

Style analysis by machine learning reveals that two authors likely shared the writing of *QAnon*'s messages at two different periods in time.

Group of Stylometric analysis,
for the experts Claude-Alain Roten
OrphanAnalytics
www.orphananalytics.com
Case postale 805
1800 Vevey, Switzerland

+41 79 785 25 33
contact@orphananalytics.com

Summary

Between October 28, 2017 and November 13, 2020, 4952 anonymous messages signed Q were published first on the *4chan* site and then on the *8chan* site, which recently became *8kun*. This corpus of messages is designated *QAnon*, for their Q-signature letter and their anonymous character. For the *QAnon* movement surrounding this corpus, the signatory is a U.S. official who would have a Q clearance giving him access to confidential information and this official would anonymously publish information intended to prepare the population for important political changes. To our knowledge, the *QAnon* corpus has not yet been subjected to a stylometric analysis capable of characterizing the style of a writer. The unsupervised machine learning approach using multivariate statistical analysis allows the appearance of text sequences clustered by the author's own style.

The whole corpus is collected in order to challenge the proposal that a single writer be the sole author of Q-drops specific to *QAnon*. To this goal, texts that are too short, quotations and speeches of historical figures for example, that do not have a personal syntax to the anonymous writer have been discarded for analysis in chronological order. Multivariate statistical analyses focused on comparing three-letter pattern profiles to define the styles used in *QAnon*.

The stylometrics of the 7.5k chr **de** Q-drops concatenates classified chronologically reveals two clusters, characteristic of two different styles corresponding to the two periods of publication of the Q-drops on the *4chan* and *8chan* forum. This observation sheds light on the background information of the media investigations. The signal is mostly carried by Q-drops of less than 1k chr and clustering does not seem to interfere with the analyses. The other type of concatenation tested, concatenation by size, was unable to cluster reasonably. A success rate of more than 90% was calculated for the stylometrics of Q-drops concatenates classified chronologically by non-hierarchical clustering analysis. This rate is comparable to that measured in a criminal case under investigation and to that obtained on texts from a solved case.

Other algorithms are available that would allow to specify the results already obtained. If texts from candidate writers are provided, this knowledge of the *QAnon* corpus would make it easier to recognize the styles of people likely to have written Q-drops.

Introduction

The *QAnon* corpus and its signatory Q

As of November 13, 2020, the *QAnon* corpus contained 4952 anonymous messages (qresearch.ch) which appeared first on the *4chan* site, then on *8chan* which became *8kun* three months after its final stop. The name *QAnon* refers to Q - the 17th letter of the alphabet - signature of the messages called Q-drops, and to Anon, for the anonymity surrounding the writing and publication of these messages.

According to the eponymous movement surrounding *QAnon*, the Q signature designates a senior official with clearance level Q from the U.S. Department of Energy in charge of nuclear policy who would give him or her access to information about the U.S. nuclear arsenal. In order to prepare the public for significant policy changes, Q would anonymously publish confidential government information.

At present, the identity of Q and his activity as a senior government official has not been established, nor has the number of editors actually involved in the production of this corpus been determined: one or more individuals?

Results of media investigations

Wang & Click (2020) analyzed the style and content of just under 5000 *QAnon* messages written over the past three years. These messages, called *Q-drops*, claim that a secret war is raging between President Donald Trump and members of the so-called deep-state American elite who are allegedly satanic pedophiles trafficking in human beings. The criminal acts of some members of the *QAnon* movement have led the FBI to consider this group as a domestic terrorist threat. Most of these Q-drops are written in a school style (3rd-4th grade) at the level of an 8 to 10 year old student. The personalities most mentioned in the Q-drops are first Hillary Clinton and then Barack Obama. President Obama is mentioned in the Q-drops by his middle name Hussein as well as by the name "Renegade", allegedly chosen by his secret service. xxxxxx

According to Zadrozny & Collins (2018), Q-drops are first published on October 28, 2017 by the anonymous American forum *4chan*. A trio manages the start of the publication of Q-drops on *4chan*, including one of the members who is moderator on *4chan*, the South African Paul Furber, and he is the first to spot messages of interest signed by the letter Q that seem to come from a high-ranking official. For the creation of *QAnon* on *4chan*, Furber, whose alias on is *Baruch the Scribe*, is associated with Americans Coleman Rogers (aka *Pamphlet Anon*) and, for communication, Tracy Diaz (aka *Tracybeanz*). For some, Furber could be Q (Vogt, 2020).

According to Zadrozny & Collins (2018), the confidential audience for the first Q-drops grew rapidly thanks to Diaz's effective communication and the creation of the Reddit CBTS discussion group (for Calm Before The Storm), which was easier to access for an older audience (boomers) made up of normies not used to the practices of geeks/gamers.

The application of this strategy quickly allows *QAnon* messages to distinguish themselves from the anonymous feeds that were broadcasting on *4chan* information allegedly coming from government sources (*FBIAnon*, *HLIAnon* for High Level Insider, *CIAAnon* and *CIA intern*).

The existence of the *Reddit CBTS_Stream* group facilitates the discussion and consultation of Q-drops. Upon the announcement of an infiltration on the *QAnon* page, the publication of Q-drops is transferred on December 1, 2017 from the *4chan* forum to the controversial *8chan*, owned by the American Jim Watkins. The interest around *QAnon* increases with the interview of Furber and Coleman on *InfoWars*. Furber then manages the publication of Q-drops on *8chan cbts* board, but he is quickly excluded from the private discussions of the *Reddit CBTS* group for threatening to reveal a user's data.

Furber announced that imposters have now taken control of the *Reddit* newsgroup. As of January 6, 2018, Q-drops are published on the *8chan qresearch* feed managed by Jim Watkins' son Ron Watkins. In March 2018, *Reddit* closed the Q-drops, which then had 20,000 subscribers, for advocating violence and publishing personal and confidential information. In order to get Q in the media, Rogers created a youtube channel with his wife Christina Urso within a month. The couple is suspected by some activists of benefiting financially from the *QAnon* movement. Soon after, two public incidents revealed that Rogers had access to the Q account.

According to Wang & Click (2020) Ron Watkins secures the publication of Q-drops on *8chan* by *tripcode* as of August 10, 2018. The *8chan* forum which then publishes Q-drops is closed on August 1, 2019 following publications related to mass massacres, to reappear three months later, on November 2, 2019, under the name *8kun*. When Q does not publish anything in 2019 during these three months of the inactivity of *8chan*, some observers have hypothesized that since Jim and Ron Watkins are respectively the owner of *8chan* and manager of the *QAnon* publication page, they know Q or write Q-drops for Q. Their web activity analyzed by Wang & Click (2020) supports this hypothesis. This takeover of the account by the Watkins would explain the change in language observed by Vogt (2020) in a *Reply All* podcast: compared to the first Q-drops, the others abuse aggressive writing in capital letters and are made up of less coherent sentences. Furber confirms this hypothesis with a tweet announcing that the Watkinses control the Q-count (Wang & Click, 2020).

Ron Watkins resigned on November 13, 2020 from his position at *8chan*. A message probably from his hand denounces his father as the editor of Q-drops (Stanley, 2020). Since then Ron Watkins has been working as an election fraud expert for Donald Trump ().

None of the five individuals mentioned above - i.e. Furber, Rogers, Diaz, Jim and Ron Watkins - has publicly acknowledged being or knowing Q.

OrphAnalytics' know-how

The present work analyzes by stylometry the writing history of the *QAnon* corpus. Knowing, for example, that the *4chan* forum did not guarantee that only one author could use the Q signature, the question of the number of possible writers in the *QAnon* corpus is as relevant as the one concerning the authentication of these messages. These issues of authorship are topical, as new Q-drops are being published and their content is intended to interfere with the transition from the administration of President Donald Trump to the next. OrphAnalytics' algorithmic stylometric tools (www.orphanalytics.com) allow to measure the number of authors who have written a corpus and to authenticate them if reference texts of candidate authors are available by comparing the character pattern profiles of each text fragment. The stylometric tools used have been developed from statistical approaches used and validated by comparative genomics.

OrphAnalytics' servers and stylometric expertise approach are designed for various author attribution needs: establishing academic fraud by plagiarism or ghostwriting, determining contract amendments for arbitration tribunals, even identifying the authors of anonymous letters in unsolved criminal cases for investigative services. The results of these investigations are and remain confidential. Made public, the following two cases from the company's practice show this.

Stylometric analysis of Elena Ferrante's novels

These tools have been used for demonstration purposes on various literary works. In response to rumors that the 2016 Nobel Prize in Literature could be awarded to the author of the novels signed with the pseudonym Elena Ferrante, OrphAnalytics experts compared the style of these works with those of authors suspected of publishing under the name Ferrante. The results obtained reveal that the author Domenico Starnone writes in a style strongly similar to that of Elena Ferrante.

The OrphAnalytics expert team is the third statistical analysis group to announce in the media that Starnone could be the author of Ferrante's novels, but it is the first to publish the statistical results. To respond to the media urgency of the moment, the results are published by press release¹.

This attribution of authorship is clearly confirmed in September 2017 by the similar results obtained by eight internationally renowned experts at the workshop organized by the University of Padua on the statistical analysis of Ferrante's work.

Analyses of the Hearings on the Appointment of Justice Kavanaugh to the Supreme Court in the U.S. Senate

In order to publicly confirm the effectiveness of the stylometric tools, an analysis is being conducted at the time of the appointment of Judge Brett M. Kavanaugh in 2019. Christine Bradley Ford complained in a letter sent to California Senator Dianne Feinstein that Ford had been sexually assaulted by Kavanaugh when they were both high school students. During her public hearing in the U.S. Senate, Republican senators challenged the authenticity of this letter, asking her several times if she was the writer of the letter. Ford responded in the affirmative.

OrphAnalytics' stylometric results, obtained within a few days from the transcripts of the available hearings, reveal that the style of this letter is similar to the text of her introductory message during her Senate hearing, which most certainly appears in her hand. The style of Ford's texts differs from that of Kavanaugh's texts used for his Senate hearings and their point cloud is similar in size. Finally, a comparison of the clusters of the transcripts of the Senate hearing responses of Ford and Kavanaugh also allows us to differentiate the oral styles of the two individuals and to measure that the Ford and Kavanaugh writing and oral transcription clouds behave in the same way.

The analysis conducted in a few days answers the question about the authenticity of Ford's letter to Senator Feinstein. The stylometric comparison thus confirms that Ford is most certainly the writer of this letter: if Ford had written this letter, the stylometric analyses would have measured a very strong similarity between the style of the letter sent to Senator Feinstein and that of the text of her introduction to the Senate hearing. Simply put, Ford is most certainly the writer of this letter. On the

¹ <https://www.orphanalytics.com/en/news/ferrante-starnone>

eve of Judge Kavanaugh's appointment by the Senate, a press release presenting these stylometric results was issued by OrphAnalytics².

Brett M. Kavanaugh, who has since been elected to the Supreme Court, denied at the time of his nomination that he attempted to sexually assault Christine Bradley Ford.

Material and Methods

Questioned texts

The Q-drops used for the analyses have been downloaded on the sites hosting them³. They correspond to the 4952 Q-drops available on November 13, 2020. The corpus *QAnon* of 4952 Q-drops is purified to keep only the messages with sufficient syntax specific to an author in this context. Messages of less than 50 chr (with the exception of the 1308 Q-drops of less than 50 chr included in the size analysis: Fig. 1), quotations and speeches by politicians, open letters and lists of banking institutions, for example, are therefore discarded. In addition, web links are removed from the messages. At the end of this process, the analyzed corpus consists of 2457 messages of more than 50 chr.

Reference texts

In order to measure an author's stylometric signal, 51 texts by Alexander Hamilton were compared to those of the *QAnon*. These texts were urgently written as articles for New York State newspapers to convince the people of New York State to accept the U.S. Federal Constitution put to a vote in 1789. They were published as 51 chapters in the *Federalist papers* (1788-1789) under the pseudonym *Publius*, containing 85 chapters in all. The other 34 chapters of the *Federalist papers* were written by James Madison and John Jay together and under the same conditions.

Written by three founding fathers of American History, the *Federalist papers* have been considered since their publication as the cardinal work for understanding the American Constitution. The pioneering work in stylometry (i.e. statistical analysis of style) by the statisticians Mosteller and Wallace (1964) made it possible to confirm the author of each chapter of the *Federalist papers* by using the maximum computing capacity of the computers of the time⁴. The 51 chapters of the *Federalist papers* attributed to Hamilton used in our analyses are as follows: 1, 6-9, 11-13, 15-17, 21-36, 59-61, 65-85.

Preparation of the texts

The text sequences are then processed to keep only one space character between the words and the 26 letters of the message alphabet, ignoring the initial case. Depending on the choice of analysis, the texts can be analyzed either separately if their sequence exceeds 1k chr, or concatenated (assembled end to end) by size and/or chronological order to be cut into 7.5k chr pieces.

Profile of text sequences.

² <https://www.orphanalytics.com/en/news/press-release-october-4-2018>

³ <https://qreseat.ch/>

⁴ <https://priceconomics.com/how-statistics-solved-a-175-year-old-mystery-about/>

A three-character pattern usage profile is established for each message or concatenate used in the following statistical comparisons. Each text sequence can be represented by a dot in a space with a number of dimensions equal to the number of three-character pattern types used in the analyses.

Statistical analyses

The profiles of each text sequence are compared two by two by establishing the distance between the two text fragments. This distance between two text sequences is established by summing the block distances separating the coordinates of the points representing the pair of texts being compared, i.e. frequencies of three-character structures.

Unsupervised machine learning

A multivariate analysis by principal coordinates on the block distances separating each text pair projects each point representing a text sequence into a new space with a reduced number of dimensions. This statistical analysis retains the relative variance between dots representing a text sequence and prioritizes the resulting dimensions according to the associated relative variance.

Each text sequence is represented by a dot in the two-dimensional graph constructed with the two new dimensions of maximum variance obtained after the reduction of dimensionality. The two axes of the graph are calibrated in the relative units defined by the multivariate analysis.

Results

Comparison of two types of Q-drops clustering: by size or by chronological order

The first two analyses compared Q-drops of more than 50 concatenated characters per size group (Fig. 1) or in chronological order (Fig. 2). Groups and subgroups were created according to the periods defined by the article by Wang & Click (2020). These chronological periods are marked by the publication of Q-drops on the *4chan* or *8chan* sites, which became *8kun*, on the dedicated *8chan* pages (*cbts* vs *qresearch*) and by the definitive cessation of *8chan* for three months, which sees *8chan* become *8kun* after 3 months.

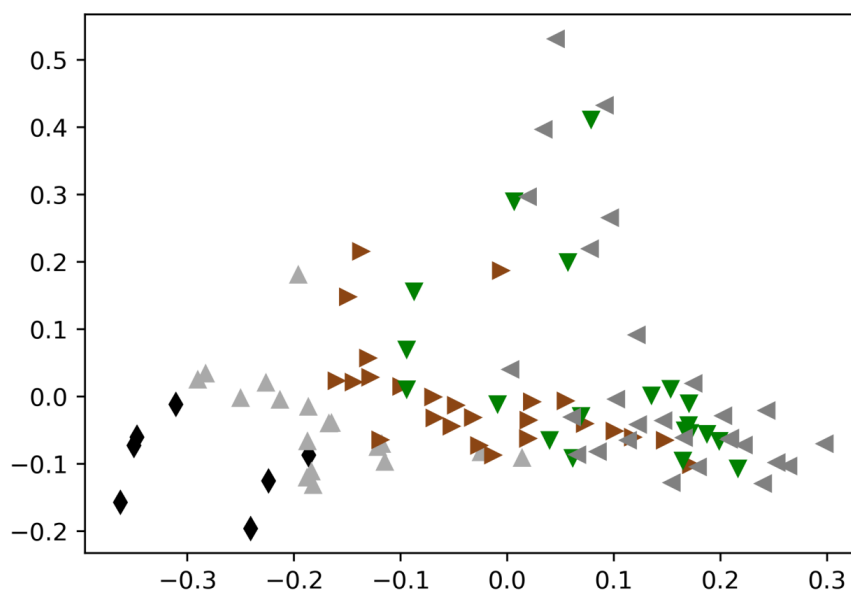


Figure 1: Comparison of Q-drops concatenates classified by size. Multivariate statistical analysis comparing the use of three-character patterns. The comparison is made on the block distances between 7.5k Q-drops concatenates packed in five different length classes. To create the 7.5k chr concatenates, the Q-drops are first sorted by size to form 5 size classes and then, within each class, sorted in chronological order. The Q-drops of each class are then concatenated into a single sequence which is then cut into fragments of 7.5k chr. The precise length of the 7.5k chr concatenates is algorithmically determined to split the sequence of all Q-drops of each class into an integer number of segments of about 7.5k chr.

According to their length class, the concatenates are illustrated by the symbols (diamonds then four triangles rotating clockwise) associated with the following colors: class from 1 to 50 chr in black diamond (average size 27 chr), class from 51 to 150 chr in light grey triangles pointing upwards (average size 92 chr), class from 151 to 400 chr in brown triangles pointing to the right (average size 241 chr); class from 401 to 900 chr in green triangles pointing downwards (average size 585 chr); class from 901 to 4100 chr in light grey triangles pointing to the left (average size 1532 chr). The graph shows the results in the two dimensions of maximum variance.

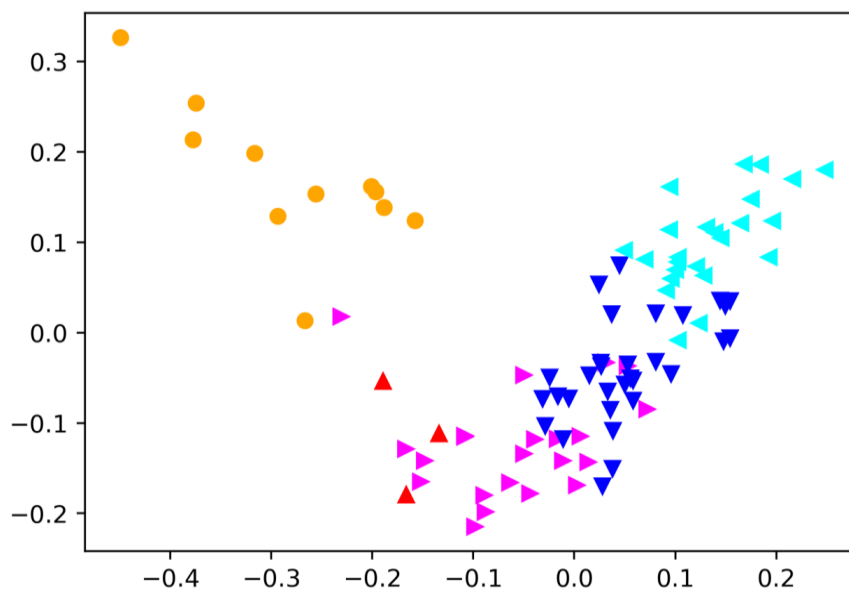


Figure 2: Comparison of Q-drops concatenates in chronological order. Multivariate statistical analysis of block distances between 7.5k chr Q-drops concatenates of five different chronological classes (Wang & Click, 2020). The Q-drops are first sorted in chronological order to be sorted according to the five determined periods, then concatenated into a sequence per class which is then cut into 7.5k chr fragments as in Fig. 1: Symbols (circles and triangles rotating clockwise) and colors arranged according to the chromatic circle (orange, red, magenta, blue, cyan):

- orange circles, from 28-10 to 1-12-2017, 4chan, Q-drop #1-240 (average size 502 chr);
 - red triangles pointing upwards, from 1-12-2017 to 6-1-2018, 8chan, cbts, #241-466 (average size 157 chr);
 - magenta triangles pointing right, from 6-1 to 10-8-2018, 8chan, qresearch, #468-1824 (average size 220 chr);
 - blue triangles pointing down, from 10-8-2018 to 1-8-2019, 8chan, qresearch, Q-drops secured by tripcode, #1829-3570 (average size 253 chr);
 - cyan triangles oriented to the left, from 2-11 to 13-11-2019, 8kun, qresearch, #3572-4951 (average size 307 chr).
- The graph presents the results in the two dimensions of maximum variance.

The analysis of Fig. 1 reveals that multivariate statistical analysis does not cluster the 7.5k chr concatenates because Q-drops classified by size. However, the concatenates gradually spread out as a function of length.

A difference in style is measurable according to the size of the Q-drops, mainly along the horizontal axis: short Q-drops show a syntax that gradually differs from long Q-drops. In the vertical axis, the

points of the class consisting of short Q-drops are more grouped than the points of the classes of longer Q-drops. This difference in behavior reveals that the stylometric signal is more stable in short Q-drops than in long Q-drops. The short Q-drops therefore appear more constrained in style than the long Q-drops.

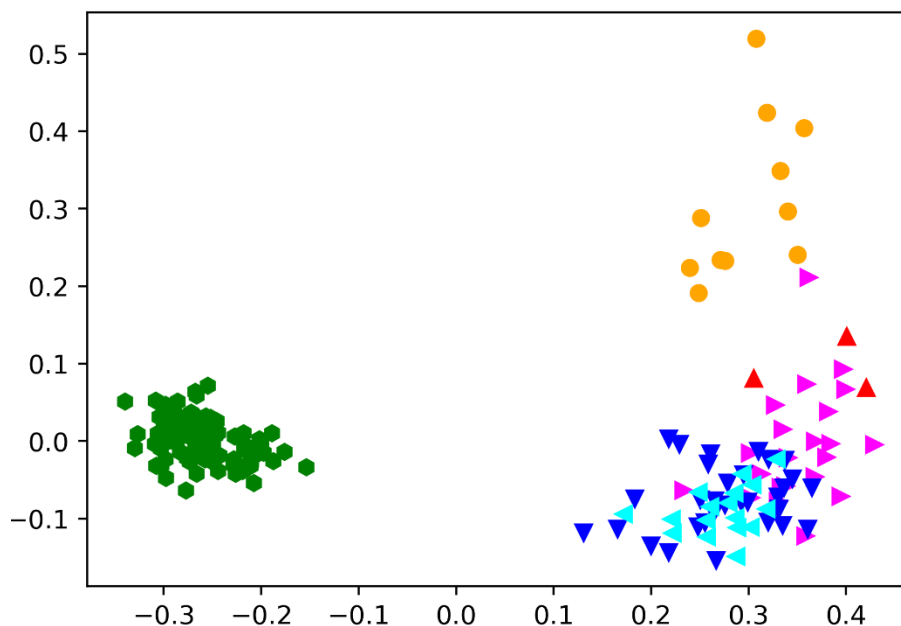
Contrary to Fig. 1, Fig. 2 shows the similar analysis on concatenates constructed in chronological order. Two clusters appear: the first one, orange, corresponds to the period when Q-drops were published on *4chan*, the other one, variegated (red, magenta, blue, cyan), being made of Q-drops published on *8chan* which became *8kun*. These two orange and colorful clusters will be designated by the publication forum of the Q-drops that published them: *4chan* and *8chan* respectively.

This grouping in two clusters reveals that two styles measurable by multivariate statistical analysis are observable in the Q-drops arranged in chronological order. In addition, a fine analysis of the *8chan* cluster reveals a drift over time (from 1-12-2017 to 13-11-2020). According to our experience, this slight drift is similar to that observed for an author who regularly follows a format that he imposes himself. The different orientation of two ellipses that can be superimposed on the *4chan* and *8chan* clusters is another indication of two different styles. With the exception of the choice of symbols and colors, the results in Fig. 2 are the same as those shown in the figure of the press release that accompanies this publication of results⁵.

The comparison of Figs. 1 and 2 reveals that only the concatenation of the Q-drops in chronological order groups the Q-drops into two styles.

Text clustering by author

In order to measure how multivariate statistical analyses group texts according to their authors, a similar experiment is organized in which 7.5k concatenates of Q-drops arranged in chronological order are associated with the 7.5k concatenates of the 51 chapters of Alexander Hamilton's Federalist Papers arranged in chronological order.



⁵ <https://www.orphanalytics.com/en/news/proa20201215>

Figure 3: Comparison of Q-drops concatenates and Alexander Hamilton's papers in chronological order. Same conditions as those used for the results in Fig. 2, the only difference being the addition of the 7.5k concatenates obtained from the chronological concatenation of the 51 chapters of Alexander Hamilton: green hexagons.

Multivariate statistical analyses of the 7.5k chr concatenates of the Q-drops and Hamilton chapters reveal that the sequences are grouped into three clusters, two of which are of the same composition as the two shown in Fig. 2.

A non-hierarchical UPGMA clustering of the results of this multivariate statistical analysis by cosine theta similarity measure unambiguously distinguishes (co-phenetic coefficient 0.995%) the three clusters of 7.5k chr concatenates of the Hamilton, *4chan* and *8chan* texts, consisting respectively of 97, 11 and 73 concatenates. Only one point of the *8chan* cluster is associated with the *4chan* cluster. The error rates computable by this clustering are 0, 1/12 and 1/74 for the Hamilton, *4chan* and *8chan* clusters respectively. The clustering success rate is over 90%. This non-hierarchical clustering analysis successfully confirms that the three styles are distinct.

The multivariate statistical analysis associated with clustering reasonably suggests that the two clearly different styles observed in the Q-drops sequences in Fig. 2 are the product of two different authors. Therefore, according to the results of this statistical analysis, an author writing with one of the two *QAnon* styles, with the *4chan* style for example, would most likely (more than 90%) produce a text with the *4chan* style.

Effect of concatenation on concatenates of more than 1k because they are chronologically classified

In order to produce results quickly, in accordance with the needs of the news (i.e. the confirmation of December 14, 2020 by the Electoral College of the election to the presidency of Joe Biden contested by Donald Trump), the strategy of concatenate analysis has been selected. The concatenates arranged in chronological order of Q-drops form the *4chan* and *8chan* clusters. A stylometric measurement is organized to measure the impact of concatenation on the statistics of the three Figures 1 to 3. In this analysis, only sequences of more than 1k chr of the *QAnon* corpus in chronological order are compared to each other, without concatenation.

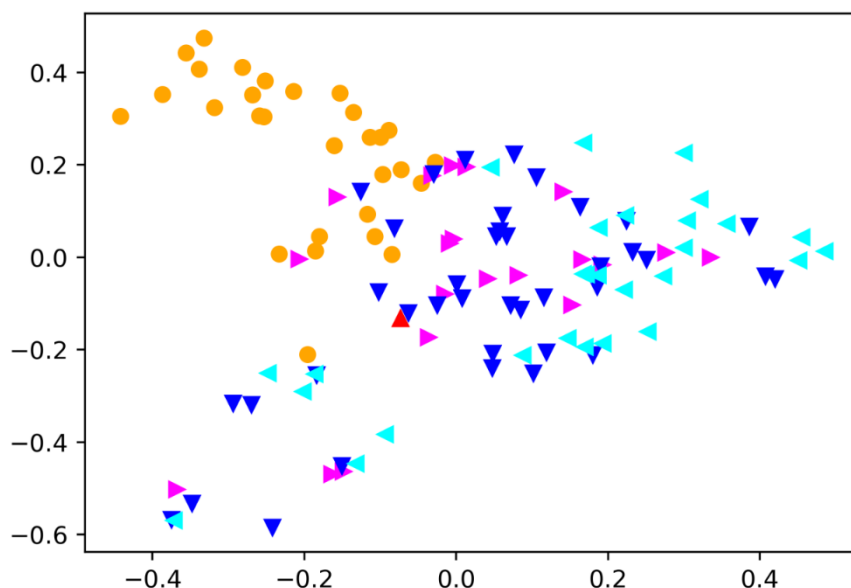


Figure 4: Comparison of individual Q-drops of more than 1k as ranked in chronological order. As soon as the Q-drops of more than 1k chr are selected, these Q-drops are sorted in chronological order and grouped according to the five chronology classes of Wang & Click (2020). Colors and symbols according to Fig. 2. The statistical analysis parameterized according to Fig. 2 compares the sequences directly with each other, without concatenation. The graph shows the results in the two dimensions of maximum variance.

The results in Fig. 4 are similar to those presented in Fig. 2, even if the cluster resolution is lower. This loss of resolution is due to the smaller size of the compared sequences: between 1 and 5 k car. The two clusters *4chan* and *8chan* are recognizable in Fig. 4. The difference in style between the two clusters is therefore due to the Q-drops of more than 1k car compared in isolation.

Effect of size on the comparison of concatenates of Q-drops of different sizes

In order to measure the effect of the size of chronologically concatenated Q-drops, two statistical comparisons are organized similar to the one producing the results shown in Fig. 2. A comparison of the statistical results is organized to measure the difference in result between 7.5 k car concatenates of more than 1k car (Fig. 5) and 51 to 1000 car concatenates (Fig. 6).

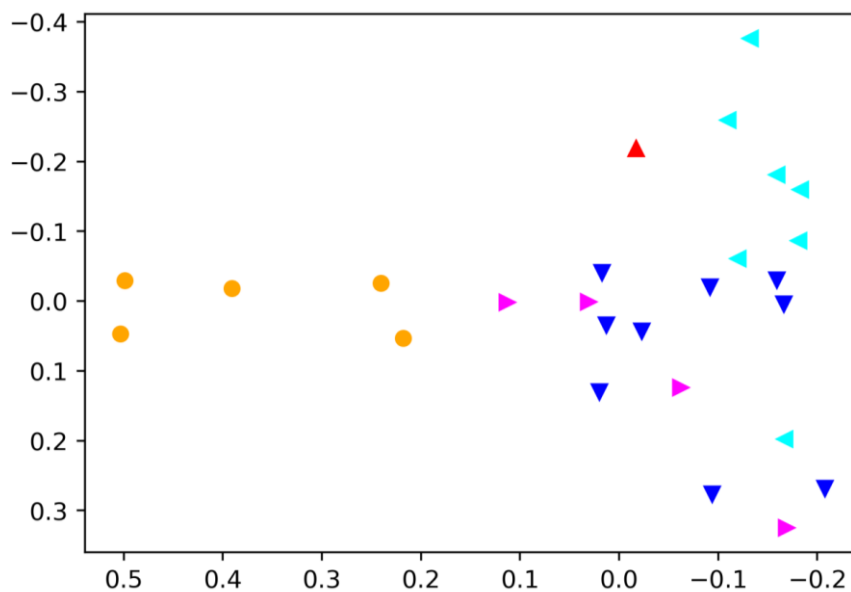


Figure 5: Comparison of Q-drops concatenates of more than 1k as classified in chronological order. Multivariate statistical analysis according to Fig.4. The sequences compared are those in Fig. 4, but concatenated as in Fig. 2. In order to easily compare the results in Fig. 5 with similar analyses (Figs. 2, 4 and 6), the results presented are those of the first and third dimension of large variance and their axes are inverted (positive units before negative ones) the results. These graphical choices do not affect the statistical results obtained.

Five classes of 118 Q-drops >1 of more than 1000 chr. #1-239, n = 29, mean 1357 chr \pm 256; #240-466, n = 1, mean 1486 chr; #467-1827, n = 20, mean 1502 chr \pm 1377; #1828-3570, n = 39, mean 1628 chr \pm 564; #3571-4952: n = 29, mean 1982 chr \pm 942.

The analysis of the 7.5 k chr concatenates of Q-drops of more than 1k car chronologically clustered in Fig. 5 reveals that the two orange *4chan* and variegated *8chan* clusters observed in Figs. 2 to 4 are perfectly separated in Fig. 5. The comparison of Fig. 4 and 5 illustrating the statistical analyses on a set of identical texts, i.e. Q-drops of more than 1k car, confirms that the lower clustering resolution observed in Fig. 4 is due to the smaller size of the compared text sequences: the sequences being individual and not concatenated.

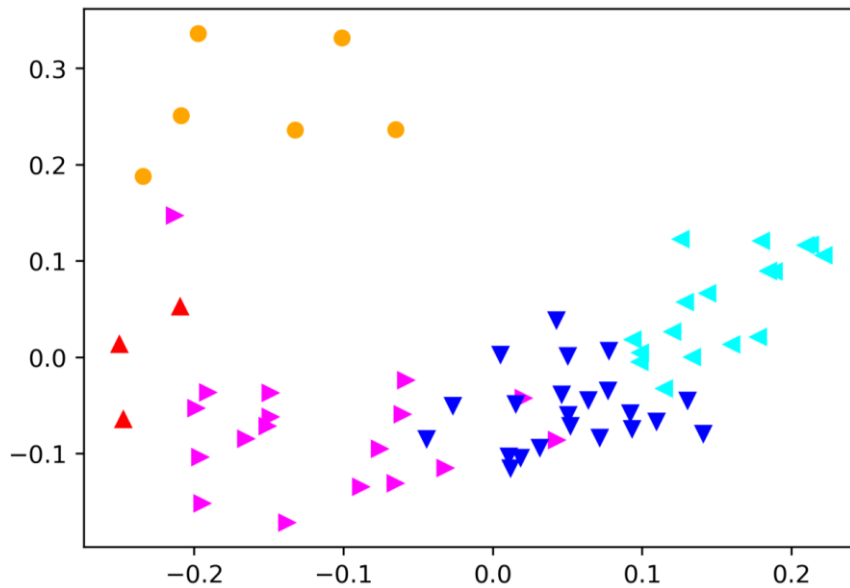


Figure 6: Comparison of Q-drops concatenates of less than 1k as classified in chronological order. Multivariate statistical analysis according to Fig. 5. Only the compared sequences differ: the 7.5k car concatenates of Q-drops from 51 to 1000 car are interlocked and concatenated as in Fig. 5. The graph shows the results in the two dimensions of maximum variance.

Five classes of 2325 Q-drops 50-1000 chr. #1-239, $n = 134$, mean 320 chr ± 244 ; #240-466, $n = 124$, mean 146 chr ± 130 ; #467-1827, $n = 690$, mean 183 chr ± 157 ; #1828-3570, $n = 839$, mean 189 chr ± 178 ; #3571-4952: $n = 538$, mean 224 chr ± 186 .

The analysis of the 7.5 k car concatenates of Q-drops of 51 to 1000 car chronologically arranged in Fig. 5 reveals that the two orange *4chan* and variegated *8chan* clusters observed in Figs. 2 to 4 are perfectly separated in Fig. 6. This resolution between the two clusters *4chan* and *8chan* is certainly the best obtained in the framework of this research if one compares the dispersion of the points in Fig. 6 to that of Figs. 1 to 5.

The comparison of the statistical results shown in Figs. 5 and 6 reveals unambiguously that the part specific to each of the two styles observed in *QAnon* is mostly carried by the Q-Drops from 51 to 1000 chr. Furthermore, a junction zone appears between the two clusters which reveals either a third person or a relatively short collaboration in time between the two people who would have written each with one of the two styles of *QAnon*. Finer algorithms capable of providing a more precise answer are available and will be used in the near future.

Discussion

Analysis of the *QAnon* corpus by unsupervised machine learning shows that a stylometric analysis of the 7.5k concatenates of chronologically classified Q-drops is able to highlight two clusters, characteristic of two different styles (Fig. 2). These two styles coincide with two periods of Q-drops publication, each style corresponding to a single forum: *4chan* or *8chan*. This observation sheds light on the background information available in the media surveys, which showed that more than one writer probably wrote all of the Q-drops: Vogt (2020), Wang & Click (2020). Another type of concatenation tested, size concatenation, proved unable to provide a usable clusterization.

The analysis of Q-drops larger than 1k as compared individually (Fig. 4) reveals that the clusterization observed in the concatenate comparison analysis shown in Fig. 2 is independent of concatenation. Comparison of the signals obtained from the concatenate analyses of Q-drops smaller than 1k chr

(Fig. 6) and those larger than 1k chr (Fig. 5) reveals that the signal is predominantly carried by Q-drops of size is less than 1k chr.

Stylometric analysis of the Q-drops of *QAnon* associated with the texts of Alexandre Hamilton (Fig. 3) reveals that the size of the clusters of points obtained, as well as the resolution of the analysis, allows the separation of styles specific to different authors. The success rate of more than 90% measured by non-hierarchical UPGMA clustering for this multivariate statistical analysis corresponds to those obtained by OrphAnalytics experts in criminal investigation settings.

In the context of an unsolved crime that requires the comparison of criminal messages with reference texts, the success rate obtained is 95%. This rate is calculated by comparing the results of stylometric analyses and their validation by the context. A similar efficiency rate is found by a probability analysis of the clustering results. The complexity of the corpus questioned, the size and the type of messages analyzed correspond to those of *QAnon*.

Provided by an international survey organization, a solved case was used as a test to measure the effectiveness of OrphAnalytics' stylometry. Messages similar to *QAnon*'s were used to measure a 95% success rate in an experiment without context knowledge.

Conclusion

In view of the results presented in this contribution, as well as the insights provided by the practice of the OrphAnalytics expert group, the presence of two styles in the *QAnon* corpus is clearly established. These two styles most likely correspond to two distinct authors. Since a certain number of authors have been proposed as possible authors of this corpus, an analysis of reference texts would allow to challenge the results obtained in this contribution. The continuation of our work would correspond to that done on Elena Ferrante's corpus. This additional analysis would benefit from algorithms capable of providing results with a higher resolution, as well as from a quantification of the probability of the results obtained.

Bibliography

Breland A. (2020) *Trump's Newest Expert on Vote Fraud Spent Years Boosting QAnon. The president has seized on "unsubstantiated... technically incoherent" claims.* November 19, 2020, www.motherjones.com
<https://www.motherjones.com/2020-elections/2020/11/ron-watkins-trump-dominion-fraud/>

Mosteller F. & Wallace D.L. (1964). *Inference and disputed authorship* The Federalist, Addison-Wesley, Reading, Mass.

Christopher B. (2016). *How Statistics Solved a 175-Year-Old Mystery About Alexander Hamilton.* October 31, *Priceonomics*.
<https://priceonomics.com/how-statistics-solved-a-175-year-old-mystery-about/>

Stanley A. (2020) *Wait... Did Ron Watkins Just Rat His Dad Out as Q ?* Nov 15, *gizmodo.com*
<https://gizmodo.com/wait-did-ron-watkins-just-rat-his-dad-out-as-q-1845683225>

Vogt P.J. (2020) *Country of Liars.* September 18, *Gimlet*, Reply All, Transcript #166.
<https://gimletmedia.com/shows/reply-all/llhe5nm>

Wang A. & Click S. (2020). *We analyzed every message ever posted by 'Q,' the enigmatic persona that started the QAnon conspiracy theory.* November 3, 2020, *Business Insider*, <https://www.businessinsider.com/every-QAnon-message-q-drop-analyzed-2020-10>

Zadrozny B. & Collins B. (2018). *How three conspiracy theorists took 'Q' and sparked QAnon.* August 14, 2018, *NBC report*.
<https://www.nbcnews.com/tech/tech-news/how-three-conspiracy-theorists-took-q-sparked-QAnon-n900531>