 0).$$

Proof. The result is immediate from Theorem 2 and the properties of the chromatic polynomial mentioned above. Indeed, $p_{E_n}(x) = x^n$, and for every edge $e \in E(G)$,

$$p_G(x) = \begin{cases} \dfrac{x-1}{x} p_{G-e}(x) & \text{if } e \text{ is a bridge,} \\ 0 & \text{if } e \text{ is a loop,} \\ p_{G-e}(x) - p_{G/e}(x) & \text{if } e \text{ is neither a bridge nor a loop.} \end{cases}$$

Hence, by Theorem 2, $p_G(x) = U(G; \frac{x-1}{x}, 0, x, 1, -1) = x^{k(G)}(-1)^{r(G)} \times T_G(1-x, 0)$, as claimed. □

Setting $x = 0$ in the Tutte polynomial, the polynomial in y we obtain also has a natural interpretation: it is the so-called *flow polynomial* of the graph, suitably normalized.

To define the flow polynomial, we shall consider flows in our graphs (multigraphs with loops) with values in a finite additively written Abelian group A. We call such a flow an *A-flow* if it satisfies Kirchhoff's current law at *each* vertex. Thus an A-flow is really a circulation (see Exercise 7 in Chapter III), but it would be wrong not to use the accepted terminology.

As in our study of flows in graphs in Chapters II and III, one may pick an orientation of the edges, giving a set $\vec{E}$ of directed edges and loops, and then an A-flow is a map $f : \vec{E} \to A$ such that the total flow *out* of a vertex is equal to the total flow *into* the vertex. Equivalently, each edge $e = xy$ has a certain flow f_{xy} in it from x to y, with the convention that it is the same as a flow $-f_{yx}$ in the same edge from y to x.

An A-flow is said to be *nowhere-zero* if it has a non-zero value in every edge. Let us write $q_G(A)$ for the *number of nowhere-zero A-flows* in G. As we shall see shortly, $q_G(A)$ is a polynomial in the order of the group A, so we are justified in calling it the flow polynomial.

Occasionally, $q_G(A)$ is easily determined. For example, $q_{E_n}(A) = 1$ for every $n \geq 1$ since there is only one A-flow in E_n, the unique map from the empty set into A, and that A-flow is nowhere-zero. If $G = C_n$ is an n-cycle $x_1x_2 \ldots x_n$ with $n \geq 1$ then $q_G(A) = |A| - 1$, since an A-flow assigns the same current to each of $x_1x_2, x_2x_3, \ldots, x_nx_1$.

Let us see what we can say about $q_G(A)$, $q_{G-e}(A)$ and $q_{G/e}(A)$ for an edge $e \in E$. First, if e is a bridge then $q_G(A) = 0$. Indeed, suppose that e is the only edge from V_1 to $V_2 = V(G) - V_1$. In an A-flow the total current from V_1 to V_2 is 0, so there is no current in the bridge e. Hence $q_G(A) = 0$.

Secondly, if e is a loop then every A-flow on G is obtained from an A-flow on $G - e$ by sending an arbitrary current through the loop e. In order to obtain a nowhere-zero A-flow on G, we have to start with a nowhere-zero A-flow on $G - e$ and then choose one of $|A| - 1$ non-zero values for the current in the loop e. Consequently,

$$q_G(A) = (|A| - 1)q_{G-e}(A).$$

Finally, suppose that e is neither a loop nor a bridge, and joins u to v. Consider a nowhere-zero A-flow f on G/e. As the edge sets $E(G/e)$ and $E(G - e)$ are naturally identified, f can be viewed as a flow f' on $G - e$. Clearly, either f' is a nowhere-zero A-flow on $G - e$, or else f' fails Kirchhoff's current law at u and v, and nowhere else. In the latter case, there is a *unique* extension of this flow to a nowhere-zero A-flow f'' on G: the current in e has to be chosen to make Kirchhoff's current law hold at u (and then it holds at v as well). Furthermore, every nowhere-zero A-flow f' on $G - e$ is obtained in this way and so is every nowhere-zero A-flow f'' on G. Consequently,

$$q_{G/e}(A) = q_{G-e}(A) + q_G(A).$$

A priori, $q_G(A)$ seems to depend on the structure of A; as the next result implies, rather surprisingly, this is not the case: $q_G(A)$ depends *only* on the order of A.

Theorem 7 *Let A be a finite Abelian group and G a multigraph. Then*

$$q_G(A) = (-1)^{n(G)} T_G(0, 1 - |A|).$$

Proof. The result is, once again, immediate from Theorem 2 and the properties of the flow polynomial noted above. Indeed, we have shown that $q_{E_n}(A) = 1$ for every $n \geq 1$, and if $e \in E(G)$ then

$$q_G(A) = \begin{cases} 0 & \text{if } e \text{ is a bridge,} \\ (|A| - 1) q_{G-e}(A) & \text{if } e \text{ is a loop,} \\ -q_{G-e}(A) + q_{G/e}(A) & \text{if } e \text{ is neither a bridge nor a loop.} \end{cases}$$

Hence, by Theorem 2,

$$q_G(A) = U(G; 0, |A| - 1, 1, -1, 1) = (-1)^{n(G)} T_G(0, 1 - |A|),$$

as required. □

As an easy consequence of Theorem 7, we see that $q_G(A)$ depends *only* on the order of A and not on its structure. In view of this, we denote $q_G(A)$ by $q_G(|A|)$. Furthermore, this function $q_G(k)$ is a polynomial in $k \in \mathbb{N}$, so it gives a polynomial $q_G(x) \in \mathbb{Z}[x]$. We call $q_G(x)$ the *flow polynomial* of G. As $q_G(A)$ depends only on $|A|$, it is customary to talk of a *k-flow*, meaning an A-flow with $|A| = k$ or simply a $\mathbb{Z}_k$-flow.

It is natural to consider the flow polynomial as the dual of the chromatic polynomial (see Exercise 16). In particular, the four colour theorem is equivalent to the fact that every *bridgeless* planar graph has a nowhere-zero 4-flow. Furthermore, corresponding to Hadwiger's conjecture concerning colourings, Tutte conjectured in 1954 that *every* bridgeless graph has a nowhere-zero 5-flow. In fact, it is far from obvious that one can even guarantee a nowhere-zero k-flow for *some* k. This was proved by Jaeger in 1979 when he showed that one can guarantee a nowhere-zero 8-flow. A little later, Seymour proved that, in fact, every bridgeless graph has a nowhere-zero 6-flow. This remarkable result is still rather far from a proof of Tutte's 5-flow conjecture.

Our penultimate example resembles the last result: as a byproduct of the evaluation of a function, we shall obtain its independence from one of its variables.

We know from Exercise 52 of Chapter V that the number $a(G)$ of acyclic orientations of a graph G is given by the chromatic polynomial at $x = -1$: $a(G) = |p_G(-1)| = (-1)^{|G|} p_G(-1)$. Hence, by Theorem 6, $a(G) = T_G(2, 0)$. As we are about to see, $T_G(1, 0)$ counts *certain* acyclic orientations of connected graphs. Given a connected graph G and a vertex u of G, write $a_u(G)$ for the number of acyclic orientations of G in which there is only one source, and that source is u.

Theorem 8 *For every connected graph G and every vertex $u \in V(G)$ we have*

$$a_u(G) = T_G(1, 0).$$

Proof. We shall deduce the assertion from the following four properties of the function $a_u(G)$.

(i) If $G = E_1$ then $a_u(G) = 1$.

(ii) If G contains a loop e then G has no acyclic orientation so $a_u(G) = 0$.

(iii) Suppose that $e = uv$ is a bridge of G, and consider an acyclic orientation of G, with u the only source. Then in the component of $G - e$ containing v, the only source has to be v, so the acyclic orientations of G with u the only source are in 1–to–1 correspondence with the acyclic orientations of G/e, with u (which in G/e is the same as v or (uv)) the only source. Hence

$$a_u(G) = a_u(G/e).$$

(iv) Finally, suppose that $e = uv \in E(G)$ is neither a loop nor a bridge. Consider an acyclic orientation of G, with u the only source. Let us ask the question: is uv the *only* edge directed into v? If it is, then our orientation gives an acyclic orientation of G/e in which u is the only source; otherwise, it gives an orientation of $G-e$ in which u is the only source. Also, all appropriate orientations of G/e and $G - e$ arise in this way. Consequently, in this case we have

$$a_u(G) = a_u(G - e) + a_u(G/e).$$

Note now that if u is a vertex of a connected graph G with $e(G) > 0$ then there is an edge $e \in E(G)$ incident with u. But then $a_u(G)$ is determined by the 'nature' of this edge (loop, bridge or neither) and the values $a_u(G-e)$ and $a_u(G/e)$. Hence there is a *unique* function $a_u(G)$ on the set of (equivalence classes of) *connected graphs G with a distinguished vertex u* that has properties (i) - (iv). Recalling that $T_{G-e} = T_{G/e}$ whenever e is a bridge or a loop, we see that $T_G(1, 0)$ is such a function, so $a_u(G) = T_G(1, 0)$, as claimed. □

As our final example, we shall show that the Tutte polynomial can be used to determine the probability of connectedness of a random subgraph of a connected graph. In fact, a slightly more general assertion will be an immediate consequence of Theorem 2.

Let $G = (V, E)$ be a graph, and let $0 < p < 1$. With a slight abuse of notation, we shall write E_p for a random subset of E obtained by retaining the edges with probability p, independently of each other (and so deleting them with probability $q = 1 - p$). Thus if $G = K_n$ then $\langle E_p \rangle$ is precisely an element of $\mathcal{G}(n, \mathbb{P}(\text{edge}) = p)$, studied in Chapter VII; in general, $\langle E_p \rangle$ is a random subgraph of G. We write $\mathbb{P}$ for the probability on these random subgraphs $\langle E_p \rangle$.

Theorem 9 *Let $G = (V, E)$, $0 < p < 1$, $q = 1 - p$ and E_p be as above. Then*

$$\mathbb{P}(r\langle E_p \rangle = r(G)) = p^{r(G)} q^{n(G)} T_G(1, 1/q).$$

Proof. In view of Theorem 2, it suffices to check that the function $C(G) = \mathbb{P}(r\langle E_p \rangle = r(G))$ satisfies the conditions of Theorem 2 with $x = p$, $y = 1$, $\alpha = 1$, $\sigma = q$ and $\tau = p$.

Although this is very easily seen, we shall spell it out.

If G is the empty graph E_n then $r\langle E_p \rangle = r(G) = 0$, so $C(E_n) = 1$.

If $e \in E$ is a bridge of G then $r\langle E_p \rangle = r\langle E \rangle$ implies that $e \in E_p$. Consequently, $C(G) = pC(G - e)$.

If $e \in E$ is not a bridge of G then $r(G) = r(G - e)$, so $C(G) = pC(G/e) + qC(G - e)$. Also, if e is a loop then $G/e = G - e$, so $C(G) = C(G - e)$, a fact obvious from first principles as well. □

Among the exercises we give two other beautiful instances of enumeration by the Tutte polynomial (see Exercises 10–11).

X.5 A Spanning Tree Expansion of the Tutte Polynomial

In this section we shall give Tutte's original definition of his polynomial, and we shall show that it agrees with our definition. As we know the basic properties of the Tutte polynomial, our task will be much easier than it would be if we started from first principles.

Although one tends to talk of writing the Tutte polynomial as a sum over all spanning trees, this is possible only if our graph is connected. (Surprise, surprise!) As the Tutte polynomial of a graph is the product of the Tutte polynomials of its components, we are led to an expansion in terms of forests whose components are spanning trees of the components of the graph. With a slight (but definite) abuse of terminology, we call such a forest a *spanning forest*.

Thus a graph $F = (V', E')$ is a *spanning forest* of a graph $G = (V, E)$ if $V' = V$, $E' \subset E$, and each component of F is a spanning tree of a component of G. Equivalently, $V' = V$, $E' \subset E$, $r(F) = r(G)$ and $n(F) = 0$. Putting it slightly differently, a spanning forest of a graph is a subforest with the same number of vertices and the same number of components as the original graph.

Let G be a graph and let us consider an order on its edge set: say, $E(G) = \{e_1, e_2, \dots, e_m\}$, with e_i preceding e_j if $i < j$. Also, let F be a spanning forest of G. Following the terminology in Section II. 1, for $e_i \in E(F)$ we call $U_F(e_i) = \{e_j \in E(G) : (F - e_i) + e_j \text{ is a spanning forest}\}$ the *cut* defined by e_i. Furthermore, for $e_i \in E(G) - E(F)$, the *cycle* defined by e_i is the unique cycle of $F + e_i$; we write $Z_F(e_i)$ for the edge set of this cycle.

Call an edge $e_i \in E(F)$ an *internally active* edge (of F, with respect to the ordering of the edges of G) if e_i is the smallest edge of the cut it defines. Thus $e_i \in E(F)$ is internally active if $i \le j$ whenever $e_j \in U_F(e_i)$, that is, if $e_i \in Z_F(e_j)$ implies $i \le j$. Similarly, call $e_j \in E(G) - E(F)$ *externally active* if e_j is the smallest edge of the cycle it defines, that is, if $i \ge j$ whenever $e_i \in Z_F(e_j)$. For an illustration, see Fig. X.2.

We say that a spanning forest has *internal activity* i and *external activity* j if there are precisely i internally active edges and precisely j externally active edges. Also, as a shorthand, by an (i, j)-forest we mean a spanning forest of internal activity i and external activity j (with respect to some given ordering of the edges.) If we know that our forest is, in fact, a spanning tree, then we shall talk of an (i, j)*-tree*.

Note that, upholding the ancient tradition of pure mathematicians, we use the same generic indices in two different contexts: we have edges e_i on a forest, edges

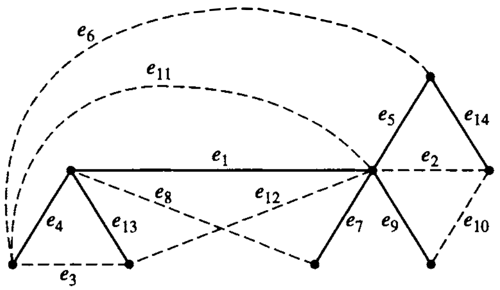


FIGURE X.2. The subgraph in bold is a spanning tree with internally active edges e_1, e_7, e_9 and externally active edges e_2, e_3.

e_j not on a forest, and we also have (i, j)-forests, that is, spanning forests with i special edges of one kind (denoted by e_i) and j special edges of another kind (denoted by e_j). Hopefully, this does not lead to any confusion.

Theorem 10 *Let G be a graph with an ordering of its edges. Write t_{ij} for the number of spanning forests with internal activity i and external activity j. Then $\sum_{i,j} t_{ij}x^i y^j$ is precisely the Tutte polynomial $T_G(x, y)$ of G.*

In particular, $t_{ij} = t_{ij}(G)$ is independent of the ordering of the edges and depends only on the graph G.

Proof. We shall prove the assertion by induction on the number of edges of G. For $G = E_n$ we have $t_{00} = 1$ and $t_{ij} = 0$ if $i + j > 0$, so $\sum_{i,j} t_{ij}x^i y^j = 1 = T_{E_n}(x, y)$.

Let $G = (V, E)$, $e = \{e_1, \ldots, e_m\}$, $m \geq 1$, and assume that the assertion holds for graphs with at most $m - 1$ edges. Set $G' = G - e_m$ and $G'' = G/e_m$, so that $E(G') = E(G'') = \{e_1, \ldots, e_{m-1}\}$. This is the order we shall take on the edge sets of G' and G'' which, as usual, we take to be identical. Let t'_{ij} be the number of (i, j)-forests in G', and let $t''_{i,j}$ be the number in G''. By the induction hypothesis, $\sum_{i,j} t'_{ij}x^i y^j = T_{G-e_m}(x, y)$ and $\sum_{i,j} t''_{ij}x^i y^j = T_{G/e_m}(x, y)$. We shall distinguish the three usual cases according to the nature of e_m. The arguments in (i), when e_m is a bridge, and (ii), when e_m is a loop, agree almost verbatim, and so do the arguments in (iii) for $G - e_m$ and G/e_m; nevertheless, due to the importance of the result, we give all the details of this practically trivial proof.

(i) *Suppose that e_m is a bridge.* Then e_m is in every spanning forest of G, and a subgraph F of G is a spanning forest iff $e_m \in E(F)$ and $F - e_m$ is a spanning forest of $G - e_m$. Also, e_m is internally active in every spanning forest F of G, since it is in no cycle $Z_F(e_j)$.

Clearly, for $1 \leq i \leq m - 1$ the edge e_i is internally (externally) active in G with respect to F iff it is internally (externally) active in $G - e_m$ with respect to $F - e_m$, since in checking the activity of e_i exactly the same set of edges is involved in each case. Hence F is an (i, j)-forest of G iff $e_m \in E(F)$ and $F - e_m$ is an $(i - 1, j)$-forest of $G - e_m$, and so, *a fortiori* $t_{ij} = t'_{i-1,j}$. Therefore, by the induction hypothesis and a basic property of the Tutte polynomial,

$$\sum_{i,j} t_{ij} x^i y^j = \sum_{i,j} t'_{i-1,j} x^i y^j = x \sum_{i,j} t'_{i-1,j} x^{i-1} y^j = x \sum_{i,j} t'_{ij} x^i y^j$$
$$= x T_{G-e_m}(x, y) = T_G(x, y),$$

as required.

(ii) *Suppose that e_m is a loop.* Then e_m is in no spanning forest of G, and a subgraph F of G is a spanning forest of G iff it is a spanning forest of $G-e_m$. Also, e_m is externally active in every spanning forest F of G since it is the only edge of $Z_F(e_m)$. In addition, for $1 \le i \le m-1$ the edge e_i is internally (externally) active in G with respect to F iff it is internally (externally) active in $G - e_m$ with respect to the same spanning forest F. Hence F is an (i, j)-forest of G iff it is an $(i, j-1)$-forest of $G - e_m$. Consequently, $t_{ij} = t'_{i,j-1}$, so

$$\sum_{i,j} t_{ij} x^i y^j = \sum_{i,j} t'_{i,j-1} x^i y^j = y \sum_{i,j} t'_{ij} x^i y^j$$
$$= y T_{G-e_m}(x, y) = T_G(x, y).$$

(iii) *Suppose that e_m is neither a bridge nor a loop.* Trivially, F is a spanning forest of $G - e_m$ iff it is a spanning forest of G and $e_m \notin E(F)$. Also, if F is an (i, j)-forest of $G - e_m$, then it is an (i, j)-forest of G, since every other edge precedes e_m and, as e_m is not a loop, $Z_F(e_m)$ has other edges in addition to e_m.

Similarly, F is a spanning forest of G/e_m iff $e_m \notin E(F)$ and $F + e_m$ is a spanning forest of G. Also, if F is an (i, j)-forest of G/e_m then $F + e_m$ is an (i, j)-forest of G, since every other edge precedes e_m and, as e_m is not a bridge, $U_{F+e_m}(e_m)$ has other edges in addition to e_m.

From these it follows that $t_{ij} = t'_{ij} + t''_{ij}$ and so, by the induction hypothesis and a basic property of the Tutte polynomial,

$$\sum_{i,j} t_{ij} x^i y^j = \sum_{i,j} t'_{ij} x^i y^j + \sum_{i,j} t''_{ij} x^i y^j$$
$$= T_{G-e_m}(x, y) + T_{G/e_m}(x, y) = T_G(x, y),$$

as claimed.

This completes the proof of the induction step. □

The theorem above can be taken to be another definition of the Tutte polynomial. As we have already mentioned, precisely this was Tutte's original definition: for a connected graph G with an order on the edges,

$$T_G(x, y) = \sum_{i,j} t_{ij} x^i y^j,$$

where $t_{ij} = t_{ij}(G)$ is the number of spanning trees of G with internal activity i and external activity j. It is trivial that for every order $\prec$ on the edges there is a polynomial $T_\prec$, but it is remarkable that this polynomial $T_\prec$ is *independent of the order* on the edges we take. In proving Theorem 10, we were greatly helped by the earlier definition of the polynomial, that is, by the knowledge that there *is* a polynomial with appropriate properties.

To gain more insight into the nature of internal and external activities, let us see how Theorem 10 can be proved from first principles, without appealing to any of the earlier results.

Second proof of Theorem 10. Let us start with the independence of $T_\prec$ from the order $\prec$ on the edges. To simplify the notation, we shall assume that our graph is connected, as the extension to the general case is trivial.

Let $\prec$ be the order $e_1 \prec e_2 \prec \cdots \prec e_m$, where $E(G) = \{e_1, e_2, \ldots, e_m\}$, and let $\prec'$ be the order obtained from $\prec$ by interchanging e_h and e_{h+1}. Thus in the order $\prec'$ we have $e_1 \prec' e_2 \prec' \cdots \prec' e_{h-1} \prec' e_{h+1} \prec' e_h \prec' e_{h+2} \prec' \cdots \prec' e_{m-1} \prec' e_m$.

Let us define *weights* w and w' on the set $\mathcal{T}$ of spanning trees of G: for $T \in \mathcal{T}$ set $w(T) = x^i y^j$ if T is an (i, j)-tree with respect to $\prec$, and $w'(T) = x^i y^j$ if T is an (i, j)-tree with respect to $\prec'$. Then $T_\prec(G)$ and $T_{\prec'}(G)$ are the sums of the w-weights and w'-weights: $T_\prec(G) = \sum_{\mathcal{T}} w(T)$ and $T_{\prec'}(G) = \sum_{\mathcal{T}} w'(T)$. We shall partition $\mathcal{T}$ into small sets (into sets of sizes one and two) and show that for each set the total w-weight of the set is precisely its total w'-weight. Then, *a fortiori*, $T_\prec(G) = T_{\prec'}(G)$.

The small sets in question are precisely the orbits of an involution $\mathcal{T} \to \mathcal{T}'$, $T \mapsto T'$, given as follows. For $T \in \mathcal{T}$, if one of e_h and e_{h+1} is in the cycle set of the other (that is, $e_h \in Z_T(e_{h+1})$ or $e_{h+1} \in Z_T(e_h)$), then let T' be obtained from T by interchanging e_h and e_{h+1}; otherwise, set $T' = T$. It is immediate that the map $\mathcal{T} \to \mathcal{T}'$ given by $T \mapsto T'$ is indeed an involution, that is, $(T')' = T$.

An edge e different from e_h and e_{h+1} has the same activity for T in $\prec$ as in $\prec'$, so $w(T) = w'(T)$ unless at least one of e_h and e_{h+1} has different activities for T in $\prec$ and $\prec'$, which can happen only if $T \neq T'$. Consequently, in order to prove that $T_\prec(G) = T_{\prec'}(G)$, it suffices to show that

$$w(T) + w(T') = w'(T) + w'(T') \tag{8}$$

when $T \in \mathcal{T}, T \neq T'$.

Suppose that every edge $e \neq e_h,\ e_{h+1}$ has the same activity for T as for T'. Note that it is irrelevant whether we take $\prec$ or $\prec'$, as $e \prec f$ if and only if $e \prec' f$. Since the cuts and cycles are the same, the activities of e_h for T and T' in $\prec$ are the same as the activities of e_{h+1} for T' and T in $\prec'$, so in this case $w(T) = w'(T')$ and $w(T') = w'(T)$. Therefore, we may assume that some edge $e \neq e_h,\ e_{h+1}$ is active for T and inactive for T'. By interchanging the labels of e_h and e_{h+1}, and swapping $\prec$ with $\prec'$, we may also assume that $e_h \in E(T)$, $e_{h+1} \in E(T')$.

Suppose then that $e \in E(T)$, $e \neq e_h$, is active for T but not for T'. Thus e is the minimal element of $U_T(e)$, but the minimal element f of $U_{T'}(e)$ is smaller than e. Thus $U_T(e) \neq U_{T'}(e)$, so $e_{h+1} \in U_T(e)$, and as e is active for T, $e \prec e_{h+1}$. Now, as f is in one of $U_T(e)$, $U_{T'}(e)$ but not the other, $f \in U_T(e_h) = U_{T'}(e_{h+1})$, while $e_{h+1} \in U_T(e)$ is equivalent to $e \in Z_T(e_{h+1}) = Z_{T'}(e_h)$. But then e_h, e_{h+1} are inactive for T, T' in either order, as e, f precede e_h, e_{h+1} in either order, and

e lies in the relevant cycles and f in the relevant cuts. Thus $w(T) = w'(T)$ and $w(T') = w'(T')$.

Finally, suppose that $e \notin E(T)$, $e \neq e_{h+1}$, is active for T but not for T'. Thus e is the minimal element of $Z_T(e)$, but the minimal element f of $Z_{T'}(e)$ is smaller than e. Then $e_h \in Z_T(e)$, i.e., $e \in U_T(e_h) = U_{T'}(e_{h+1})$, and $f \in Z_T(e_{h+1}) = Z_{T'}(e_h)$. Therefore, as before, e_h, e_{h+1} are inactive in all cases, so $w(T) = w'(T)$ and $w(T') = w'(T')$, proving (8). This completes the proof of $T_{\prec}(G) = T_{\prec'}(G)$.

Having proved the independence of t_{ij} of the order, so that $t_{ij} = t_{ij}(G)$ is only a function of the graph G, let us see that the graph polynomial $\tilde{T}(G) = \sum_{i,j} t_{ij}(G)x^i y^j$ satisfies the appropriate recurrence relations and boundary conditions, so $\tilde{T}(G)$ is indeed the Tutte polynomial T_G.

As $t_{00}(E_n) = 1$ and $t_{ij}(E_n) = 0$ if $i + j > 0$, we do have $\tilde{T}(E_n) = 1$.

Now let us turn to the recurrence relations. Let us pick an element of $E(G)$, and let us distinguish three cases according to the nature of this element. We shall greatly benefit from the order independence of $t_{ij}(G)$: given $e \in E(G)$, we may and shall assume that e is the very last element in the order, that is $E(G) = \{e_1, e_2, \dots, e_m\}$ and the edge we are interested in is precisely $e = e_m$.

(i) Suppose that e_m is a bridge. Then e_m is in every spanning forest of G, and it is an active edge in every spanning forest. Hence the spanning forests of $G - e_m$ are in 1–to–1 correspondence with the spanning forests of G, with an (i, j)-forest of $G - e_m$ corresponding to an $(i + 1, j)$-forest of G. But then

$$\tilde{T}(G) = \sum t_{i+1,j}(G)x^{i+1}y^j = x\sum t_{ij}(G - e_m)x^i y^j = x\tilde{T}(G - e_m).$$

(ii) Suppose that e_m is a loop. Then e_m is an externally active edge with respect to every spanning forest of G, so the spanning forests of $G - e_m$ are in 1–to–1 correspondence with the spanning forests of G, with an (i, j)-forest of $G - e_m$ corresponding to an $(i, j + 1)$-forest of G. But then

$$\tilde{T}(G) = \sum t_{i,j+1}(G)x^i y^{j+1} = y\sum t_{ij}(G - e_m)x^i y^j = y\tilde{T}(G - e_m).$$

(iii) Suppose that e_m is neither a bridge nor a loop. Then partition the set $\mathcal{F}_{ij}$ of (i, j)-forests as $\mathcal{F}'_{ij} \cup \mathcal{F}''_{ij}$, with $\mathcal{F}'_{ij} = \{F \in \mathcal{F}_{ij} : e_m \in E(F)\}$. Then $\{F - e_m : F \in \mathcal{F}'_{ij}\}$ corresponds to the set of (i, j)-forests of G/e_m, since e_m is not internally active in any forest $F \in \mathcal{F}'_{ij}$. Also, e_m is not externally active with respect to any forest $F \in \mathcal{F}''_{ij}$, so $\mathcal{F}''_{ij}$ corresponds to the set of (i, j)-forests of $G - e_m$. Consequently,

$$\tilde{T}(G) = \tilde{T}(G - e_m) + \tilde{T}(G/e_m),$$

so we are done. □

The new definition, the spanning forest expansion, has several advantages over the old one, namely

$$T_G(x, y) = S(G; x - 1, y - 1) = \sum_{F \subset E} (x - 1)^{r\langle E\rangle - r\langle F\rangle}(y - 1)^{n\langle F\rangle}.$$

First, the new sum has (usually) many fewer terms than the old one. Secondly, and more importantly, in the new expansion the coefficient of $x^i y^j$, rather than the coefficient of $(x-1)^i(y-1)^j$, is defined explicitly, and turns out to be non-negative. (We do know this from the recursive definition, but that does not give us an explicit expression for the coefficients.) Thirdly, a judicious choice of the *order* on the edges is frequently advantageous in proving results about the coefficients $t_{ij} = t_{ij}(G)$.

As a simple application of Theorem 10, let us see what we can say about the coefficients of the lowest terms, namely t_{00}, t_{10} and t_{01}. First of all, as $T_{E_n}(x, y) = 1$, let us assume that $E(G) = \{e_1, \dots, e_m\}$ with $m \geq 1$, this being also the order on $E(G)$. Note that whatever our spanning forest F is, e_1 is certainly an active edge: internally active if $e_1 \in E(F)$, and externally active otherwise. In particular, $t_{00} = 0$. As, trivially, $t_{10}(K_2) = 1$ and $t_{01}(K_2) = 0$, let us assume that $m \geq 2$. If G has at least two blocks containing at least one edge each then we can choose an order on $E(G)$ such that e_1 and e_2 belong to distinct blocks of G. Then both e_1 and e_2 are active with respect to every spanning forest, so $t_{10} = t_{01} = 0$.

Suppose then that G consists of a 2-connected graph and isolated vertices. Since the addition of isolated vertices does not alter the Tutte polynomial, in our study of t_{10} and t_{01} we may assume that G itself is 2-connected. This is only so that we can call a spade a spade: t_{ij} is the number of (i, j)-trees of G rather than the number of (i, j)-forests. However, as far as the argument is concerned, this is irrelevant.

As noted earlier, e_1 is active with respect to every spanning tree. Furthermore, the edge e_2 is also active unless its cut or cycle (whichever is appropriate) contains e_1. Hence if T is a $(1, 0)$-tree then $e_1 \in E(F)$ and $e_2 \in U_T(e_1)$. Let T^* be obtained from T by interchanging e_1 and e_2: as $e_2 \in U_T(e_1)$ (that is, $e_1 \in Z_T(e_2)$), T^* is also a spanning tree. Clearly, T^* is a $(0, 1)$-tree: its only active edge is e_1, and that edge is externally active.

It is immediate that the process above can be reversed: a $(0, 1)$-tree T does not contain e_1 but contains e_2, and interchanging e_1 and e_2 we get a $(1, 0)$-tree T^*. Hence the map $T \mapsto T^*$ gives a 1–to–1 correspondence between the set of $(1, 0)$-trees and the set of $(0, 1)$-trees. In particular, this gives the following result.

Theorem 11 *Let $\sum_{i,j} t_{ij}x^iy^j$ be the Tutte polynomial of a graph G with at least two edges. Then $t_{10} = t_{01}$.* □

The identity $t_{10} = t_{01}$ is one of an infinite family of identities holding for the coefficients t_{ij} of the Tutte polynomial. Brylawski proved that if $e(G) > h$ then

$$\sum_{i=0}^{h}\sum_{j=0}^{h-i}(-1)^j\binom{h-i}{j}t_{ij} = 0.$$

Thus if $e(G) > 0$ then $t_{00} = 0$; if $e(G) > 1$ then $t_{10} = t_{01}$; if $e(G) > 2$ then $t_{20} - t_{11} + t_{02} = t_{10}$; and so on (see Exercise 8).

In fact, $t_{10} = t_{10}(G)$ is a significant graph invariant in its own right: it is usually called the *chromatic invariant* of G and is denoted by $\theta(G)$. The terminology is justified by the following simple theorem. The first part shows the connection with the chromatic polynomial, and the second shows that θ is invariant under subdivisions of graphs with at least two edges.

Theorem 12 (i) *For every graph G the derivative $p'_G(1)$ of the chromatic polynomial $p_G(x)$ satisfies*

$$p'_G(1) = (-1)^{r(G)+1}\theta(G).$$

(ii) *Let G and H be homeomorphic graphs, each with at least two edges. Then $\theta(G) = \theta(H)$.*

Proof. (i) This is immediate from

$$p_G(x) = (-1)^{r(G)}x^{k(G)}T_G(1-x,\,0) = (-1)^{r(G)}x^{k(G)}\sum_{i=0}^{n-1} t_{i0}(1-x)^i.$$

(ii) We know that G and H have isomorphic subdivisions. Hence it suffices to show that if H is obtained from G by subdividing an edge $e \in E(G)$ into two edges, e_1 and e_2, say, then $\theta(H) = \theta(G)$.

If e is a bridge of G then $\theta(G) = \theta(H) = 0$ since the chromatic invariant of a graph with at least two non-trivial blocks is 0. If e is not a bridge of H then e_1 is neither a bridge nor a loop of H so, from the recursion of the Tutte polynomial,

$$\theta(H) = \theta(H - e_1) + \theta(H/e_1).$$

Clearly, e_2 is a block of $H - e_1$, so this graph has at least two non-trivial blocks, implying $\theta(H - e_1) = 0$. Also, $H/e_1 \cong G$, so we are done. □

The spanning tree expansion also gives some information about the sizes of the coefficients of the chromatic polynomial.

Theorem 13 *Let G be a connected graph of order n, with chromatic polynomial $p_G(x) = \sum_{j=0}^{n-1}(-1)^j a_j x^{n-j}$. Then $a_0 = 1 \le a_1 \le \cdots \le a_l$ for $l = \lfloor n/2 \rfloor$.*

Proof. We know that

$$p_G(x) = (-1)^{n-1}x\sum_{i=0}^{n-1} t_{i0}(-x+1)^i,$$

so

$$(-1)^j a_j = (-1)^{n-1}\sum_{i=n-j-1}^{n-1}(-1)^{n-j-1}t_{i0}\binom{i}{n-j-1},$$

that is,

$$a_j = \sum_{i=n-j-1}^{n-1} t_{i0}\binom{i}{n-j-1}.$$

Hence if $1 \le j \le n/2$ then

$$a_j - a_{j-1} = t_{n-j-1,0} + \sum_{i=n-j}^{n-1} t_{i0}\left\{\binom{i}{n-j-1} - \binom{i}{n-j}\right\} \ge t_{n-j-1,0},$$

since $n-j-1 \ge n/2-1 \ge (i-1)/2$ for all $i \le n-1$, so $\binom{i}{n-j-1} \ge \binom{i}{n-j}$. □

The relations $\deg_x T_G(x, y) = r(G)$ and $\deg_y T_G(x, y) = n(G)$ that we encountered earlier are also immediate from the fact that $t_{ij}(G)$ is the number of spanning forests of internal activity i and external activity j. Furthermore, if G is loopless and F is a spanning forest of G then with respect to an order in which every edge of F comes before every other edge, F has internal activity $r(G)$ and external activity 0. Similarly, if G is bridgeless and F is a spanning forest of G then with respect to an order in which every edge of F comes after every other edge, F has internal activity 0 and external activity $n(G)$. In particular, $t_{r(G),0} \geq 1$ and $t_{0,n(G)}(G) \geq 1$. Consequently, $\max\{i - j : t_{ij}(G) \neq 0\} = r(G)$ and $\max\{j - i : t_{ij}(G) \neq 0\} = n(G)$. By considering less obvious orders on the edges, one can show that several other coefficients are non-zero. We shall give two examples of this.

Theorem 14 *Let G be a 2-connected loopless graph with n vertices and m edges, and let $T_G(x, y) = \sum t_{ij} x^i y^j$. Then $t_{i0} > 0$ for each i, $1 \leq i \leq n-1$, and $t_{0j} > 0$ for each j, $1 \leq j \leq m - n + 1$.*

Proof. Given a spanning tree T, for $E_0 \subset E(G)$ let $\gamma_T(E_0)$ be the set of chords whose cycles meet E_0, together with the set of tree-edges whose cuts meet E_0:

$$\gamma_T(E_0) = \{e \in E(G) : Z_T(e) \cap E_0 \neq \emptyset\} \cup \{e \in E(G) : U_T(e) \cap E_0 \neq \emptyset\}.$$

Note that $\gamma_T(E_0) \supset E_0$. Let $\overline{\gamma}_T(E_0)$ be the *closure* of E_0 with respect to this γ-operation:

$$\overline{\gamma}_T(E_0) = E_0 \cup E_1 \cup \cdots,$$

where $E_{k+1} = \gamma_T(E_k)$. As G is 2-connected, $\overline{\gamma}_T(E_0) = E(G)$ whenever E_0 is a non-empty set of edges (see Exercise 18), so that $E_0 \subset E_1 \subset \ldots \subset E_l = E(G)$ for some l.

(i) For $1 \leq i \leq n - 1$, let E_0 be a set of i edges of T, and let $E_1 = \gamma_T(E_0)$, $E_2 = \gamma_T(E_1)$, and so on. We know that $E_l = E(G)$ for some l. Let $\prec$ be an order compatible with the sequence $E_0 \subset E_1 \subset \cdots \subset E_l = E(G)$, that is, an order in which the edges of E_0 come first, followed by the edges of $E_1 \setminus E_0$, the edges of $E_2 \setminus E_1$, and so on, ending with the edges of $E_l \setminus E_{l-1}$. Then each edge of E_0 is active, and no other edge is active. Hence T is an $(i,\ 0)$-tree in the order $\prec$, so $t_{i0} > 0$.

(ii) For $1 \leq j \leq m - n + 1$ we start with a set E_0 of j chords of T, that is with a set $E_0 \subset E(G) \setminus E(T)$, and proceed as in (i). Once again, the active edges are precisely the edges of E_0, so T is a $(0,\ j)$-tree, proving $t_{0j} > 0$.

To see the last assertion, recall that $\deg_x T_G(x, y) = r(G) = n - 1$ and $\deg_y T_G(x, y) = n(G) = m - n + 1$. □

What about the coefficient t_{11}? By Exercise 1, for a cycle C_n we have $T_{C_n} = y+x+x^2+\cdots+x^{n-1}$, so $t_{11}(C_n) = 0$. Also, for the *thick edge* I_k consisting of two vertices joined by $k \geq 2$ edges, by Exercise 2 we have $T_{I_k} = x+y+y^2+\cdots y^{k-1}$, so $t_{11}(I_k) = 0$ as well. As we shall show now, for every other 2-connected loopless graph G we have $t_{11} \geq 1$.

Theorem 15 *Let G be a 2-connected loopless graph that is neither a cycle nor a thick edge. Then $t_{11}(G) > 0$.*

Proof. It is easily seen that G contains a cycle C and an edge e_1 joining a vertex of C to a vertex not on C. Let T be a spanning tree that contains e_1 and all edges of C except for an edge e_2, and set $E_0 = \{e_1, e_2\}$. Let $E_0 \subset E_1 \subset \ldots \subset E_l = E(G)$ be as in the proof of Theorem 14, with $E_{k+1} = \gamma_T(E_k)$, and let $\prec$ be an order compatible with this nested sequence. It is immediate that, with respect to this order, T has precisely one internally active edge, namely e_1, and precisely one externally active edge, namely e_2. Hence $t_{11}(G) > 0$, as claimed. □

Read conjectured in 1968 that the sequence of moduli of the coefficients of the chromatic polynomial is *unimodal*, i.e., with the notation of Theorem 13, $a_0 \leq a_1 \leq \cdots \leq a_m \geq a_{m+1} \geq \cdots \geq a_{n-1}$ for some m. This conjecture is still open, although Theorem 13 goes some way towards proving it. A related conjecture of Tutte, stating that the t_{ij} form unimodal sequences in i and j separately, and the analogous conjecture of Seymour and Welsh for matroids, were disproved by Schwärzler in 1993.

X.6 Polynomials of Knots and Links

Knots and links as mathematical objects were introduced by Listing in 1847 and, independently, by Thomson in 1869. In his paper on vortex motion, Thomson suggested that in order to understand space properly, we have to investigate knots and links (see Fig. X.3). The challenge was taken up by Thomson's collaborator, Tait, in a lecture delivered in 1876, and in a subsequent series of papers. In the 1880s, Tait, Kirkman and Little attempted to give a census of knots with at most ten crossings. However, knot theory proper was really started only in the 1920s, with the work of Dehn, Alexander and Reidemeister, who introduced a variety of knot invariants. These knot invariants enabled Alexander and Briggs to complete the rigorous classification of knots with up to nine crossings. Curiously, a mistake in Little's table was corrected by Perko, an amateur mathematician, only in 1974, when he showed that two knots with ten crossings claimed to be different

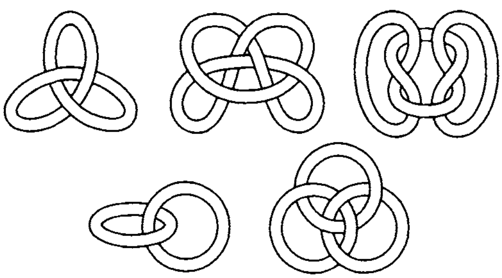


FIGURE X.3. Knots and links from Thomson's 1869 paper; the first knot is the (right-handed) trefoil knot, the two links are the Hopf link and the Borromean rings.

by Little were, in fact, equivalent. The most commonly used notation for knots is still the one introduced by Alexander and Briggs.

All this belongs to the classical period of knot theory, and we shall say very little about it. Our interest in knot theory stems from the fact that in the 1980s Jones started a revolution in knot theory by introducing a new polynomial invariant, which later was shown to be closely related to the Tutte polynomial. Our aim in this section is to introduce this polynomial together with some related polynomials and to indicate their connection to the Tutte polynomial.

Although in Section VIII.1 we fleetingly touched upon knots, when we discussed their fundamental groups, for this section we have to set the scene with a little more care. A *link* L of n components is a subset of $\mathbb{R}^3 \subset \mathbb{R}^3 \cup \{\infty\} = S^3$, consisting of n disjoint piecewise linear simple closed curves. A *knot* is a connected link. The 3-dimensional sphere S^3 is always oriented; occasionally the components of L are also oriented, in which case we have an *oriented link*. We demand that our links are piecewise linear only to avoid infinite sequences of twists; it would be equally good to assume that our links are smooth submanifolds of S^3 - in fact, in our drawings we shall always follow this convention (see Fig. X.4).

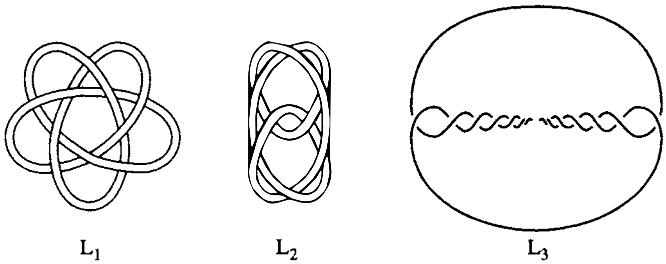


FIGURE X.4. The knots L_1 and L_2 are equivalent; L_3 is not a (tame) link.

Two links L_1 and L_2 are *equivalent* if there exists an orientation-preserving homeomorphism $h : \mathbb{R}^3 \to \mathbb{R}^3$ such that $h(L_1) = L_2$. If L_1 and L_2 are oriented then $h(L_1)$ must be oriented the same way as L_2. Equivalence can also be defined in terms of homotopy: two links are said to be *equivalent* or *ambient isotopic* if they can be deformed into each other. For example, if L_1 and L_2 are knots given by piecewise-linear maps $h_i : [0, 1] \to \mathbb{R}^3$ with $h_i(0) = h_i(1)$, $i = 0, 1$, then L_1 and L_2 are equivalent iff there is a piecewise-linear map $h : [0, 1]^2 \to \mathbb{R}^3$ such that $h_0(x) = h(0, x)$, $h_1(x) = h(1, x)$, and $h_t(x) = h(t, x)$ gives a knot for every t, that is $h(t, 0) = h(t, 1)$ for every t.

In fact, no topological subtlety need enter knot theory, as it is very easy to view links as purely combinatorial objects. What we do is consider *link diagrams*: *projections* of links into $\mathbb{R}^2$. We may assume that the projection is a 4-regular plane multigraph with loops, with a loop adding 2 to the degree of its vertex. At each vertex the edges form strictly positive angles so that the projections of the two parts of the link 'cross cleanly', rather than touch. Thus a *link diagram* is a (finite) 4-regular plane graph with some extra structure, namely at each vertex the

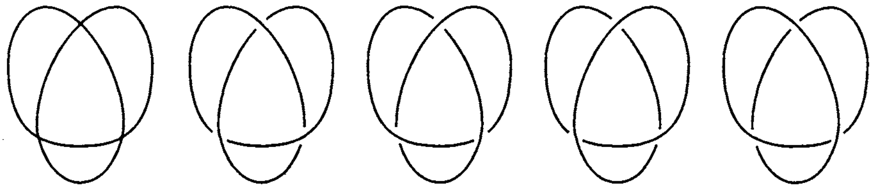


FIGURE X.5. Diagrams arising from the universe U of a triangle with double edges; there are four more that arise from the last two by rotation. The first is the standard diagram of the right-handed trefoil knot, and the second is the standard diagram of the left-handed trefoil knot.

two pairs of edges cross in one of two ways: one goes either under or over the other (see L_1 in Fig. X.4). Note that a 4-regular plane graph of order n gives rise to 2^n link diagrams; the 4-regular graph is the *universe* of the link diagrams that arise from it (see Fig. X.5).

Throughout our study of link diagrams, we do not distinguish between planar isotopic link diagrams; thus two link diagrams are considered to be the same if they come from isomorphic *plane* graphs with the same extra structure. For the sake of convenience, the same letter is used for a link and its diagram.

Reidemeister proved in the 1920s that two links are ambient isotopic iff their diagrams can be transformed into each other by planar isotopy and the three *Reidemeister moves* illustrated in Fig. X.6. The Reidemeister moves are used also to define a finer equivalence of link diagrams: two link diagrams are said to be *regularly isotopic* if they can be transformed into each other by planar isotopy and moves of Types II and III.

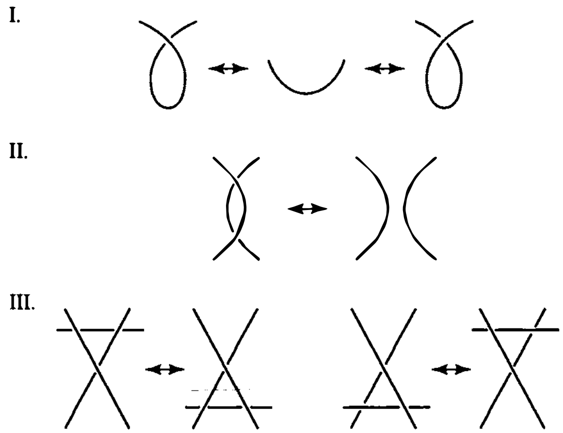


FIGURE X.6. The three Reidemeister moves. A move of Type I adds or removes a curl, a move of Type II removes or adds two consecutive undercrossings or overcrossings, and a move of Type III, a triangle move, changes the position of two undercrossings or overcrossings.

Ideally, one would like to find a simple and complete classification of knots and links. In theory, knots were classified by Haken in 1962, but that classification is in terms of a very elaborate algorithm which is too unwieldy to use in practice.

Lowering our sights a little, instead of trying to come up with a simple complete classification, we may try to introduce link invariants that are fairly simple to calculate and yet are fine enough to distinguish 'many' knots and links. This is the approach we shall adopt here. In particular, we would like our invariants to help us to answer the following basic questions. Is a link of several components really *linked* or is it (equivalent to) a link with a diagram having at least two components? Is a knot really *'knotted'* or is it (equivalent to) the *unknot*, the trivial knot whose diagram has no crossings? More generally, does a link have a diagram with no crossings? Is a link equivalent to its mirror image?

A trivial invariant of links is the number of components: it is easily read out of a diagram of a link. The non-trivial invariants we shall find are defined as invariants of link diagrams, and they are all based on examining and possibly changing a crossing in a link diagram, while keeping the diagram unchanged outside a small neighbourhood of this crossing.

Some beautiful combinatorial invariants of links are based on colouring their *'strands'*, the arcs from one undercrossing to another. Although this approach is outside the main thrust of this section, let us describe briefly the simplest of these invariants. Call a 3-colouring of the strands of a diagram *proper* if at no crossing do we find precisely two colours. A 3-colouring is *non-trivial* if at least two colours are used (see Fig. X.7). Now it is easily seen that if a Reidemeister move transforms a diagram L into L', and L has a non-trivial proper 3-colouring, then so does L'. Hence the existence of a non-trivial proper 3-colouring is an ambient isotopy invariant. The unknot does not have a non-trivial proper 3-colouring, and therefore a knot whose diagram has a non-trivial proper 3-colouring is not equivalent to the unknot. In particular, the knot in X.7 is knotted, and so is the (right-handed) trefoil knot in Fig.X.3 (and also in Fig. VIII.4). For extensions of these ideas, see Exercises 25-33.

But let us turn to the main thread of this section.

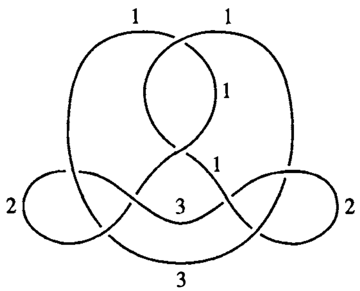


FIGURE X.7. A non-trivial proper 3-colouring of a knot diagram (namely, of a diagram of the knot 8_{15} in the Alexander-Briggs notation).

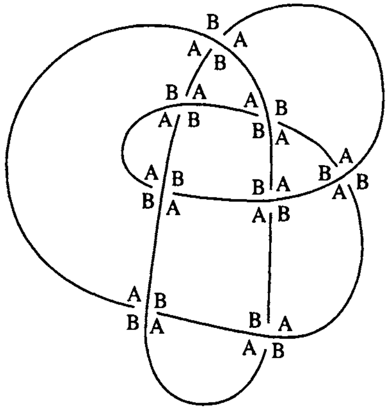
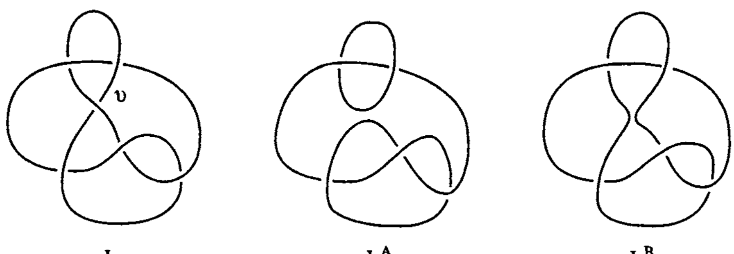


FIGURE X.8. The A and B regions in a diagram of the knot 8_{19}.

The Tutte polynomial was defined by 'resolving' an edge in two different ways: deleting it and contracting it. There are also two ways of resolving a crossing in an (unoriented) link diagram, as we shall see now. Every unoriented crossing distinguishes two out of the four regions incident at its vertex. Rotate the over-crossing arc counterclockwise until the under-crossing arc is reached, and call the two regions swept out the *A regions* and the other two the *B regions* (see Fig. X.8). Sometimes instead of regions one talks about *channels.*

What are the two ways of resolving a crossing? We may slice it open at the A regions, so that the two A regions unite, or we may slice it open at the B regions, so that the two B regions unite. Briefly, at every crossing we may have an A*-slice* or a B*-slice.*

As a self-explanatory shorthand, let us write $\times$ for a link diagram, with emphasis on a particular crossing in it. After an A-slice we get $\asymp$, after a B-slice we get)(. If we wish to be a little more rigorous (and avoid typographical difficulties), given a link diagram L with a crossing at v, we write L_v^A for the link diagram obtained from L by an A-slice at v, and L_v^B for the link diagram obtained by a B-slice at v (see Fig. X.9).



L $\qquad$ L_v^A $\qquad$ L_v^B

FIGURE X.9. A diagram L (of the knot 6_3) with a crossing at v, and the diagrams L_v^A and L_v^B.

Imitating the Tutte polynomial and its variants, we shall define a polynomial, the *Kauffman bracket*, or just *bracket*, whose value on a link diagram L is a fixed linear combination of its values on L_v^A and L_v^B. The polynomial is uniquely determined by this and some natural boundary conditions but, just like the Tutte polynomial, it can also be given explicitly. In order to give this explicit expression, we need some more definitions.

A *state* S of a link diagram is a choice of *slicing* for each crossing of the diagram; thus a diagram with n crossings has 2^n states. The S-splitting of a link diagram L is the result of slicing all crossings according to the state S; clearly a splitting is a collection of unlinked trivial knots, that is a link without any crossings (see Fig. X.10). Writing V for the set of crossings, a state S is a function $S : V \to \{A,\ B\}$, with $S(v) = A$ meaning that at v we take an A-slice. Thus the set of all states is $\{A,\ B\}^V$.

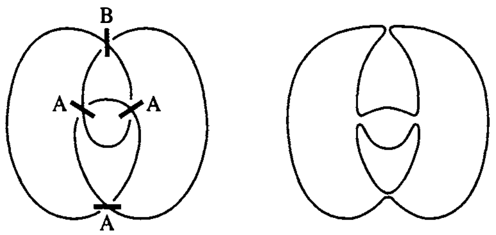


FIGURE X.10. A state S of the standard diagram, K, of the figure of eight, and the splitting it gives; $a_K(S) = 3$, $b_K(S) = 1$ and $c_K(S) = 2$.

Given a link diagram L, and a state S of L, write $a_L(S)$ for the *number of A-slices* in S, and $b_L(S)$ for the *number of B-slices*. Also, let $c_L(S)$ be the *number of components* of the S-splitting of L.

After all this preparation, the *Kauffman square bracket* $[L]$ of a link diagram L, with value in $\mathbb{Z}[A,\ B,\ d]$, is easily defined:

$$[L] = \sum_S A^{a_L(S)} B^{b_L(S)} d^{c_L(S)-1},$$

where the summation is over all states of the diagram L.

The basic properties of the Kauffman square bracket are given in the theorem below, reminiscent of Theorem 2. Let us write $\bigcirc$ for a connected link diagram without crossings, and $L_1 \cup L_2$ for a diagram which is a disjoint union of L_1 and L_2. Let $\mathcal{L}$ be the set of link diagrams.

Theorem 16 *There is a unique map $\varphi : \mathcal{L} \to \mathbb{Z}[A,\ B,\ d]$ such that*

(i) if L and L' are planar homotopic link diagrams then $\varphi(L) = \varphi(L')$,

(ii) $\varphi(\bigcirc) = 1$,

(iii) $\varphi(L \cup \bigcirc) = d\varphi(L)$ for every link diagram L,

(iv) $\varphi(L) = A\varphi\left(L_v^A\right) + B\varphi\left(L_v^B\right)$ for every link diagram L with a crossing at v. Furthermore, $\varphi(L) = [L]$.

Proof. It is clear that conditions (i) - (iv) determine a *unique* map, *if* there is such a map. Hence all we have to check is that the Kauffman square bracket [.] has properties (i)-(iv). The first three are immediate from the definition.

Property (iv) is also almost immediate. Indeed, let v be a crossing of L. Then, writing $L' = L_v^A$ and $L'' = L_v^B$,

$$\begin{aligned}
[L] &= \sum_S A^{a_L(S)} B^{b_L(S)} d^{c_L(S)-1} \\
&= \sum_{S,\, S(v)=A} A^{a_L(S)} B^{b_L(S)} d^{c_L(S)-1} + \sum_{S,\, S(v)=B} A^{a_L(S)} B^{b_L(S)} d^{c_L(S)-1} \\
&= A \sum_{S'}{}' A^{a_{L'}(S')} B^{b_{L'}(S')} d^{c_{L'}(S')-1} + B \sum_{S''}{}'' A^{a_{L''}(S'')} B^{b_{L''}(S'')} d^{c_{L''}(S'')-1} \\
&= A\left[L_v^A\right] + B\left[L_v^B\right],
\end{aligned}$$

where $\sum_{S'}'$ denotes summation over all states S' of L' and $\sum_{S''}''$ denotes summation over all states S'' of L''. □

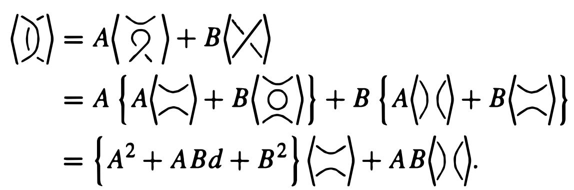
An equivalent form of Theorem 16 is that if R is a commutative ring and $A, B, d \in R$ then the Kauffman square bracket with parameters A, B and d satisfies conditions (i)-(iv). In general, the square bracket is far from being invariant under ambient isotopy or even regular isotopy, but if A, B and d satisfy certain conditions then it is a regular isotopy invariant. To be precise, define the *Kauffman angle bracket* or simply *Kauffman bracket* $\langle L\rangle \in \mathbb{Z}[A, A^{-1}]$ of a link L by setting

$$\langle L\rangle(A) = [L](A, A^{-1}, -A^2 - A^{-2}).$$

Thus $\langle L\rangle$ is a *Laurent polynomial* in A, and it is simply the evaluation of $[L]$ at A, B and d satisfying the conditions $AB = 1$ and $d = -A^2 - A^{-2}$.

Lemma 17 *The Kauffman bracket is invariant under regular isotopy.*

Proof. Let B and d be as above, so that $AB = 1$ and $d = -A^2 - A^{-2}$ and, under these conditions, $\langle L\rangle(A) = [L](A, B, d)$. First, let us evaluate the effect of a Type II move on the angle bracket by resolving crossings by (iv) and applying (iii):



As $AB = 1$ and $A^2 + ABd + B^2 = 0$, the right-hand side is $\langle\,)(\,\rangle$, so the bracket is invariant under Type II moves.

To complete the proof, we shall show that Type II invariance implies Type III invariance. Indeed, by (iv),

and

and the two right-hand sides are equal by Type II invariance. Invariance under the other Type III move is checked similarly. □

As it happens, it is easy to alter slightly the Kauffman (angle) bracket $\langle L\rangle$ to turn it into an ambient isotopy invariant of links. In order to do this, we make use of a simple invariant of *oriented* link diagrams. Oriented crossings can be assigned values of ± 1, usually called *signs*, according to the rules in Fig. X.11. The sum of the signs of all the crossings in an oriented link diagram L is the *twist number* or *writhe* $w(L)$ of L. The writhe is *not* an ambient isotopy invariant, but it *is* a *regular isotopy invariant*, as can be seen instantly by inspecting the effect of the Reidemeister moves of Types II and III.

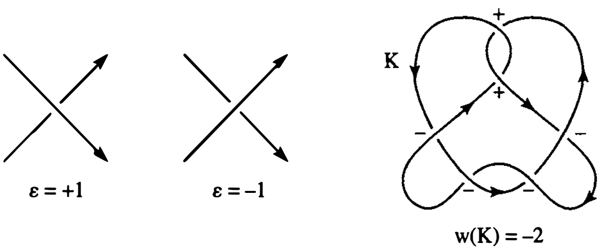


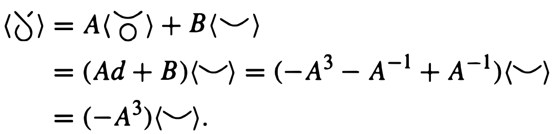
FIGURE X.11. The convention of signs and the writhe of a diagram (of the knot 6_1).

In fact, as we are mainly interested in unoriented links rather than oriented ones, we shall use the self-writhe rather than the writhe. Given an oriented link diagram L with components $L_1, \ldots, L_k$, the *self-writhe* of L is $s(L) = w(L_1) + \ldots + w(L_k)$. Since the sign of a crossing does not change if we change the orientations of both arcs at the crossing, the self-writhe is *independent of the orientation*. Therefore we may define the self-writhe $s(L)$ of an *unoriented link diagram* as the self-writhe of any orientation of L. As the next result shows, if we multiply $\langle L\rangle$ by a simple function of the self-writhe then we obtain an ambient isotopy invariant.

Theorem 18 *The Laurent polynomial* $f[L] = (-A)^{-3s(L)}\langle L\rangle \in \mathbb{Z}[A, A^{-1}]$ *is an invariant of ambient isotopy for unoriented links.*

Proof. Since $s(L)$ and $\langle L\rangle$ are invariants of regular isotopy, so is $f[L]$. Thus all we have to check is that $f[L]$ is invariant under Type I Reidemeister moves.

Note first that



A similar expansion gives

$$\langle \curlyvee_- \rangle = (-A^{-3})\langle \smile \rangle.$$

Since $s(\curlyvee_+) = s(\smile)+1$ and $s(\curlyvee_-) = s(\smile)-1$, independently of the orientation, the Laurent polynomial $f[L]$ is invariant under Type I moves as well:

$$\begin{aligned} f[\curlyvee_+] &= (-A)^{-3s(\curlyvee_+)}\langle \curlyvee_+ \rangle = (-A)^{-3\{s(\smile)+1\}}(-A^3)\langle \smile \rangle \\ &= (-A)^{-3s(\smile)}\langle \smile \rangle = f[\smile] \end{aligned}$$

and, analogously, $f[\curlyvee_-] = f[\smile]$. □

Similar functions can be defined for oriented links. Thus, the Kauffman bracket $\langle L \rangle$ of an oriented link L is simply the Kauffman bracket of the link without its orientation. This is again a regular isotopy invariant of oriented links: to turn it into an ambient isotopy invariant $f[L]$, we usually multiply it by $(-A)^{-3w(L)}$ rather than by $(-A)^{-3s(L)}$.

The Laurent polynomial $f[L]$, which we shall call the *one-variable Kauffman polynomial* of a link or an oriented link, is perhaps the nicest of many closely related link polynomials. The first member of this family, the *Jones polynomial* $V_L(t)$, constructed by Vaughn Jones in 1985, is an ambient isotopy invariant of oriented links defined by the identities $V_{\bigcirc} = 1$ and

$$t^{-1}V_{\times_+} - tV_{\times_-} = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right)V_{\asymp}. \tag{9}$$

It is easily seen that there is at most one Laurent polynomial $V_L \in \mathbb{Z}[t^{1/2}, t^{-1/2}]$ satisfying the relations above: the problem is to show that there *is* such a polynomial. In fact, a simple change of variable turns $f[L]$ into the Jones polynomial.

Theorem 19 *The Jones polynomial $V_L(t)$ of an oriented link L is given by $V_L(t) = f[L]\left(t^{-1/4}\right)$, where $f[L] = (-A)^{-3w(L)}\langle L \rangle(A)$.*

Proof. Since $f[\bigcirc] = 1$, all we have to check is that $f[L]\left(t^{-1/4}\right)$ satisfies (9). By property (iv) of the bracket polynomial, as $B = A^{-1}$ we have

$$\langle \times_+ \rangle = A\langle \asymp \rangle + A^{-1}\langle)(\rangle$$

and

$$\langle \times_- \rangle = A\langle)(\rangle + A^{-1}\langle \asymp \rangle.$$

Hence

$$A\langle \times_+ \rangle - A^{-1}\langle \times_- \rangle = (A^2 - A^{-2})\langle \asymp \rangle,$$

and so

$$\begin{aligned} A^4 f[\times_+] - A^{-4} f[\times_-] = {} & A^4(-A)^{-3(w(\asymp)+1)}\langle \times_+ \rangle \\ & - A^{-4}(-A)^{-3(w(\asymp)-1)}\langle \times_- \rangle \end{aligned}$$

$$= (-A)^{-3w(\;)} \left\{ -A\langle\;\rangle + A^{-1}\langle\;\rangle \right\}$$
$$= \left(A^{-2} - A^{2}\right) f[\;].$$

On substituting $A = t^{-1/4}$, we find that

$$t^{-1} f[\;] - t f[\;] = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right) f[\;]$$

as required. □

The Jones polynomial, especially in its form as the one-variable Kauffman polynomial, can be used to show the inequivalence of many knots and links. For example, the bracket of the right-handed trefoil knot is $A^{-7} - A^{-3} - A^{5}$, and as its writhe is $+3$, its Kauffman polynomial is $-A^{-16} + A^{-12} + A^{-4}$. In particular, the right-handed trefoil knot is not (equivalent to) the unknot (the trivial knot): it is *knotted*. As it happens, this particular assertion is easier proved by other means, for example, by the colouring argument mentioned above (see Exercise 26). However, as we shall see in a moment, the Kauffman polynomial of the right-handed trefoil knot can also be used to show that the right-handed trefoil knot is *not amphicheiral* (also said to be *chiral*): it is not equivalent to its mirror image, the left-handed trefoil knot.

Given a link diagram L, let us write L^* for its *reflection* or *mirror image* obtained by reversing all the crossings (and keeping the orientation the same if L is oriented).

Theorem 20 *The bracket and one-variable Kauffman polynomial of the mirror image L^* of a link diagram L are*

$$\langle L^*\rangle(A) = \langle L\rangle(A^{-1})$$

and

$$f_{L^*}[A] = f_L[A^{-1}].$$

The same holds for oriented link diagrams.

Proof. Note that reversing all the crossings results in swapping A and B, that is A and A^{-1}, in the expansion of the bracket. Hence $\langle L^*\rangle(A) = \langle L\rangle(A^{-1})$. Also, $s(L^*) = -s(L)$ and $w(L^*) = -w(L)$, so the second assertion follows. □

As the one-variable Kauffman polynomial of the trefoil knot does not remain invariant if A is replaced by A^{-1}, the trefoil knot is not amphicheiral, as claimed. (It is for this reason that we have to distinguish between a right-handed trefoil knot and a left-handed trefoil knot.)

The pronounced similarity between 'resolving' an edge of a graph in defining the Tutte polynomial and 'resolving' a crossing of a link diagram in defining the Kauffman bracket is far from superficial: the link polynomials described above can easily be obtained from the Tutte polynomials of certain coloured graphs associated with the diagrams. As we shall see, this connection is especially striking in the case of the so-called alternating link diagrams since then we need only the Tutte polynomials of uncoloured graphs.

To conclude this chapter, we discuss the correspondence between link diagrams and signed plane graphs, introduce alternating diagrams and show that for an alternating diagram the Kauffman bracket is easily expressed in terms of the Tutte polynomial of the associated graph.

Recall that the map of the universe of a link is two-colourable (see Exercise V.24). It is customary to take a black and white colouring of the faces, call those coloured black *shaded*. A *shaded link diagram* is a link diagram with such a proper two-colouring of the faces, i.e., with alternate regions shaded. Note that every plane diagram has precisely two shadings. Also, in the neighbourhood of every crossing there are two shaded regions and two unshaded regions, although these regions are not necessarily different (see Fig. X.12). To each connected shaded plane diagram D we associate an edge-coloured multigraph $G(D)$ as follows. For each shaded face F, take a vertex v_F in F, and for each crossing at which F_1 and F_2 meet, take an edge $v_{F_1}v_{F_2}$. Thus, if $F_1 = F_2 = F$ then the crossing contributes a loop at v_F. Furthermore, colour each edge $+$ or $-$ according to the type of the crossing, as shown in Fig. X.12. We call the graph obtained a *signed* plane graph.

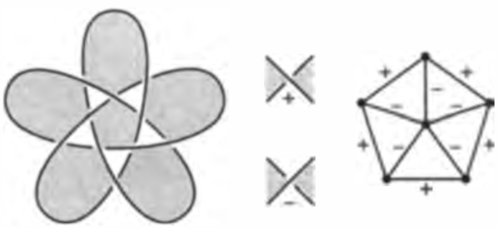


FIGURE X.12. A shading of the first diagram in Fig. X.4, the two types of crossings, and the signed graph associated to the shaded diagram.

Conversely, every connected shaded plane diagram can be reconstructed from $G(D)$. To see this, we just construct the medial graph of $G(D)$, and assign crossing information to it according to the colouring of $G(D)$. The medial graph is particularly easily defined for a plane graph in which every face is a polygon with at least three sides. For such a plane graph G, the *medial graph* $M(G)$ of G is obtained by inserting a vertex on every edge of G, and joining two new vertices by an edge lying in a face of G if the vertices are on adjacent edges of the face. Thus $M(G)$ is a 4-regular plane graph whose alternate faces contain vertices of G. The construction of $M(G)$ is similar for any plane multigraph, as illustrated in Fig. X.13.

Now, given a signed plane graph G with medial graph $M(G)$, shade those faces of $M(G)$ that contain vertices of G. To turn $M(G)$ into a link diagram $D = D(G)$, define the crossings to be over or under according to the colour of the edge at that crossing (see Fig. X.14). For disconnected diagrams, the correspondence is a little more complicated, and we shall not go into it.

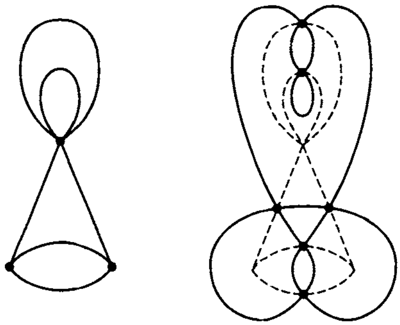


FIGURE X.13. A multigraph and its medial graph.

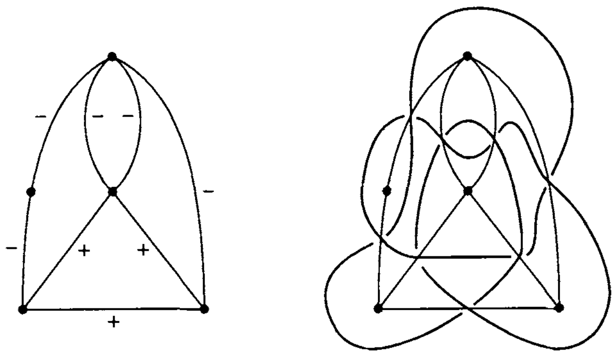


FIGURE X.14. A signed plane graph, and its link diagram, a diagram of 8_{20}.

Call two signed plane graphs *equivalent* if they are the signed graphs of equivalent link diagrams. By Reidemeister's Theorem, two signed plane graphs are equivalent if they can be obtained from each other by a sequence of transformations corresponding to Reidemeister moves; as our graphs are signed, there are, in fact, six so-called *graph Reidemeister moves.*

Having reduced the study of (equivalence classes of) link diagrams to the study of (equivalence classes of) signed plane graphs, we are interested in invariants of signed plane graphs which are constant on equivalence classes. The simplest ways of assigning signs to the edges of a plane graph are making them all $+$ or making them all $-$. The diagrams corresponding to these assignments are said to be *alternating.* Equivalently, a link diagram is alternating if its crossings alternate as one travels along the arcs of the link: over, under, over, under, ... Thus the diagrams of the trefoil and the figure of eight in Fig. VIII.4 are both alternating, and so are the diagrams of the Hopf link and the Borromean rings in Fig. X.3. However, the diagrams of the two knots in Fig. X.4 are not alternating.

It is easily seen that every 4-regular plane multigraph is the universe of an alternating link diagram (see Exercise V.24). Also, if L is a diagram of a k-component link then the universe of L is the universe of at least 2 and at most 2^k alternating diagrams.

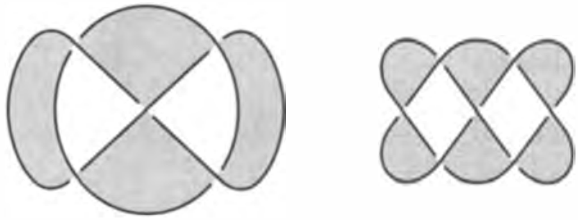


FIGURE X.15. The shaded diagrams of the Whitehead link and of the knot 7_4.

Clearly, a connected link diagram is alternating if, and only if, each of its regions has only A-channels or only B-channels. Calling a region an *A-region* if all its channels are A-channels, and a *B-region* if all its channels are B channels, there are two ways of shading a connected alternating link diagram L: we may shade all the A-regions, or we may shade all the B-regions. Let us write $G^+(L)$ for the graph obtained from the first shading, and $G^-(L)$ for the second. As a signed graph, $G^+(L)$ will be taken with $+$ signs, and $G^-(L)$ with $-$ signs.

Conversely, given a connected plane graph G, let $D^+(G)$ be the alternating link diagram obtained from G by taking each edge with $+$, and let $D^-(G)$ be obtained by taking each edge with $-$. By construction, for every connected alternating link diagram L we have $D^+(G^+(L)) = D^-(G^-(L)) = L$.

By a careful examination of the effect of the resolution of a crossing on the associated signed graph, from Theorem 16 one can show that for an alternating diagram the Kauffman polynomial and so the Jones polynomial are determined by the Tutte polynomial of the associated graph. In particular, one can prove the following result.

Theorem 21 *Let L be a connected alternating oriented link diagram with a A-regions, b B-regions, and writhe w. Then the Jones polynomial of L is given by the Tutte polynomial of $G^+(L)$:*

$$V_L(t) = (-1)^w t^{(b-a+3w)/4} T_{G^+(L)}(-t, -1/t). \qquad \square$$

A crossing is said to be an *isthmus* or a *nugatory crossing* if some two of the local regions appearing at the crossing are parts of the same region in the whole diagram, as in Fig. X.16. A nugatory crossing appears as a loop or bridge in $G^+(L)$ and $G^-(L)$; for example, in the diagram in Fig. X.16, the crossing at v gives a bridge of $G^+(L)$ and a loop of $G^-(L)$. As nugatory crossings make no contribution to 'knottedness', it is preferable to study diagrams without any

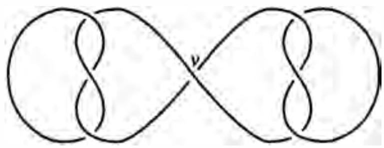


FIGURE X.16. A diagram with a nugatory crossing at v.

of these crossings. The following result was conjectured by Tait, and proved by Murasugi and Thistlethwaite independently about one hundred years later.

Theorem 22 *The number of crossings of a connected alternating link diagram without nugatory crossings is an ambient isotopy invariant.*

Proof. Let L be a connected alternating link diagram with m crossings, none of which is nugatory. We claim that m is precisely the *breadth* of the Laurent polynomial $V_L(t)$, i.e. the difference between the maximum degree and the minimum degree. As the Jones polynomial is ambient isotopy invariant, this is, in fact, more than our theorem claims.

To prove our claim, denote by $a = a(L)$ the number of A-regions. Then $G = G^+(L)$ has a vertices and m edges; also, there are no loops or bridges since L has no nugatory crossings. By Theorem 21, the breadth of $V_L(t)$ is

$$\begin{aligned}\text{breadth } V_L(t) &= \max\deg V_L(t) - \min\deg V_L(t)\\ &= \max\{i - j : t_{ij}(G) \neq 0\} - \min\{i - j : t_{ij}(G) \neq 0\}\\ &= (a-1) - (-m + a - 1) = m,\end{aligned}$$

as claimed. The penultimate equality followed from Theorem 14. □

In fact, it is clear from the proof that if a connected alternating diagram L has m crossings, m' of which are nugatory, then $m - m'$ is the breadth of the Jones polynomial $V_L(t)$, so $m - m'$ is an ambient isotopy invariant.

In fact, using similar methods, Murasugi and Thistlethwaite proved another classical conjecture of Tait: every alternating link has an alternating link diagram with the minimal number of crossings.

To conclude this section, let us remark that, in over a century, knot theory has come full circle. It started out with Thomson's hope of applying it to the study of space, and now it is of great importance in the study of the knots formed by DNA molecules, in the synthesis of various knotted molecules, and, through the Tutte polynomial, in statistical mechanics and topological quantum field theory.

X.7 Exercises

1. Use the contraction-deletion formula to compute the Tutte polynomial of the n-cycle: $T_{C_n}(x, y) = y + x + x^2 + \cdots + x^{n-1}$. Note that this holds for $n = 1$ and 2 as well, with the appropriate interpretation of C_1 and C_2.

2. Use the contraction-deletion formula to show that $T_{I_k} = x + y + y^2 + \cdots + y^{k-1}$, where I_k is the thick edge consisting of two vertices joined by k edges.

3. Let T_{k_1, k_2, k_3} be the 'thick triangle' consisting of three vertices and $k_1 + k_2 + k_3$ edges, with k_1, k_2 and k_3 edges joining the three pairs of vertices. Show that if $k_1, k_2, k_3 \geq 1$ then the Tutte polynomial of T_{k_1, k_2, k_3} is

$$x^2 - 2x + x\{y^{k_1} + y^{k_2} + y^{k_3} - 3\}/(y - 1)$$

$$+ \{y^{k_1+k_2+k_3} - y^{k_1+1} - y^{k_2+1} - y^{k_3+1} + y^2 + y\}/(y-1)^2.$$

4. Show that if $B_1, B_2, \ldots, B_\ell$ are the blocks of a graph G with $e(G) > 0$ then
$$T_G(x, y) = \prod_{i=1}^{\ell} T_{B_i}(x, y).$$

5. Check that the universal polynomial U satisfies formulae (3)-(5) when at least one of σ and τ is zero.

6. Check that the dichromatic polynomial $Z_G(q, v)$ introduced after Theorem 2 can be obtained from U and T, as claimed there.

7. Show that for each enumeration of the edges of G there are spanning trees T_1 and T_2 such that every edge of T_1 is internally active in T_1, and every edge of G not in T_2 is externally active in T_2.

8. Show that if G is a graph with at least three edges then its tree-numbers $t_{ij}(G)$ satisfy $t_{20} - t_{11} + t_{02} = t_{10}$.

9. Let x and y be distinct vertices of a graph G_0, and let u and v be distinct vertices of a graph G_1 which is vertex-disjoint from G_0. Let G' be obtained from $G_0 \cup G_1$ by identifying x with u and y with v, and let G'' be obtained from $G_0 \cup G_1$ by identifying x with v and y with u. Prove that $T_{G'} = T_{G''}$.

10. An orientation of a graph is *totally cyclic* if every edge is contained in some oriented cycle. Prove that the number of totally cyclic orientations of a bridgeless graph is $T_G(0, 2)$.

11. Let G be a graph with vertex set $\{v_1, v_2, \ldots, v_n\}$. Given an orientation of G, let s_i be the *score* of v_i: the number of edges incident with v_i that are directed *away* from v_i, and let $\bar{s} = (s_i)_1^n$ be the *score vector* of the orientation. Show that the total number $s(G)$ of score vectors is $T_G(2, 1)$, the number of forests in G.
[*Hint*. Show first that the set $S(G)$ of score vectors is "convex" in the following sense: if $(s_1, s_2, s_3, \ldots, s_n)$ and $(s_1', s_2', s_3, \ldots, s_n)$ are score vectors with $s_1' > s_1$ and so $s_2' < s_2$, then $(s_1 + 1, s_2 - 1, s_3, \ldots, s_n)$ is a score vector as well. Use this to prove that if $e \in E(G)$ is neither a loop nor a bridge then $s(G) = s(G - e) + e(G/e)$.]

12. Let G be a graph on which there is a nowhere-zero $\mathbb{Z}_k$-flow. Show that there is also a nowhere-zero $\mathbb{Z}$-flow such that in each edge the value of the flow is at most $k - 1$ in modulus. Deduce that if G has a nowhere-zero k-flow then it also has a nowhere-zero $(k + 1)$-flow.

13. Show that the Petersen graph does not have a nowhere-zero 4-flow. [*Hint.* The edge-chromatic number of the Petersen graph is 4.]

14. Let G be a connected plane graph (with multiple edges and loops), with edges $e_1, e_2, \ldots, e_m$ and faces $F_1, F_2, \ldots, F_q$. The *dual* G^* of G has vertices

$v_1, v_2, \ldots, v_q$ and edges $f_1, f_2, \ldots, f_m$, with f_i joining v_j to v_k if F_j and F_k have e_i in their boundaries (see Fig. X.17). Check that the dual of the cycle C_n is the thick edge I_n. What can you say about (the graphs of) the Platonic solids?

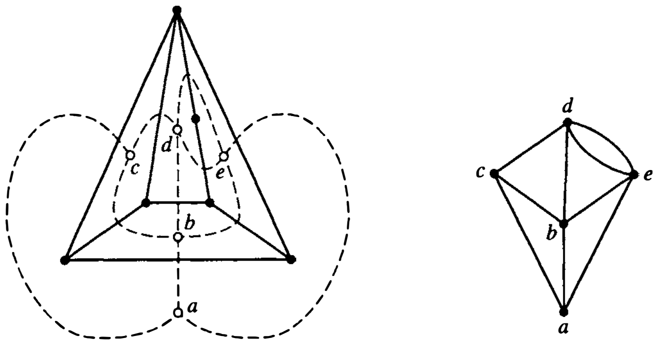


FIGURE X.17. A plane graph and its dual.

15. Let G be a connected plane graph with dual G^*. Prove that

$$T_{G^*}(x, y) = T_G(y, x).$$

16. Show that the chromatic polynomial and the flow polynomial are related by duality:

$$p_G(x) = q_{G^*}(x)x^{k(G)}$$

whenever G is a connected plane graph with dual G^*.

17. Show that the four colour theorem is equivalent to the assertion that every bridgeless planar graph has a nowhere-zero 4-flow.

18. Let T be a spanning tree of a 2-connected loopless graph $G = (V, E)$, and let E_0 be a non-empty subset of E. Show that the closure $\gamma_T(E_0)$ of E_0 defined in the proof of Theorem 14 is the whole of E.

19. Determine the class of graphs G such that $T_{11}(G) > 0$.

20. Let G be a 2-connected loopless graph of order n and girth g. Show that $t_{i1}(G) > 0$ for $0 \le i \le n - g$.

21. Use Reidemeister moves and planar isotopy to show that the knots of the diagrams L_1 and L_2 in Fig. X.4 are ambient isotopic.

22. Let $\tilde{K}$ be the knot given by a continuous function $h : [0, n] \to \mathbb{R}^3$, $h(t) = (x(t), y(t), z(t))$, which is linear on each interval $[k, k+1]$, such that $h(0) = h(n) = (0, 0, 0)$, $h(t) = (0, 0, t)$ for $0 \le t \le 1$ and $z(t) > z(t')$ if $1 \le t < t' \le n$. Show that $\tilde{K}$ is (equivalent to) the trivial knot, that is to the unknotted circle.

23. Let L be a knot diagram obtained in the simplest way we could draw it on a piece of paper: we start at a point P and lift our pencil only when we have to in order to get across a previously drawn line. Thus, starting at P in a certain direction, at every crossing first we go over (so that at a later stage we go under at that crossing). Show that L is equivalent to the trivial (unknotted circle) diagram.

24. A link $\tilde{L}$ is said to be *split* if it has a diagram whose universe is a disconnected graph. The components of a link $\tilde{L}$ with at least two components are said to be *linked* if $\tilde{L}$ is not split.
    Given an oriented link diagram with two sets of components C_1 and C_2, let $C_1 \sqcap C_2$ be the set of crossings of C_1 and C_2. The *linking number* $\mathrm{lk}(C_1, C_2)$ of C_1 and C_2 is
    $$\mathrm{lk}(C_1, C_2) = \frac{1}{2} \sum_{v \in C_1 \sqcap C_2} \varepsilon(v).$$
    Check that the linking number is an ambient isotopy invariant. Deduce that the Hopf link is indeed linked, and it has two orientations that are not ambient isotopic.

25. For a link diagram L, denote by $c_3(L)$ the number of non-trivial proper colourings of L with colours 1, 2 and 3. Thus $c_3(L)$ is the number of ways of colouring the strands with colours 1, 2 and 3 such that (i) at no crossing do we have precisely two colours, and (ii) at least two colours are used. Show that $c_3(L)$ is an ambient isotopy invariant.

26. Check that the diagram of the trefoil knot in Fig. X.5 has a non-trivial proper 3-colouring. Deduce that the trefoil is knotted: it is a non-trivial knot.

27. Give a non-trivial proper 3-colouring of the diagram of the knot 7_4 in Fig. X.15, and deduce that 7_4 is a non-trivial knot.

28. Show that neither the Hopf link in Fig. X.3, nor the Whitehead link in Fig. X.15 has a non-trivial proper 3-colouring, and deduce that the components are indeed linked in each.

29. Show that the diagram of the Borromean rings in Fig. X.3 does not have a non-trivial proper 3-colouring, so the rings are indeed linked, although no two of them are linked.

30. Show that the link of $D^+(C_4)$, the alternating link diagram obtained from the 4-cycle, has two linked components.

31. Use the invariant $c_3(L)$ of Exercise 25 to prove that the connected sum of two trefoil knots in Fig. X.18 is knotted and it is not equivalent to a trefoil knot. Show also that there are infinitely many pairwise inequivalent knots.

32. For a prime $p \geq 3$, a mod p *labelling* of a link diagram L is a labelling of the strands of L by the elements of $\mathbb{Z}_p$ such that (i) if at a crossing the over-pass is labelled x and the other two labels are y and z, then $2x = y + z$, and (ii) at

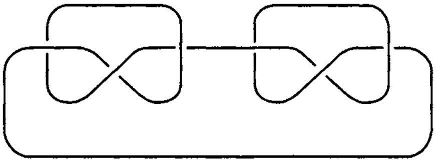


FIGURE X.18. The connected sum of two trefoil knots.

least two labels are used. Denote by $c_p(L)$ the number of mod p labellings of L. Show that for $p = 3$ this definition coincides with the definition of $c_3(L)$ above, and that $c_p(L)$ is an ambient isotopy invariant.

33. Let L be the quinquefoil in Fig. VIII.13. Show that $c_3(L) = 0$, $c_5(L) = 20$ and $c_p(L) = 0$ for every prime $p \geq 7$.

34. Show that the figure of eight knot in Fig. VIII.4 is amphicheiral, that is it is ambient isotopic to its mirror image.

35. Prove that if we redraw a link diagram by turning one of its regions into the outside region then we obtain an ambient isotopic link diagram.

36. Show that we need not have $B = A^{-1}$ and $d = -A^2 - A^{-2}$ to make the Kauffman square bracket $[L]$ a regular isotopy invariant.

37. Calculate the Jones polynomial of the Hopf link and deduce that the two circles are indeed linked.

38. Calculate the Jones polynomial of the Whitehead link and deduce that the two components are indeed linked.

39. Calculate the Jones polynomial of the Borromean rings and deduce that the rings are indeed linked although no two of them are linked.

40. Calculate the Jones polynomials of the right-handed trefoil knot and of the figure of eight knot and check that your answers tally with the ones obtained from the result of Exercise 3.

41. The knots 8_8 and 10_{129} in Fig. X.19 are not equivalent. Calculate their Jones polynomials and note that they are equal.

42. Show that the partition function $R_G(q, p)$ of the random cluster model is
$$R_G(q, p) = q^{k(G)} p^{r(G)} (1 - p)^{n(G)} T_G\left(\frac{p + q - pq}{p}, \frac{1}{1 - p}\right).$$

43. Consider the random cluster model on G, with parameters q and p. Show that for $q > 1$ new edges joining vertices in the same component are more likely than those uniting two old components. To be precise, given $F_0 \subset E(G)$ and $ab = f \in E \backslash F_0$,
$$\mathbb{P}(f \in F | F \backslash \{f\} = F_0) = \begin{cases} p & \text{if } k\langle F \cup f \rangle = k\langle F \rangle, \\ \dfrac{p}{p + q - pq} & \text{otherwise.} \end{cases}$$

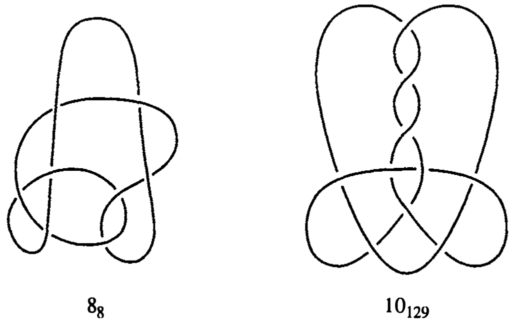


FIGURE X.19. Diagrams of the knots 8_8 and 10_{129}.

44. Note that the proof of Theorem 4 shows that

$$\sum_{F \subset E} v^{|F|} q^{k\langle F \rangle}$$

is precisely the dichromatic polynomial $Z_G(q, v)$, and so give another proof that the Tutte polynomial is well defined.

45. Prove the theorem of Conway and Gordon that every embedding of K_6 into $\mathbb{R}^3$ is intrinsically linked: there are two triangles forming a non-trivial link. [*Hint.* Note first that any embedding of K_6 can be changed to any other embedding by changing some crossings from 'over' to 'under' and vice versa. Let (T_i, T_i'), $i = 1, \ldots, 10$, be the ten pairs of disjoint triangles in K_6, and orient each triangle in an arbitrary way. With a slight abuse of notation, for a given embedding of K_6, set $\mathrm{lk}K_6 = \sum |\mathrm{lk}(T_i, T_i')|$, where $\mathrm{lk}(T_i, T_i')$ is the linking number of T_i and T_i', as in Exercise 24. Prove that the *parity* of $\mathrm{lk}K_6$ is independent of the particular embedding. Deduce from the embedding in Fig. X.20 that $\mathrm{lk}(T_i, T_i') \neq 0$ for some i.]

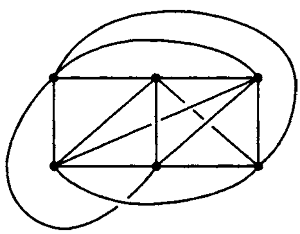


FIGURE X.20. An embedding of K_6 with $\mathrm{lk}K_6 = 3$.

46. Construct an embedding of K_6 into $\mathbb{R}^3$ in which there is only one pair of linked triangles.

47. Prove that every embedding of the Petersen graph into $\mathbb{R}^3$ contains two linked pentagons.

X.8 Notes

W.T. Tutte constructed the dichromate of a graph, the polynomial we know as the Tutte polynomial, in A contribution to the theory of chromatic polynomials, *Canad. J. Math.* **6** (1954) 80–91. This paper contains the spanning tree expansion of the polynomial as well. In fact, Tutte constructed and studied similar polynomials in A ring in graph theory, *Proc. Cambridge Phil. Soc.* **43** (1947), 26–40, building on H. Whitney, The coloring of graphs, *Ann. Math.* **33** (1932) 688–718. Theorem 2 is essentially from J.G. Oxley and D.J.A. Welsh, The Tutte polynomial and percolation, in *Graph Theory and Related Topics* (J.A. Bondy and U.S.R. Murty, eds), Academic Press, London, 1979, pp. 329–339.

The Whitney–Tutte polynomial Q_G should really be called the dichromatic polynomial. However, in many papers on statistical mechanics, the polynomial Z_G goes under that name, and we followed this unfortunate convention. In a way, it does not matter much, as one always has to give defining properties of these polynomials.

For various models in statistical mechanics, especially the Ising model, the Potts model and the random-cluster model, see C.M. Fortuin and P.W. Kasteleyn, On the random cluster model, I, Introduction and relation to other models, *Physica* **57** (1972) 536–564, and B. Bollobás, G. Grimmett and S. Janson, The random-cluster process on the complete graph, *Probab. Theory and Related Fields* **104** (1996) 283–317.

The partial solutions to Tutte's 5-flow Conjecture are in Flows and generalized colouring theorems in graphs, *J. Combinatorial Theory (B)* **26** (1979) 205–216, and P.D. Seymour, Nowhere zero 6-flows, *J. Combinatorial Theory (B)* **30** (1981) 130–135.

For the unimodality conjectures mentioned at the end of Section 5, see R.R. Read, An introduction to chromatic polynomials, *J. Combinatorial Theory* **4** (1968) 52–71, W.T. Tutte, *Graph Theory*, Encyclopaedia of Maths and Its Appl., vol. **21**, Cambridge University Press, 1984, and P.D. Seymour and D.J.A. Welsh, Combinatorial applications of an inequality of statistical mechanics, *Math. Proc. Cambridge Phil. Soc.* **77** (1975) 485–495. For the refutation of the above conjectures, see W. Schwärzler, The coefficients of the Tutte polynomial are not unimodal, *J. Combinatorial Theory (B)* **58** (1993) 240–242.

There are many interesting papers one should consult for the dawn of knot theory, including J.B. Listing, Vorstudien zur Topologie, *Göttingen Studien* **1** (1847) 811–875; Sir William Thomson, On vortex motion, *Trans. Roy. Soc. Edinburgh* **25** (1869), 217–260; P.G. Tait, On knots, *Trans. Roy. Soc. Edinburg* **28** (1879) 145–190, with two additional parts of that paper in the same journal: **32** (1887) 327–339 and 493–506; C.N. Little, On knots, with a census for order ten, *Trans. Connecticut Acad.* **7** (1885), 1–17; and Rev. T.P. Kirkman, The 364 unifilar knots of ten crossings, enumerated and described, *Trans. Roy. Soc. Edinburgh* **32** (1887) 483–491.

A more sophisticated approach to knot theory was taken by M. Dehn, Über die Topologie des dreidimensionalen Raumes, *Math. Ann.* **69** (1910) 137–168, and J.W. Alexander, Topological invariants of knots and links, *Trans. Amer. Math. Soc.* **30** (1928) 275–306. Perhaps the most influential work about the early theory is K. Reidemeister, *Knotentheorie*, Ergebn. Math. Grenzgeb., vol. **1**, Springer-Verlag, Berlin, 1932; for an English translation, see *Knot Theory*, BSC Associates, Moscow, Idaho, 1983. A comprehensive account of knot theory is G. Burde and H. Zieschang, *Knots*, Walter de Gruyter, Berlin, 1985, xi+399 pp; for an elementary introduction, see C. Livingston, *Knot Theory*, Carus Math. Mon., vol. **24**, Math. Assoc. Amer., Washington, 1993, xviii+240 pp.

V.F.R. Jones constructed his powerful new knot polynomial in the summer of 1984, upon a careful examination of a surprising result about von Neumann algebras. The construction was published in two papers: A polynomial invariant for knots via von Neumann algebras, *Bull. Amer. Math. Soc.* **12** (1985) 103–111, and A new knot polynomial and von Neumann algebras, *Notices of AMS* **33** (1986), 219–225. The Kauffman brackets were introduced in L.H. Kauffman, State models and the Jones polynomial, *Topology* **26** (1987) 395–407.

The classical conjectures of Tait mentioned at the end of §6 were proved by K. Murasugi, Jones polynomials and classical conjectures in knot theory, *Topology* **26** (1987) 187–194, and M.B. Thistlethwaite, A spanning tree expansion of the Jones polynomial, *Topology* **26** (1987) 297–309.

For applications of knot invariants to physics, see E. Witten, Quantum field theory and the Jones polynomial, *Commun. Math. Phys.* **121** (1989) 351–399, and the entire volume *The Interface of Knots and Physics* (L.H. Kauffman, ed.) *Proc. Symp. Appl. Math.* **51**, Amer. Math. Soc., Providence, 1993, x+208 pp.

The beautiful theorem of Conway and Gordon in Exercise 45 is from J.H. Conway and C.McA. Gordon, Knots and links in spatial graphs, *J. Graph Theory* **7** (1983) 445–453. The result was greatly extended by N. Robertson, P.D. Seymour and R. Thomas, Linkless embeddings of graphs in 3-space, *Bull. Amer. Math. Soc.* **28** (1993) 84–89.

Finally, for a wealth of information on the material in this chapter, see T.H. Brylawski and J.G. Oxley, The Tutte polynomial and its applications, in *Matroid Applications* (N. White, ed.), Cambridge Univ. Press, 1992, pp.123–225, and D.J.A. Welsh, *Complexity: Knots, Colourings and Counting*, London Math. Soc. Lect. Note Ser., vol. **186**, Cambridge Univ. Press, 1993, viii+163 pp. For generalizations of the Tutte polynomial as far a possible, including conditions under which it gives rise to link polynomials, see B. Bollobás and O. Riordan, A Tutte polynomial for coloured graphs, *Combinatorics, Probability and Computing* **7** (1998).

Symbol Index

Name Index

Subject Index

Graduate Texts in Mathematics

(continued from page ii)

124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings. 2nd ed.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory. 2nd ed.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic *K*-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory. 2nd ed.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.
179 DOUGLAS. Banach Algebra Techniques in Operator Theory. 2nd ed.
180 SRIVASTAVA. A Course on Borel Sets.
181 KRESS. Numerical Analysis.
182 WALTER. Ordinary Differential Equations.

183 MEGGINSON. An Introduction to Banach Space Theory.
184 BOLLOBAS. Modern Graph Theory.
185 COX/LITTLE/O'SHEA. Using Algebraic Geometry. 2nd ed.
186 RAMAKRISHNAN/VALENZA. Fourier Analysis on Number Fields.
187 HARRIS/MORRISON. Moduli of Curves.
188 GOLDBLATT. Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.
189 LAM. Lectures on Modules and Rings.
190 ESMONDE/MURTY. Problems in Algebraic Number Theory. 2nd ed.
191 LANG. Fundamentals of Differential Geometry.
192 HIRSCH/LACOMBE. Elements of Functional Analysis.
193 COHEN. Advanced Topics in Computational Number Theory.
194 ENGEL/NAGEL. One-Parameter Semigroups for Linear Evolution Equations.
195 NATHANSON. Elementary Methods in Number Theory.
196 OSBORNE. Basic Homo!ogical Algebra.
197 EISENBUD/HARRIS. The Geometry of Schemes.
198 ROBERT. A Course in p-adic Analysis.
199 HEDENMALM/KORENBLUM/ZHU. Theory of Bergman Spaces.
200 BAO/CHERN/SHEN. An Introduction to Riemann–Finsler Geometry.
201 HINDRY/SILVERMAN. Diophantine Geometry: An Introduction.
202 LEE. Introduction to Topological Manifolds.
203 SAGAN. The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions.
204 ESCOFIER. Galois Theory.
205 FÉLIX/HALPERIN/THOMAS. Rational Homotopy Theory. 2nd ed.
206 MURTY. Problems in Analytic Number Theory.
Readings in Mathematics
207 GODSIL/ROYLE. Algebraic Graph Theory.
208 CHENEY. Analysis for Applied Mathematics.
209 ARVESON. A Short Course on Spectral Theory.
210 ROSEN. Number Theory in Function Fields.
211 LANG. Algebra. Revised 3rd ed.
212 MATOUŠEK. Lectures on Discrete Geometry.
213 FRITZSCHE/GRAUERT. From Holomorphic Functions to Complex Manifolds.
214 JOST. Partial Differential Equations.
215 GOLDSCHMIDT. Algebraic Functions and Projective Curves.
216 D. SERRE. Matrices: Theory and Applications.
217 MARKER. Model Theory: An Introduction.
218 LEE. Introduction to Smooth Manifolds.
219 MACLACHLAN/REID. The Arithmetic of Hyperbolic 3-Manifolds.
220 NESTRUEV. Smooth Manifolds and Observables.
221 GRÜNBAUM. Convex Polytopes. 2nd ed.
222 HALL. Lie Groups, Lie Algebras, and Representations: An Elementary Introduction.
223 VRETBLAD. Fourier Analysis and Its Applications.
224 WALSCHAP. Metric Structures in Differential Geometry.
225 BUMP: Lie Groups
226 ZHU. Spaces of Holomorphic Functions in the Unit Ball.
227 MILLER/STURMFELS. Combinatorial Commutative Algebra.
228 DIAMOND/SHURMAN. A First Course in Modular Forms.
229 EISENBUD. The Geometry of Syzygies.
230 STROOCK. An Introduction to Markov Processes.