QANON ON TWITTER: AN OVERVIEW

Institute for Data, Democracy & Politics
THE GEORGE WASHINGTON UNIVERSITY

Sam Jackson
Brandon Gorman
Mayuko Nakatsuka
University at Albany

Abstract

This report provides an overview of activity related to QAnon on Twitter, using a novel dataset of nearly 85 million tweets sent between July 9, 2018 and October 31, 2020. The report focuses on tweets that Twitter identified as being in English; future work will examine the portion of the dataset in other languages. We present results of several preliminary lines of inquiry. First, we validate the dataset by identifying irrelevant tweets that were mistakenly collected. We then describe how many of the tweets express belonging to the community of adherents of the QAnon conspiracy theories, how many express criticism of QAnon, and how many instead use QAnon-adjacent terms to express belonging to the community of Trump supporters. Next, we summarize the dataset, examining features such as the number of tweets per day and the number of retweets per tweet. We attempt a preliminary characterization of the user base using data drawn from user bios and present the most frequent words and two-word phrases that appear in the text of tweets. Finally, we note several areas for future work.

Introduction

Since the election of Donald Trump to the presidency, a new conspiracy theory has emerged in the United States. According to this theory, a high-ranking federal official known only as “Q” provides coded hints about covert action taken by Donald Trump against sinister elites who engage in malevolent behaviors ranging from planning a coup to running an international child sex trafficking ring. QAnon adherents have been active on social media platforms including fringe sites like 4chan, where the first “Q drops” appeared, as well as mainstream sites like Twitter and Reddit, where users post thousands of QAnon-related messages each day.

Broadly speaking, conspiracy theories are “an effort to explain some event or practice by reference to the machinations of powerful people, who attempt to conceal their role” (Sunstein & Vermeule, 2009, p. 205). These powerful people are often depicted as outsiders (ranging from the financial elite to the Illuminati to those who would orchestrate a one-world-government under the New World Order) who ruthlessly pursue their own benefit at the expense of everyday people (Barkun, 1996, p. 51). Though conspiracy theories are often highly complex in their detail, they provide an explanation for things that are difficult to understand by describing them as the result of evil actions by a group of secret actors.

Conspiracy theories pose numerous social and political problems. For example, contrary evidence cannot change followers’ minds easily, and they are resistant to anything that invalidates their beliefs. Conspiracy theorists tend to interpret new evidence in favor of confirming their beliefs (Sunstein & Vermeule, 2009). Additionally, conspiracy theories can have detrimental effects on political discourse (DiGrazia, 2017; Introne et al., 2017; Pitcavage, 2001) and even motivate violence and criminal acts (Pitcavage, 2001; Sunstein & Vermeule, 2009), as has been the case with QAnon.

Scholars have investigated both individual traits and structural conditions that are associated with conspiratorial beliefs. Individual traits (such as political cynicism, negative attitudes towards authority, and feelings of social alienation) and demographics (including age and gender) can partly explain conspiratorial ideation (Swami et al., 2010, 2011). Those who believe in one set of conspiracy theories tend to accept other conspiracy theories (Barkun, 2016; Swami et al., 2010, 2011). In addition, macro-level social conditions and structural social changes such as immigration and unemployment that invoke feelings of fear and insecurity are associated with increased interests and beliefs in conspiratorial theories (DiGrazia, 2017). Outrage and fear triggered by social and political events can encourage people to turn to conspiracy theories for explanations (Sunstein & Vermeule, 2009). Furthermore, there is a strong link between political extremism and conspiratorial ideation (Berger, 2018; van Prooijen et al., 2015).

While conspiracy theories predate the internet, online forums and social networking websites facilitate their diffusion (Berger, 2018; Introne et al., 2017). The QAnon conspiracy theory emerged in October 2017 on 4chan, an anonymous message board, when a user identifying themselves as “Q” claimed that President Trump was taking covert action against nefarious enemies (Martineau, 2017). Argentino and Amarasingam (2020) describe QAnon as “a conspiratorial and anti-establishment ideology rooted in a quasi-apocalyptic desire to destroy the existing, corrupt world and usher in a promised golden age.” QAnon followers believe that a group of pedophiles including well-known Democrats and Hollywood celebrities operate a child sex trafficking ring (Roose, 2020). They venerate President Trump as a hero who is secretly working to defeat the “deep state” (Sardarizadeh & Goodman, 2020). QAnon followers have been active on a wide range of social media platforms, including mainstream platforms such as YouTube, Twitter, Facebook, and Reddit as well as alternative platforms such as Parler, Gab, and Voat (Smith, 2020; Timberg & Stanley-Becker, 2020).

Although it is unknown how many people actually subscribe to the QAnon conspiracy theory, it has spread quickly since its inception. Journalists believe that there are hundreds of thousands of followers (Martineau, 2017; Roose, 2020). A recent NPR/Ipsos poll shows that 17% of respondents believe that “a group of Satan-worshipping elites who run a child sex ring are trying to control our politics and media,” and another 37% don't know whether that is true (Rose, 2020). QAnon has even made its way into the mainstream political arena. According to Kaplan (2020), ninety-seven QAnon supporters ran for Congress during the 2020 election cycle, and two of them won. Furthermore, QAnon has also gained a considerable following outside the United States, mainly in Europe, Japan, and Latin America (Scott, 2020; Zimmerman, 2020).

In 2020, several social media companies independently decided that activity supporting the QAnon conspiracy theory causes harm, with YouTube, Twitter, and Facebook eventually removing thousands of QAnon accounts (Ortutay, 2020). An investigation by Facebook shows that QAnon groups had millions of members on the platform before they were removed (Sen & Zadrozny, 2020). However, even after these platforms took action, thousands of QAnon accounts remained on Twitter and Facebook (Collins & Zadrozny, 2021; Timberg, 2020). After the insurrection in Washington, D.C. on January 6, 2021, a number of platforms, including Twitter, engaged in another round of aggressive moderation against QAnon supporters (Collins & Zadrozny, 2021).

In addition to spreading false information online, QAnon followers have also engaged in a number of harmful activities offline. Several adherents have been charged with violent crimes (Amarasingam & Argentino, 2020; Roose, 2020). In June 2018, a 30-year-old armed man was arrested on several charges including making terrorist threats after blocking traffic on a bridge, demanding that the government "release the OIG report," referring to a QAnon idea that an unredacted Department of Justice inspector general report would expose the "deep state" (Baer, 2018). In March 2019, a 24-year-old man allegedly murdered a mafia boss because he thought the victim was a member of the “deep state” (Watkins, 2019). In April 2020, a 37-year-old woman was arrested for threatening to kill Joe Biden (Sommer, 2020). In January 2021, QAnon followers were spotted among the large mob that stormed the US Capitol, including one who was shot and killed by Capitol Police (Notopoulos & Lytvynenko, 2021). A Bellingcat investigation of the killed woman's Twitter account reveals that the she underwent a radicalization process: "a potent MAGA online subculture appears to have transformed this former Obama voter, who turned to Trump over a dislike of Hillary Clinton, into a QAnon follower ready to storm the Capitol" (Bellingcat, 2021). The FBI has identified conspiracy theories including QAnon as a new domestic threat. New York Times columnist Farhad Manjoo (2021) writes: “This level of influence isn’t going to disappear at noon on Jan. 20. QAnon’s vast reach, and Trump’s deep hold on it, are here to stay.”

An interlude on ethics, transparency, and reproducibility

Researchers have increasingly wrestled with the ethical implications of studying online activity (franzke et al., 2020), especially when this activity can be construed as harmful. In this section, we outline our approach to the ethics of this research project. Not all of this section applies directly to this report.[1]

Typically, researchers studying offline phenomena are not obligated to obtain informed consent when observing individuals acting in public spaces. For the purposes of this project, we consider open activity on Twitter to be analogous to public offline behavior. We believe that we do not need informed consent from the Twitter users who appear in our dataset, as our use of Twitter's API means that we only collect tweets from accounts that are not private. This is particularly true given that this project is based on "pure" observation: we do not interact with any users or tweets related to our dataset.

We consider the issue of whether to identify users as separate from the issue of whether we can study their activity without informed consent. This again draws parallels with typical ethical obligations for offline research, where researchers generally do not identify the individuals they study (with notable exceptions for politicians and other public figures). Studying harmful activity presents additional ethical complications in that some individuals who might be identified in research that depicts them as bad actors welcome such identification as a means for self-promotion (Berger, 2018, pp. 21–22).

Two strategies are commonly used to avoid identifying individuals: presenting only aggregated data and results or presenting only de-identified data and results.[2] In this project, we choose not to present data that would allow for the easy identification of users without a compelling analytical reason to do so. Generally speaking, we do not provide usernames. If reporting the results of our analysis would be strengthened by identifying specific users, we would consider doing so only in one of two cases: if the user has more than 5,000 followers or if the user is verified. We assume that users who meet either of these criteria are more likely to anticipate that people are paying attention to their tweets and that they could receive additional attention for those tweets.

---

[1] The University at Albany Institutional Review Board (IRB) determined that this project does not need IRB review because it does not meet the definition of human subjects research. IRB submission #5825. We consider this IRB determination to be an administrative requirement but not an indicator that there are no ethical questions that require further thought.

[2] De-identification procedures come in different degrees of thoroughness depending on the type of data, domain of research, and other considerations. Additionally, de-identification has repeatedly been shown to be difficult bordering on impossible in many circumstances (Hafner, 2006; Zimmer, 2010).

There are two categories of users identified in this report. The first category consists of users whose handles most commonly appear in the user bios in our dataset, all of whom are verified (or were, in the case of users who had their accounts suspended by Twitter). The second category consists of users whose handles contain our collection terms.

In addition to not specifically identifying users unless they meet one of these criteria, we largely do not provide information such as full-text tweets or user bios that would allow readers to easily identify individual users. Instead, we generally either paraphrase tweets (or bios) or create fictional composite tweets by combining quotes from different tweets. In the text, we clearly signal whether a tweet is a direct quote, a paraphrase, or an illustrative fictional composite tweet. In this report, the only such information that we provide is the top handles, unigrams, and bigrams that appear in user descriptions, along with the top unigrams and bigrams that appear in the text of tweets. However, none of this information is sufficient to identify specific users in our dataset.

This approach to research ethics complicates efforts to provide transparency and reproducibility for this project. To balance transparency and reproducibility with this ethics approach, we have created a GitHub repository[3] that contains four elements: (1) a description of the contents of the repository; (2) a link to a compressed file containing all of the tweet IDs for tweets in this dataset; (3) several Python files containing code used to generate the descriptive statistics and visualizations in this report; and (4) a YAML file containing information about the Python environment used to run the code.

---

[3] https://github.com/albany-social-media-analysis/QonTwitterIDDP

Data collection

The dataset for this project was collected using STACKS (Hemsley et al., 2019), an open-source tool written in Python to collect data from Twitter's filtered stream API, do some light processing of that data (for example, converting time from textual data to native datetime data), and store the data in MongoDB. In total, the dataset contains nearly 105 million tweets.

The filtered stream API[4] provides a persistent synchronous connection to Twitter. As long as the connection is open, tweets matching the parameters specified in the API call are delivered in near-real-time up to an unspecified limit (around 1% of the global volume of tweets at any given time; it seems that the relevant time period for this is sub-second, though exact information is not available). To build this dataset, we use collection terms as "track" parameters. We do not use any other parameters (such as language or location) in our API call. This means that Twitter returns any tweet that contains at least one of the terms used as a track parameter, whether that term appears in the text of a tweet, as a hashtag, as an at-mention, or in a hyperlink.

Initially, the data collection process began with one collection term (qanon), with a second (pizzagate) added a few days later. This dataset was originally built as a test case for FlockWatch, an open-source tool built by Jackson to identify changes in language in longitudinal text datasets (Jackson 2018; 2019). FlockWatch has been run consistently with this dataset since July 2018. It provides lists of trending unigrams and bigrams (i.e., one- and two-word phrases); it also provides a list of unigrams that co-occur with existing collection terms. Importantly, this tool does not make any decisions about whether to add new collection terms; instead, it provides recommendations in the form of the output lists of trending terms and co-occurrences, which researchers can then review and use their domain expertise to determine which of the trending and co-occurring terms to add, if any. For this project, the output lists were regularly reviewed and individual unigrams and bigrams from the output were added as new collection terms as appropriate. After starting with two collection terms, there were 176 collection terms in use at the end of October 2020 (see Appendix A for a list of all collection terms and the date that each term began to be used as a collection criterion).

---

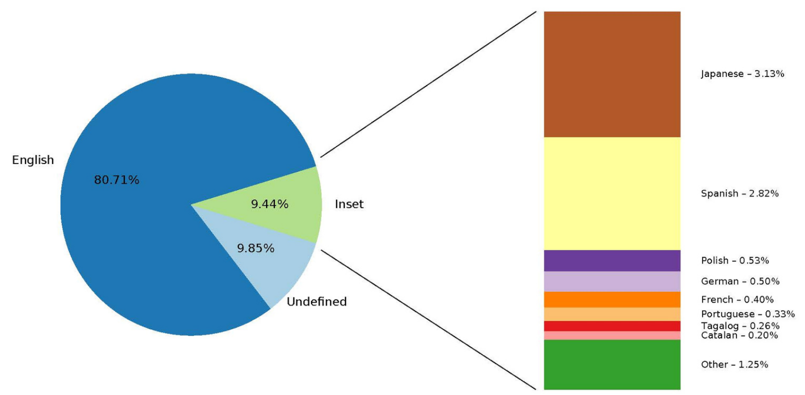
[4] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/overview

The initial data collection process began on July 9, 2018. With any long-term real-time data collection effort, there are bound to be gaps (for example, if an API connection breaks or a server crashes). We do not have complete logs that contain information about API connection breaks or server crashes; thus, we have done a post-hoc analysis to look for substantial gaps in tweet creation timestamps. In total, there are only six gaps of one hour or more. One gap is more than one day in duration: no tweets were collected from December 6, 2018 at 6:27 p.m. to December 10, 2018 at 1:55 p.m. The average number of tweets collected each day in December 2018 is 100,499, suggesting that approximately 383,041 tweets were lost during this gap.

When using the stream API, if more tweets match the API parameters than Twitter will return, we receive a message indicating how many tweets matched our collection criteria that Twitter did not send to us. We are not confident in the reliability of this information about the number of lost tweets. Our best estimate is that we missed between 332,446 and 954,852 tweets during the time period under consideration (from 0.3% to 0.9% of the total dataset). Even at the high end of this range, this is at the level of statistical noise relative to the number of tweets collected.

The total number of tweets collected between July 9, 2018 and October 31, 2020 is 104,697,999. Of these, 84,500,492 (80.71%) are identified by Twitter as being in English. For this report, we set aside tweets in other languages for future work. As mentioned above, QAnon has spread around the globe (Scott, 2020; Zimmerman, 2020); we take up the question of international spread in future analysis.

Figure 1. Proportion of tweet languages as identified by Twitter



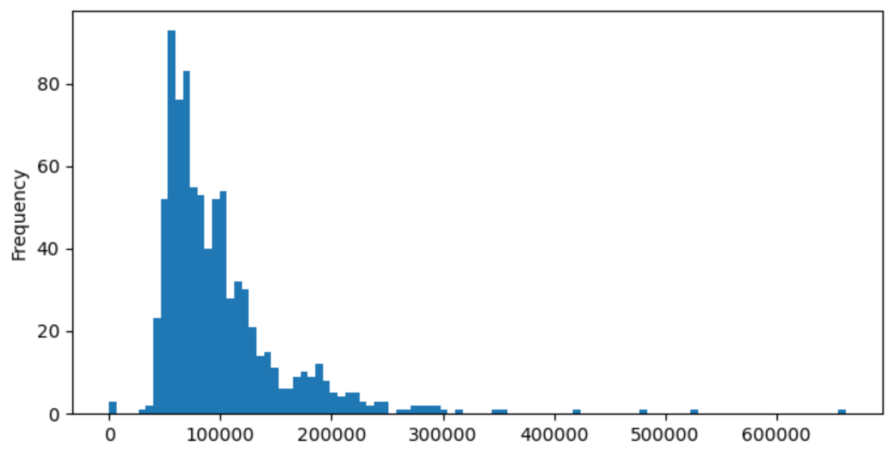
Overview

The dataset discussed here consists of nearly 85 million tweets sent over the course of 845 days. The number of tweets sent per day varied, from a non-zero low of 30,102 (on July 9, 2018, the first day of collection) to a high of 662,012 (on July 22, 2020, the day that Twitter announced wide-ranging moderation effort aimed at pro-QAnon activity). Most days (534 out of 845) saw fewer than 100,000 tweets, with the median number of tweets sent in a day being 84,057. Of the 309 days with more than 100,000 tweets, 47 had more than 200,000. This information is summarized in Table 1 and Figure 2.

Table 1. Distribution of the number of daily tweets

Mean	99,882
STD	58,297
Min	0
25%	62,910
50%	83,996
75%	117,289
Max	662,012

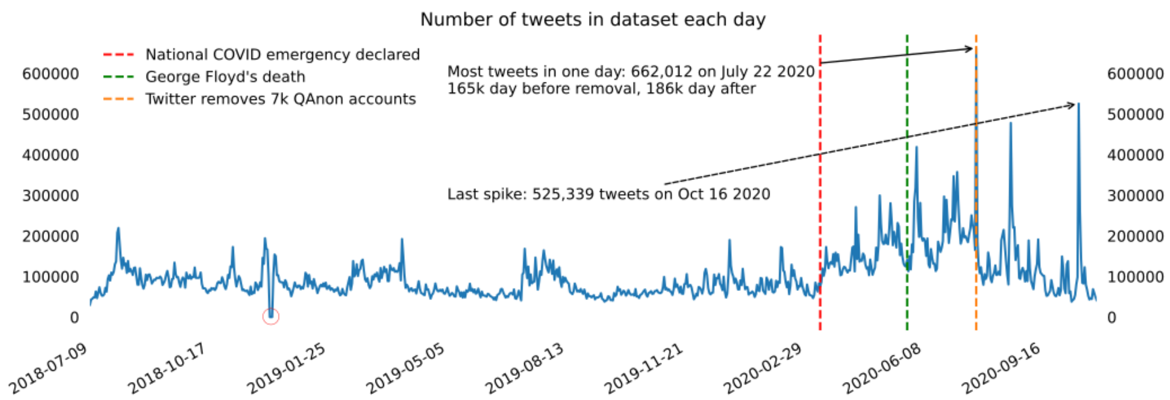
Figure 2. Histogram of the distribution of the daily number of tweets



As seen in Figure 3, there is substantial variation in the number of tweets per day. Prior to March 2020, there were occasional sudden spikes in volume. In some cases, the volume of tweets per day quickly returned to a baseline below 100,000. In other cases, the return to baseline was more gradual. From March 2020 through the end of October 2020, the volume of tweets related to QAnon was generally much higher than in the prior period, with the median number of tweets per day being 130,057; at the same time, it was more volatile, with frequent large changes in tweet volume from day to day. At first glance, we think that this increase in volume is related to two primary offline events: the increasing prominence of COVID-19, particularly after President Trump declared a national emergency on March 13, 2020 (Trump, 2020); and the widespread protests that emerged after George Floyd's death on May 25, 2020 (Bailey et al., 2020).

Early in the morning on July 22, 2020,[5] Twitter announced that the platform would step up moderation activity around QAnon, citing a potential for QAnon activity "to lead to offline harm" ("QAnon," 2020; Timberg, 2020; Twitter Safety, 2020). Media outlets reported that 7,000 accounts had been banned and some 150,000 faced suspensions. This moderation activity led to a large but short-lived uptick in the volume of tweets about QAnon. On July 21, we collected approximately 165,000 tweets. On July 22, we collected more than 650,000. On July 23, our collection dropped back down to around 185,000 tweets.

Figure 3. Volume of tweets sent over time



The small red circle in the lower left side of the figure marks the almost 4 day period for which we have zero tweets

Relevance

Before analyzing the content of the data, we evaluate the extent to which our collection criteria produced false positives during the data collection process. Given the flexibility of language, some words that are highly suggestive of QAnon could also be used to refer to other things that are completely unrelated to QAnon. We used our best discretion when reviewing FlockWatch output and choosing collection criteria. But to confirm that our dataset is in fact about QAnon, we chose to determine what portion of collected tweets are not related to QAnon.

We began this process with a close reading of several rounds of several hundred randomly sampled tweets to get a feel for their content. Over the course of these rounds, we gradually built a codebook to guide our labeling of tweets as relevant (i.e., true positives) or irrelevant (i.e., false positives). In the end, we use the following process to identify the proportion of tweets in the dataset that are relevant and the proportion that are irrelevant.

---

[5] Twitter announced this moderation activity in a series of tweets sent at midnight UTC, which is 8pm EDT or 1am British Summer Time. Most media reports of the moderation activity were released on July 22. Timestamps as reported here are in UTC.

First, we filtered out obvious false positives that were collected because of the temporary use of bad collection criteria. Two phrases are particularly important for QAnon on Twitter: "Q sent me" (or "Q sent us") and "We are Q." During our preliminary analyses, we realized that Twitter treats spaces in track parameters as logical OR operators rather than AND operators. For example, using "we are q" as a collection term leads us to capture a tweet like "Q: what are we doing on Memorial Day?" This sometimes leads to many false positives in foreign languages. For example, using "q sent" as a collection term leads to collecting many Spanish language tweets that contain "q" (short for que) and "sentir" (a verb meaning "to feel"), such as "sentí q dormi 5 minutos nomas" ("I feel that I slept for 5 minutes"). This first automatic filtering confirms that a collection term is present in the text of a tweet; tweets that do not contain a collection term are flagged as false positives.

Next, we found that 109 of our collection terms overlap with user account handles (e.g., @qanon; see Appendix B for a full list). This could increase false positives, as any tweets that mention these accounts (and any retweets of such tweets) would be collected and included in our dataset regardless of their content.

Because of the relatively small number of these account handles, we reviewed each user profile and a subset of their mentions to decide whether the handle should count as a sufficient indicator that a tweet mentioning that handle is relevant for this dataset. We consider any account handle as sufficient if it meets any of the following six criteria[6]:

1. Users whose biography, avatar, or header photo clearly evoke QAnon or other conspiracy theories.
2. Users whose activity is primarily about topics that are relevant to QAnon or other conspiracy theories.
3. Users whose handles are central to QAnon or other conspiracy theories or are unambiguously about QAnon or other conspiracy theories and are not active at all or have protected accounts.
4. Users whose handles are central to QAnon or other conspiracy theories, are active, but are not active in QAnon discourse at all.
5. Users who are @mentioned (nearly) exclusively in the context of QAnon or other conspiracy theories.
6. Users whose handles are unambiguously relevant but whose accounts are suspended.

[6] Because we use the track parameter of the streaming API, we do not automatically collect tweets sent by accounts whose handles appear in our collection criteria. We only collect tweets from such accounts if the text of those tweets (including hashtags and URLs) also contains at least one collection term.

We ignore any account handle if it meets one of the following two criteria:

1. Users whose handles are relatively peripheral to QAnon or other conspiracy theories and are not primarily active in QAnon or conspiracy-related topics.
2. Users whose handles are peripheral to QAnon or other conspiracy theories and are not active at all.

Each member of the team reviewed every account independently to come to a determination. Any disagreements were settled via majority vote during a follow-up meeting. At the end of this process, we were left with 86 account handles that we treat as sufficient indicators for distinguishing true positives from false positives. For those accounts that we deemed not to be sufficient indicators, tweets mentioning those accounts were not automatically deemed to be relevant.

Finally, excluding those tweets that at-mention an account handle deemed sufficient to indicate that a tweet is relevant to QAnon, two members of the team hand-coded a random sample of 925 tweets to evaluate their relevance to QAnon discussions. We coded a tweet as relevant if it meets any of the following criteria:

1. It contains the hashtag “#QAnon”.
2. It contains “QAnon”, “Q”, or “Anons” in a meaningful sentence.
3. It contains two or more keywords related to QAnon that are not “QAnon”.
4. It contains “QAnon” and another keyword related to conspiracy theories.
5. It contains Trump-related language (e.g., “MAGA” or “KAG”) and any term related to QAnon.
6. It contains references to magic, extraterrestrials, or supernatural beings.

High inter-rater reliability scores (pairwise agreement: 99.41%, Krippendorf's alpha: 0.82) indicate broad agreement between the coders. This suggests that approximately 2% of tweets collected are false positives. Because this is a small percentage, we consider them statistical noise and leave them in the dataset.

Affinity

Next, we turn our attention to the content of the tweets themselves. Based on an abductive assessment of the content of the tweets we analyzed during relevance coding, we identified five broad, mutually-exclusive categories of tweet:

1. Tweets that express belonging to the online QAnon community.
2. Tweets that express belonging to the wider Trump-supporting or politically conservative community but not the online QAnon community.
3. Tweets that express criticism of the online QAnon community or the QAnon conspiracy theory.

4. Tweets that are affectively neutral statements of facts about or requests for information about QAnon or related conspiracy theories.
5. Tweets with insufficient information to clearly fit into one of the first four categories.

The same two team members that performed relevance coding categorized these tweets by looking only at the text of the tweet itself without looking at the contents of links, images, and reply chains.[7]

The first category, expressing belonging to the online QAnon community, describes tweets that contain a clear, positive engagement with conspiracy theories. This is the most important of the substantive categories for the purposes of this report. We approach this by developing a list of words and expressions that indicate a high probability that the user intended the tweet to contribute to or otherwise support QAnon-related theorizing. These include any supportive mention of QAnon or related conspiracy theories (e.g., “Seth Rich”, “adrenochrome”, “pizzagate”, “pedowood”); allusions to an “awakening”, “storm”, “hidden enemy”, “plan”, or esoteric “truth”; references to “mind control”, “clones”, “body-doubles”, “demons”, “Satan”, or “Biblical” future events; apparent attempts to disseminate or “decode” supposed messages from Q (“qdrops”); supportive mentions of Gen. Michael Flynn, the phrase “2Q2Q”; or mentions of far-right conspiratorial media outlets like Gateway Pundit. We evaluate tweets holistically, as these criteria do not rise to the level of sufficient indicators of expressing belonging to the online QAnon community.

The second category, expressing belonging to the community of Trump supporters or conservatives, describes tweets that contain one or more of our collection terms, clear indicators of support for Trump, the Republican Party, or conservative groups and/or criticism of prominent liberals, the Democratic Party, or liberal groups, but no clear indicator of belonging to or support for the online QAnon community or engagement with conspiracy theories. Empirically, most of the tweets that fall into this category contain “WWG1WGA” as their only relevant collection term; this slogan appears to have been adopted by a number of mainstream conservatives and Trump supporters with no clear participation in conspiracy theory communities.

---

[7] A brief and unrepresentative check of tweets with media suggests that the presence of media rarely would change this categorization process; future work will examine tweets with links and embedded media in a robust manner.

The third category describes tweets that express criticism of the online QAnon community. The content of these tweets ranges from mockery (e.g., “QAnon cultists”, “wackos”, “crazies”) to claims that QAnon is a distraction from some other, more insidious conspiracy. In theory, this category is not mutually exclusive with the two belonging categories described above but empirically, tweets that contain both expressions of belonging and criticism are rare enough (about 0.2% of tweets) that we treat the two categories as mutually exclusive.

The fourth category describes tweets with a neutral valence. These are either requests for information about, or informational tweets on, QAnon or related conspiracy theories. In order to fit into this category, there must be no expression of belonging or criticism in the text of the tweet.

Finally, the fifth category describes tweetsthat do not contain enough information to fit into the other four categories. These tweets are often vague, ambiguous, or mixed in their affective content. Many are single words or common phrases followed by a hashtag that contains a collection term (e.g., “Thank you! #QAnon”). Others clearly express belonging or criticism, the target of which is unclear from the text of the tweet alone.

The coding procedure is as follows. First, we coded for not enough information—any tweet that falls into this category was not considered during coding for the other categories. Second, we coded for belonging to and criticism of the QAnon online community. Any tweet that is coded as belonging to or criticism of the QAnon online community is not considered during coding for belonging to the community of Trump supporters. Finally, any tweet that was coded as not fitting into any of the other categories is treated as neutral by default.

Coding was completed by two team members on a random sample of 449 tweets. Inter-rater reliability scores for most of these categories, shown in Table 2, fall short of the 0.80 suggestion for Krippendorf's alpha. For the purposes of this report, we discussed each tweet about which the coders disagreed and came to a consensus decision regarding the final category. The results that follow are based on our coding of this representative sample of tweets.

Table 2 describes the relative volume of tweets that fall into each category, along with inter-rater reliability information. The majority of tweets in our dataset are clear expressions of belonging to the QAnon online community while around one in ten is critical. A few more than one in ten express belonging to the community of conservative Trump supporters only, without any mention of QAnon or other conspiracy theories. All of these tweets contain the collection term "WWG1WGA" as compared to less than half of the tweets in the other categories. Coders agreed that about a quarter of the tweets did not contain enough information to categorize - empirically, many of these contain clear expressions of belonging to the community of conservative Trump supporters but vague or ambiguous relationships to QAnon; others express clear support or criticism, but the target of that support or criticism is unclear. Finally, 2% of tweets are neutral or purely informational.

Table 2: Substantive Tweet Categories and Reliability Information

Category	Percentage of tweets	Inter-rater reliability (Krippendorf's Alpha)
Belonging to QAnon	53%	0.72
Criticism of QAnon	10%	0.72
Belonging to Trump or conservative only	11%	0.83
Needs more information	24%	0.65
Neutral	2%	0.80

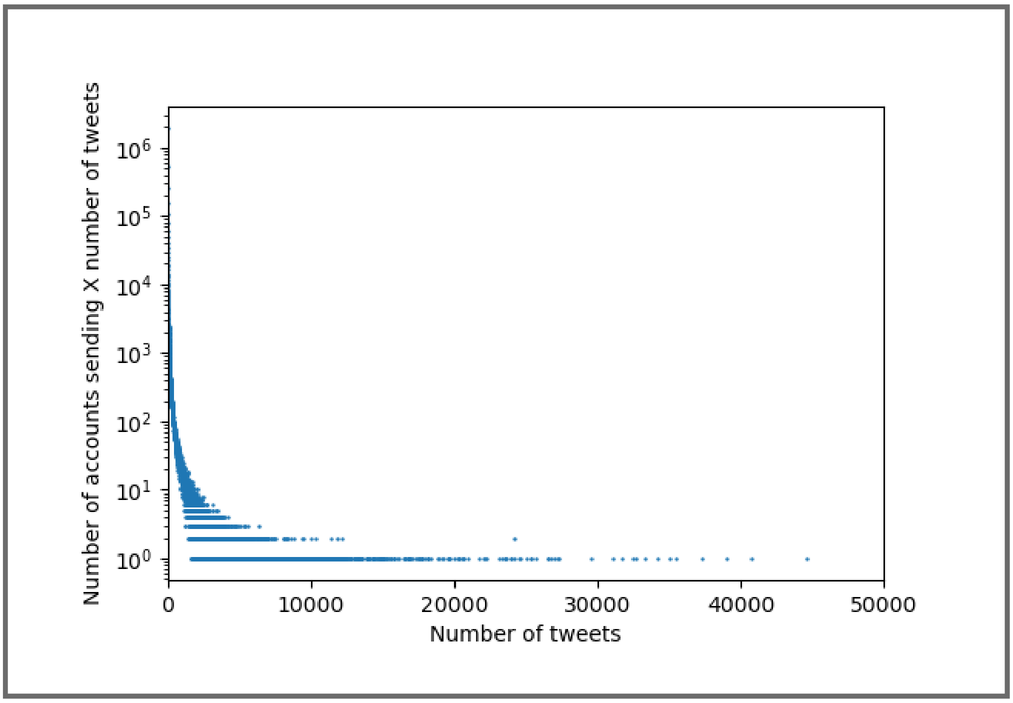
Participation

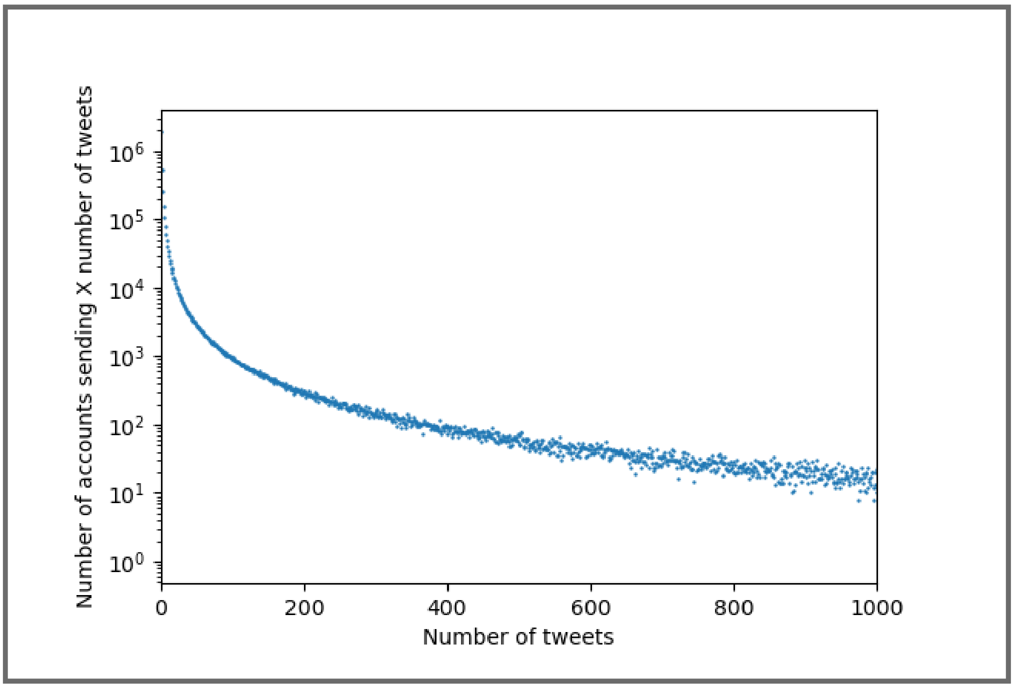
In our dataset of 84.5 million tweets, there are 3,792,037 unique users. As users can change both their display name (e.g., "BBC News (World)") and their screenname (e.g., @BBCWorld), we use the numerical user id string to identify accounts. Of these 3.8 million accounts, most (n=3,644,244, 96%) appear in our dataset with only one screenname. Seven users had more than 50 screennames in our dataset; 205 users had more than ten.

Table 3 and Figure 4 illustrate the distribution of the number of tweets per user in our dataset. As is common in social media datasets, there is a strong right skew in this distribution. A narrow majority of users tweeted only once in our dataset (n=1,936,081, 51.1%). In total, 2,987,446 users (78.8%) tweeted five or fewer times. Four users tweeted more than 50,000 times in the dataset, with the most prolific user in the dataset tweeting 96,984 times.

Table 3. Distribution of the number of tweets per user	
Mean	22.3
STD	236.1
Min	1
25%	1
50%	1
75%	4
Max	96,984

Figures 4a and b. Distribution of the number of tweets per user. Figure 4a excludes four extreme outliers. Figure 4b is a zoomed version of the far-left side of Figure 4a.





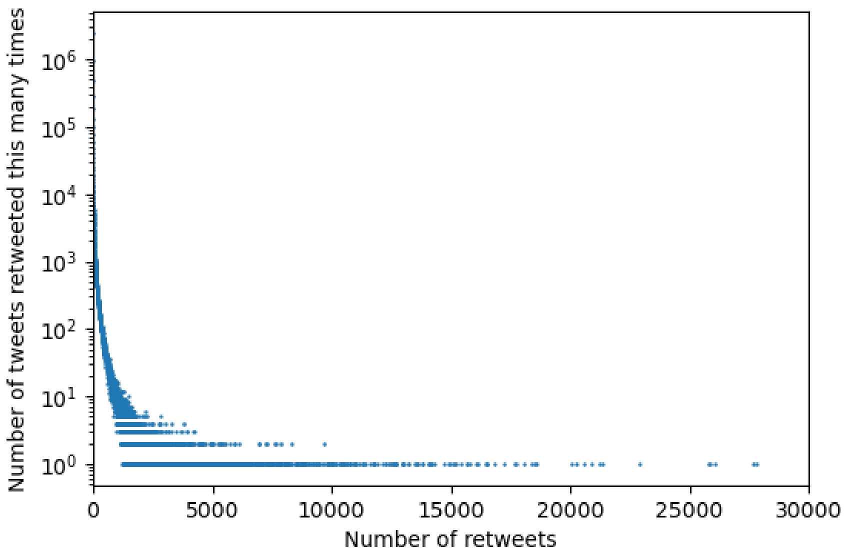
Of the 84,500,492 tweets in this dataset, 66,345,531 are retweets; 6,066,768 are quote tweets; and 12,088,193 are neither retweets nor quote tweets.

Table 4 and Figure 5 illustrate the distribution of tweet-level retweet volume in our dataset. In total, 5,365,755 tweets were retweeted at least once. As with the distribution of the number of tweets per account in this dataset, the number of retweets per tweet (excluding tweets that were never retweeted) is heavily skewed to the right. Of tweets that were retweeted at least once, 2,468,768 (46.0%) were retweeted only once; 933,773 (17.4%) were retweeted twice; and 472,089 (8.8%) were retweeted three times. 105,948 (2.0%) tweets were retweeted more than 100 times, 6,135 (.1%) were retweeted more than 1,000 times, and 107 (.002%) were retweeted more than 10,000 times. One tweet was retweeted 106,728 times; the next most retweeted tweet received 52,412 retweets.

Table 4. Distribution of the number of times tweets were retweeted

Mean	12.4
STD	131.2
Min	1
25%	1
50%	2
75%	4
Max	106,728

Figures 5a and b. Distribution of the number of times tweets were retweeted. Figure 5a excludes 5 extreme outliers. Figure 5b is a zoomed version of the far-left side of Figure 5a.



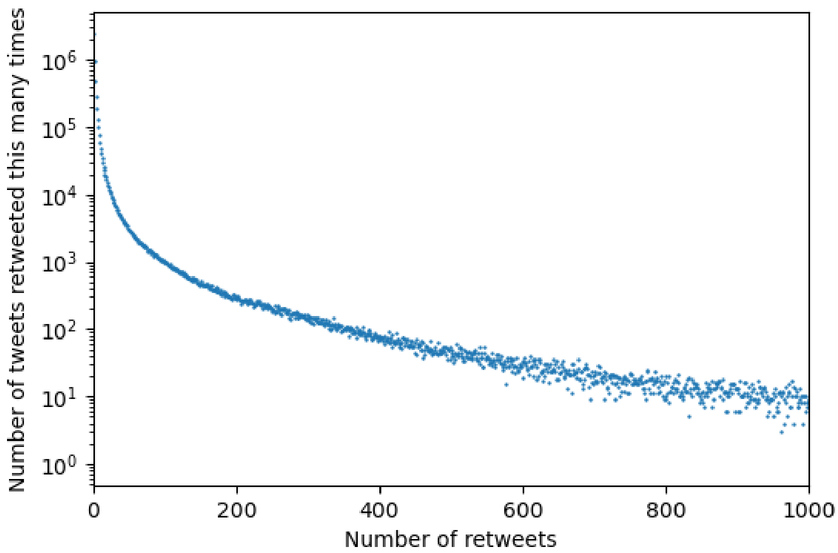
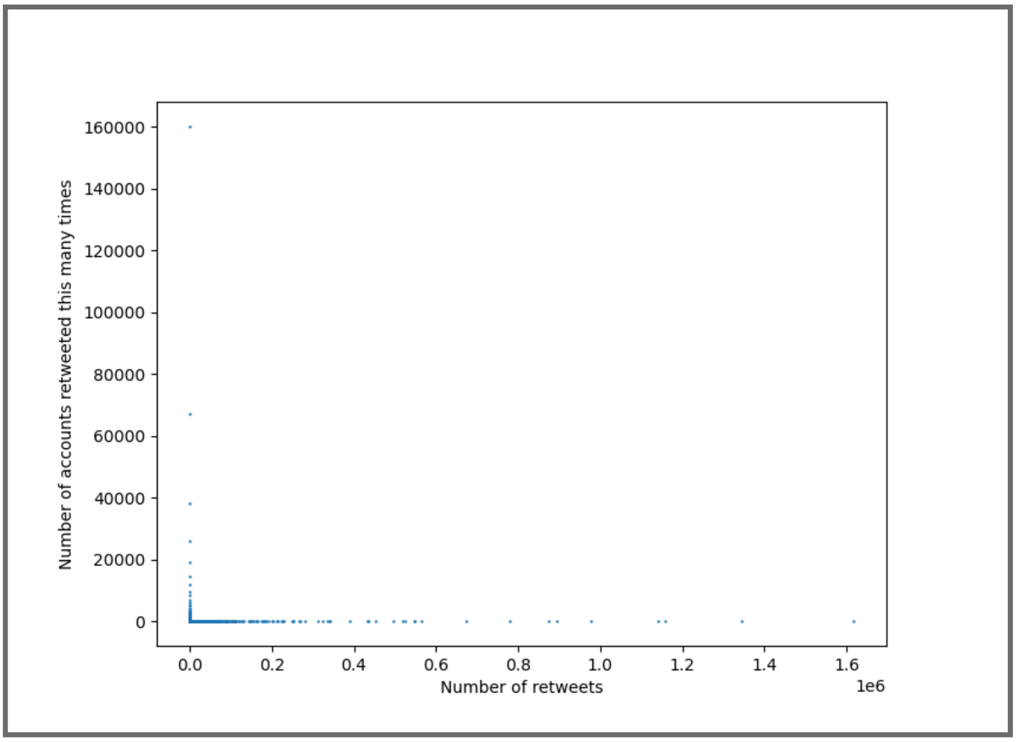


Table 5 and Figure 6 illustrate the distribution of account-level retweet volume in our dataset. The 5.3 million tweets that were retweeted at least once were sent by 489,244 different accounts. 160,049 (32.7%) accounts received only one retweet, 67,241 (13.7%) received two retweets, and 38,223 (7.8%) accounts were retweeted three times. Four accounts were retweeted more than 1 million times, 66 accounts were retweeted more than 100,000 times, 716 accounts were retweeted more than 10,000 times, and 3,533 were retweeted more than 1,000 times.

Table 5. Distribution of the number of times accounts were retweeted	
Mean	135.6
STD	5,653.9
Min	1
25%	1
50%	3
75%	11
Max	1,616,606

Figures 6a and b. Distribution of the number of times accounts were retweeted. Figure 6a shows the full distribution. Figure 6b is a zoomed version of the far-left side of Figure 6a.



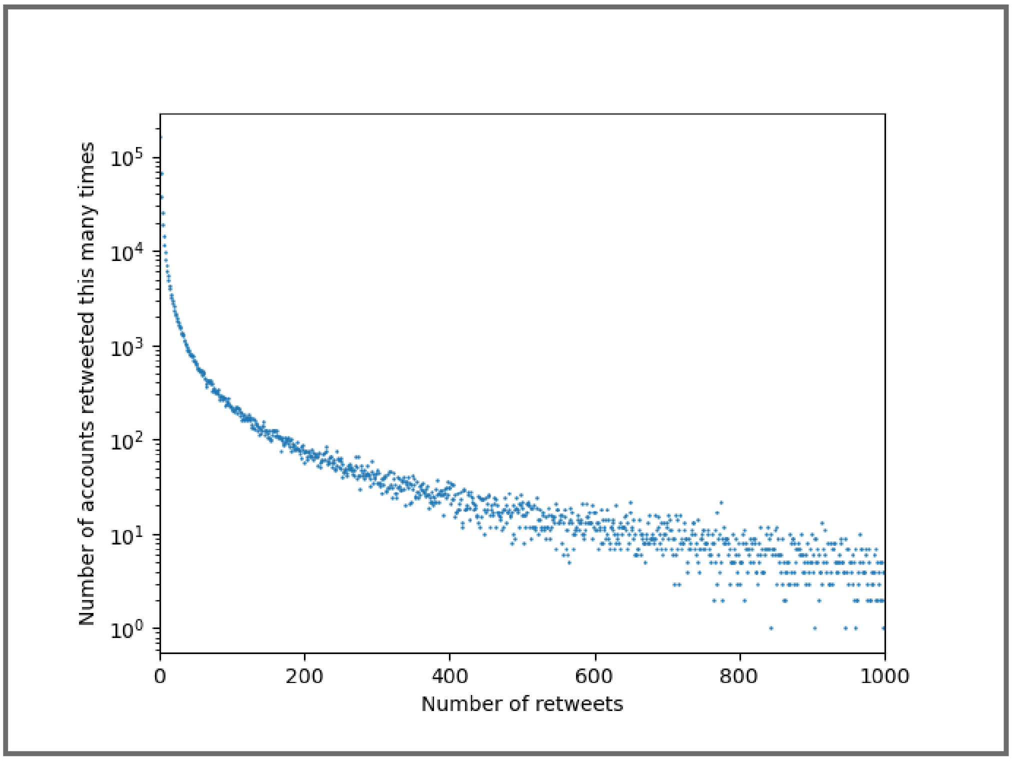
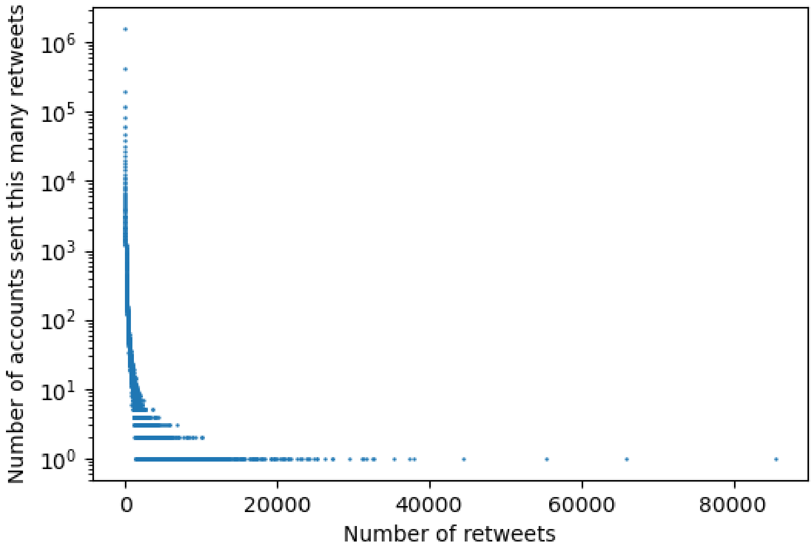
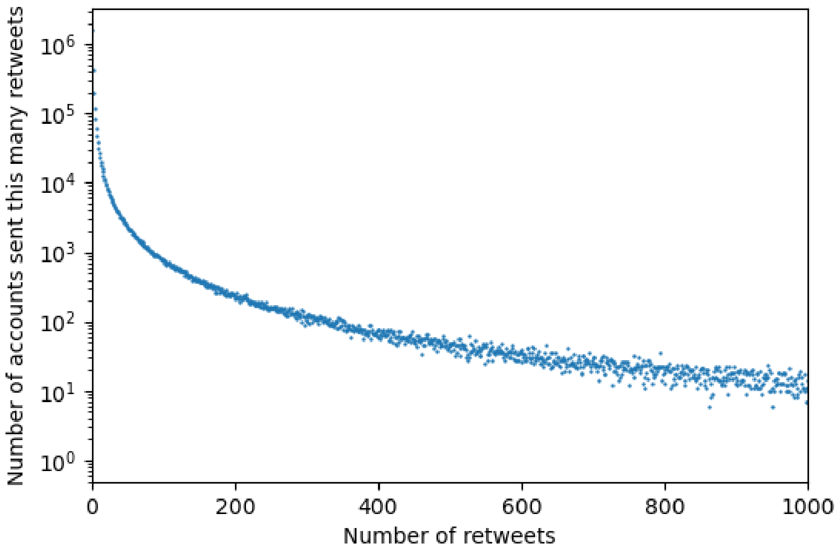


Table 6 and Figure 7 illustrate the distribution of the volume of retweets sent by accounts in our dataset. 3,041,622 accounts were responsible for the 66 million retweets. Once again, the distribution of accounts is heavily skewed to the right. 1,590,023 (52.3%) sent one retweet, 417,757 (13.7%) sent two retweets, and 197,455 (6.5%) sent three retweets. The most prolific retweeter created 85,524 retweets. 208 accounts sent more than 10,000 retweets, and 10,264 accounts sent more than 1,000.

Table 6. Distribution of the number of retweets sent by accounts	
Mean	21.8
STD	225.7
Min	1
25%	1
50%	1
75%	4
Max	85,524

Figures 7a and b. Distribution of the number of retweets sent by accounts. Figure 7a shows the full distribution. Figure 7b is a zoomed version of the far-left side of Figure 7a.





Users

Table 7 and Figure 8 show the distribution of accounts created per day in our dataset. The oldest five accounts in this dataset were created on March 21, 2006 (the day that Twitter itself was created). On average, 727 accounts in this dataset were created each day for the period March 21, 2006 through October 31, 2020. As shown in Figure 8, the number of accounts created per day is relatively steady, with occasional spikes.

Perhaps most notable are the days around June 1, 2020. The average number of accounts created per day in 2020 was 1,300 (median = 1,203). In May and June, the average number of accounts in our dataset created per day rose to 1,870 (median = 1,605). On May 28, 1,883 accounts were created, followed by days with 2,359, 2,372, and 4,097 new accounts. The most accounts created in a single day from March 2006 through October 2020 occurred on June 1, 2020, when 7,090 accounts that appear in our dataset were created. The next several days saw a rapid decline: 4,065 new accounts on June 2, then 3,053, 2,809, 2,523, and 2,044.

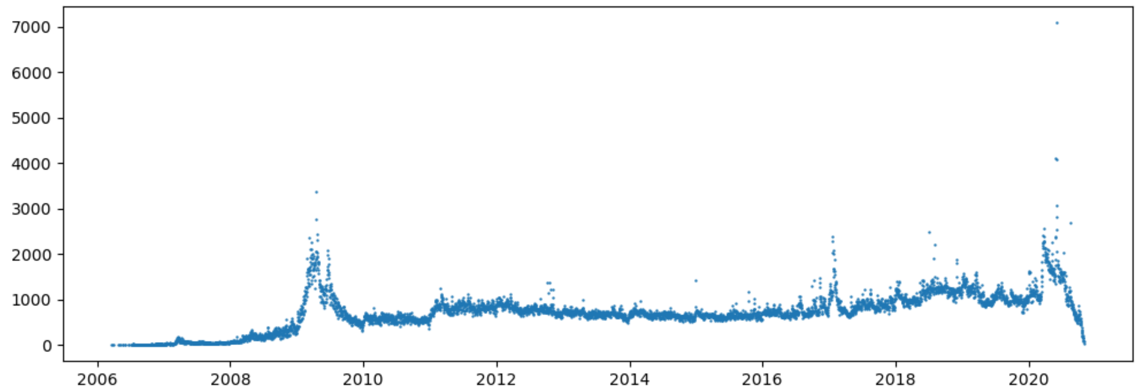
According to Twitter, eight accounts in this dataset were created at 00:00 (midnight) on January 1, 1970. This is a reference time frequently used in computing, called the UNIX Epoch [8]. This likely indicates corrupt data on Twitter's end for these eight accounts. We exclude these eight accounts from this description of when accounts were created; this is responsible for the difference between the number of unique accounts in this dataset and the sum of the number of accounts created on each day.

Table 7. Distribution of the number of accounts created per day

Mean	727.0
STD	417.8
Min	1
25%	575
50%	709
75%	920
Max	7090

[8] https://pubs.opengroup.org/onlinepubs/9699919799/xrat/V4_xbd_chap03.html#tag_21_03_00_20

Figure 8. Distribution of the number of accounts created per day



To provide preliminary insight into the users in this dataset, we examined the text of the user bio that most users provide. Users can update this field as often as they like. As mentioned before, a narrow absolute majority of users tweeted only once in this dataset (n=1,936,081, 51.1%); by definition, these users will only be associated with one user description. For the 1,855,956 users who tweeted more than once, 1,036,589 had the same user description associated with all of their tweets in our dataset. Thus, 2,972,670 (78.4%) users had one user description. Tables 8 and 9 show the distribution of unique user descriptions per user in our dataset.

Table 8. Distribution of the number of unique user descriptions per unique user

Mean	1.5
STD	1.9
Min	1
25%	1
50%	1
75%	1
Max	543

Table 9. Distribution of the number of unique user descriptions per unique user, excluding users with one unique user description	
Mean	3.5
STD	3.6
Min	2
25%	2
50%	2
75%	4
Max	543

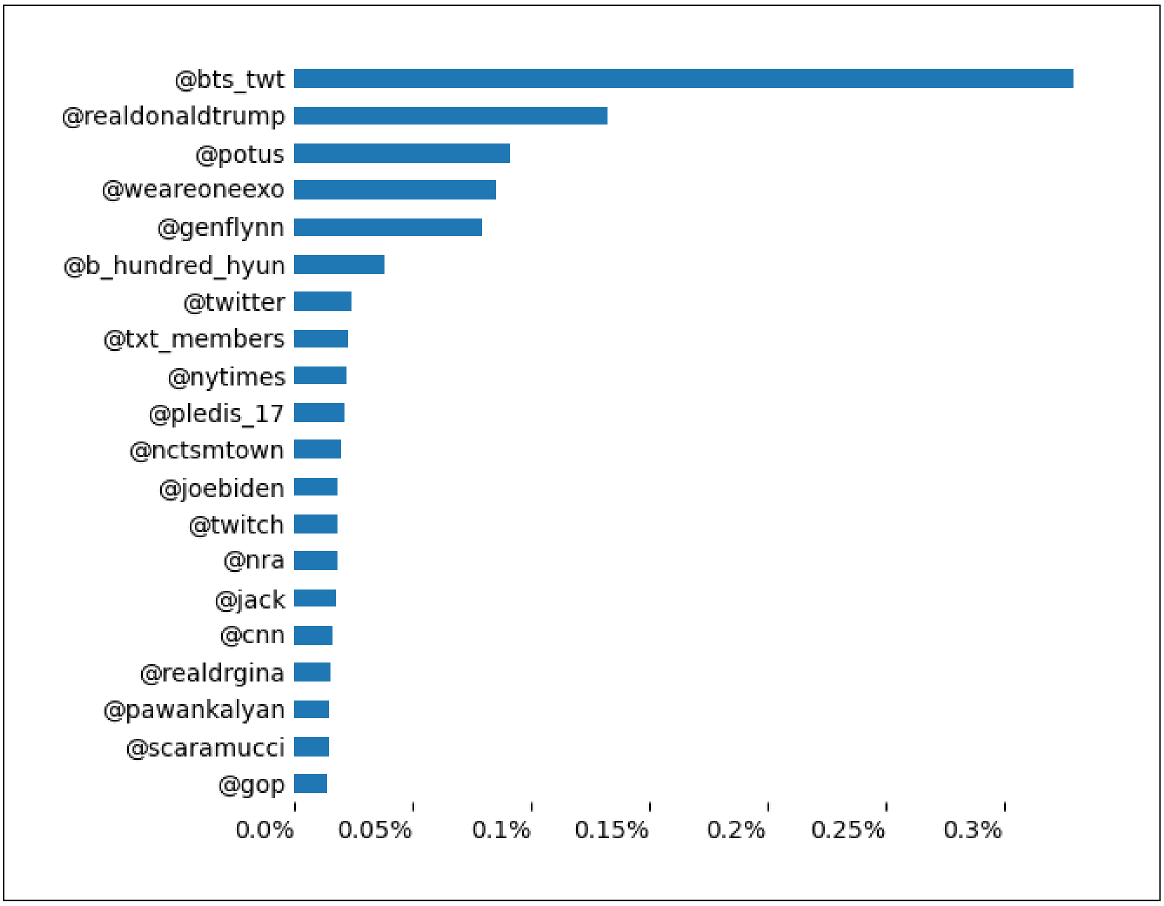
For the remaining 819,367 (21.6%) users who had more than one unique user description, we wanted to ensure that our analysis included all of the variations of their user descriptions without skewing the output to give them outsized influence. We created two unique corpuses to accomplish this. First, we created one composite user description for each user with a unique list of all of the unigrams (i.e., words) that appear in any of that user's descriptions.[9] We treat this list of composite user descriptions as the corpus for examining unigrams. Second, we used raw user descriptions to generate a list of bigrams, which we then filtered to remove duplicates for each user.

We provide two types of output from the corpus of composite descriptions: top at-mentions and top unigrams. For each type of output, we report prevalence rather than raw count.

First, we counted the number of user descriptions that include an at-mention of each user handle that was included in a user description (by identifying all of the unigrams in the corpus that begin with "@"). We converted that count to a prevalence—the top 20 at-mentions in user descriptions are in Figure 9. We removed "@" and "@gmail" from this output, as we believe these tokens indicate that someone is providing an email address rather than intending to at-mention a Twitter account.

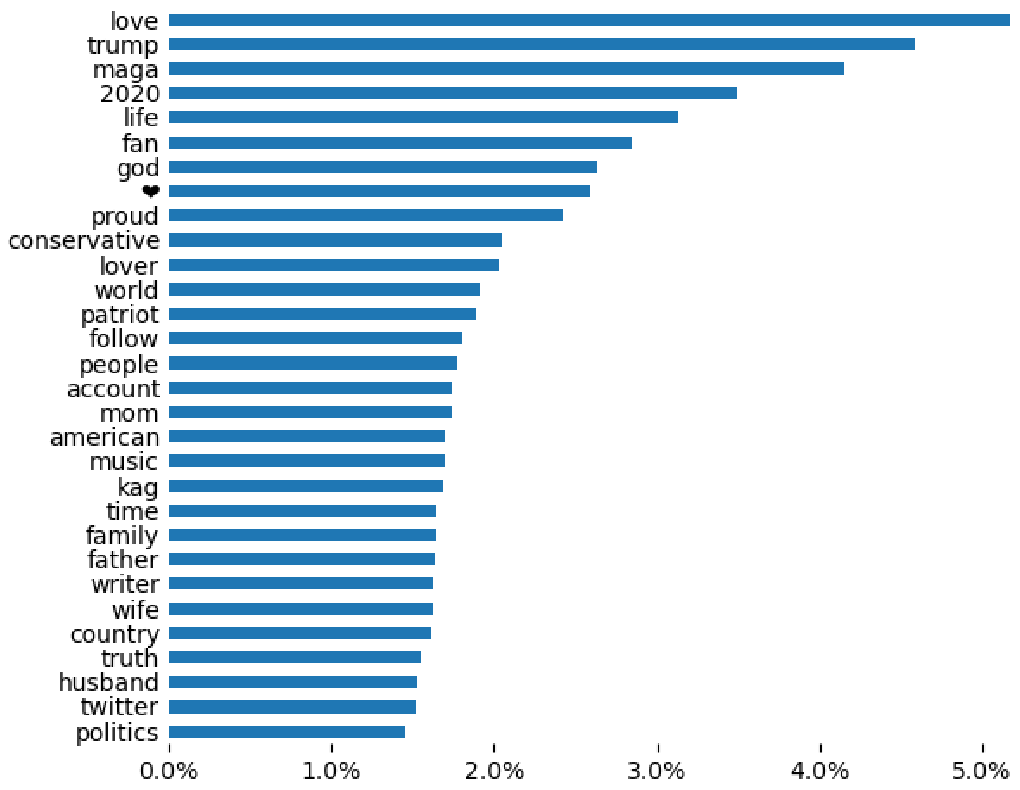
[9] Functionally, this means that we used Python to create a pandas dataframe with two columns (user id and user description) and one row for each tweet in the dataset. We dropped duplicates in the dataframe where a row was considered a duplicate only if its value in both columns matched another row's value in both columns. Next, we tokenized each remaining user description, putting that into a third column. We used groupby on user id to gather all of a user's descriptions together. Then, we joined all of the tokenized descriptions into one flat list of all of the tokens in all of the descriptions. Finally, we removed all duplicates from that list of tokens.

Figure 9. Most prevalent at-mentions in user descriptions



Second, we counted the number of user descriptions that include each unique unigram in the corpus. The top 30 unigrams in user descriptions are in Figure 10. We removed stopwords from this output, using the "English" stopword list provided as part of the NLTK Python package. We also removed punctuation and any unigram with a length of one if that unigram was a letter or number (but left in emojis). Finally, we removed several unigrams that didn't make sense on their own but that appear in the list of top bigrams.

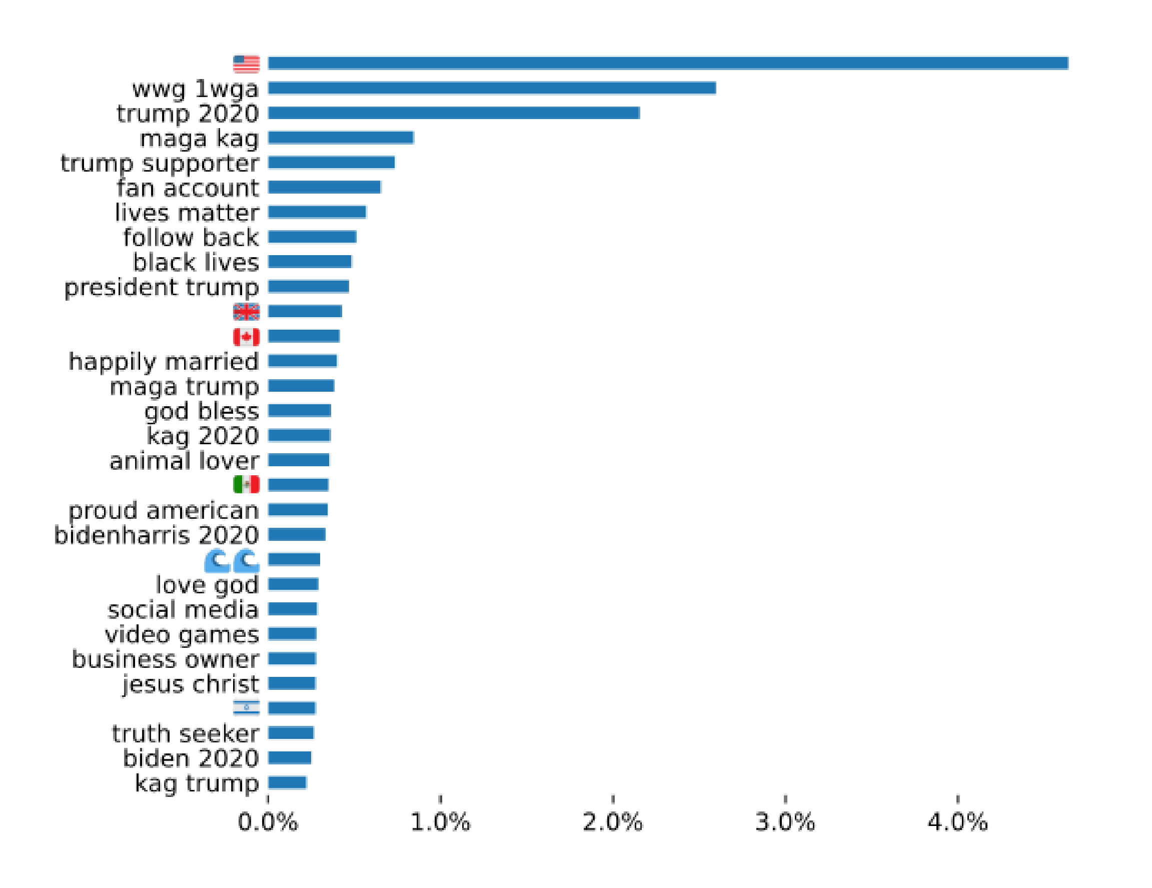
Figure 10. Most prevalent unigrams in user descriptions



Third, we counted the number of user descriptions that include each unique bigram (two-word phrase) in the corpus. The top 30 bigrams in user descriptions are in Figure 11. We removed bigrams if either unigram in the bigram was a stopword or was punctuation. We also removed bigrams if either unigram had a length of zero or was a hidden special character. Five of the top bigrams are country flag emojis; these can be rendered either as unigrams (the emoji flag) or as bigrams (two-digit region codes that are treated as two individual unigrams by NLTK's tweet tokenizer).[10] We show them here as emoji flags.

[10] https://en.wikipedia.org/w/index.php?title=Regional_indicator_symbol&oldid=994582770

Figure 11. Most prevalent bigrams in user descriptions



Taken together, these results give us a broad first look at the self-presentations of people who participate in QAnon-related discussions on Twitter. As we would expect, there are a number of indicators of political conservatism, Trump fandom, and expressions of adherence to conservative Christianity. Twitter accounts such as @realdonaldtrump, @potus, @gop, and @nra are among the most common at-mentions in user bios in our data, and words and common phrases including “trump supporter”, “maga”, “kag”, “conservative”, “god”, “patriot”, “husband”, “father”, “mother”, “wife”, “jesus christ” all point to social and political conservatism among many users. It is also clear from the prevalence of terms like “truth”, “truth seeker”, “wwg 1wga”, and at-mentions of @genflynn that many users perceive themselves as deeply engaged in conspiracy theorizing.

On the other hand, there are clear indicators of non-believers and even QAnon critics among the user base. We find ample evidence of K-Pop fans hijacking QAnon-related hashtags (Marcin, 2020); many of the top at-mentions are for popular K-Pop groups and their members (e.g., @bts_twt, @weareoneexo, @b_hundred_hyun, @txt_members, @pledis_17, @nctsmtown). We also have suggestive evidence of engagement by Democratic partisans and other left-leaning users, as @joebiden is among the top at-mentions and the phrases “bidenharris 2020”, “biden 2020”, and “black lives” are each among the top bigrams in our user bios.

The prevalence of at-mentions for @cnn and @nytimes as well as the word “writer” in user bios suggest that some of the tweets in our data are results of journalists and scholars covering the QAnon phenomenon. It is also plausible that the indicators of political liberals and journalists are actually intended as criticism of these categories in the user bios.

We also find, unsurprisingly, that QAnon-related Twitter discussions tend to be US-centric. We find the word “american” and “proud american” among the top features and the American flag emoji is the most common bigram in the user bios by a wide margin. Similarly, many of the most common features are in reference to American politics. On the other hand, we find a number of indicators that people outside of the United States are engaging in QAnon discussions online. In addition to evidence of K-Pop fans’ involvement, we also see a number of non-American flag emojis in our top bigrams (Mexico, Canada, Great Britain, and Israel) and Indian actor-turned-politician Pawan Kalyan’s Twitter account (@pawankalyan) is among the top at-mentioned accounts in the user bios. This is particularly surprising since this analysis only examines English-language tweets.

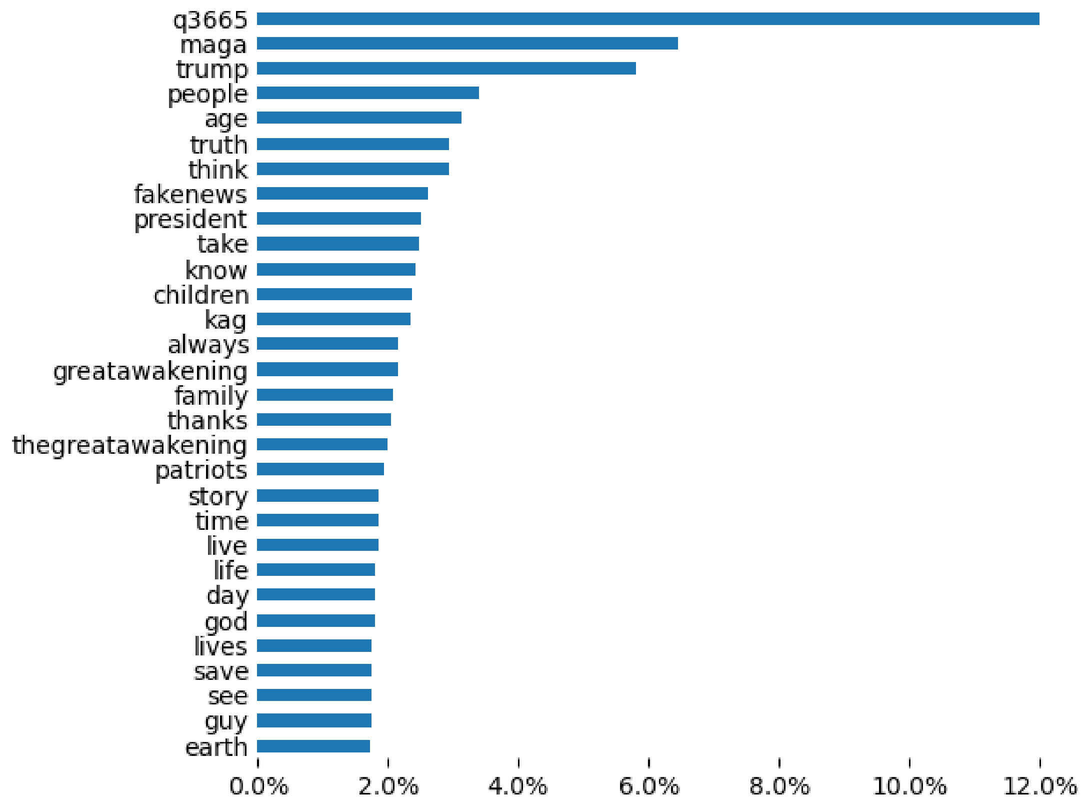
Tweet Text

Finally, we provide a preliminary investigation of the content of the nearly 85 million tweets in this dataset by identifying the most common unigrams and bigrams in the text of tweets themselves. We chose to ignore duplicate tweet text (which includes multiple retweets of the same original tweet as well as cases where one user copied the text from another tweet); in total, we found 33,068,860 tweets with unique text.[11]

As with user descriptions, we created two corpuses: one for unigrams and one for bigrams. We decided to strip the hashtag symbol ("#"), as we believe that it does not make sense to report a frequency for a word and a separate frequency for the hashtag version of that word. For the unigram corpus, we removed collection terms from the output of the most common unigrams; for the bigram corpus, we removed any bigrams where both of the unigrams that make up a given bigram are collection terms. The unigram corpus also does not contain at-mentions or emojis.

[11] This is larger than the value of the total number of quote tweets and original tweets. Presumably, for some of the retweets that we collected, we did not also collect the original tweet (for example, in cases where someone retweeted a tweet originally sent before data collection for this project); in these cases, some of the 66 million retweets contributed unique tweet text to our dataset.

Figure 12. Most prevalent unigrams in tweet text

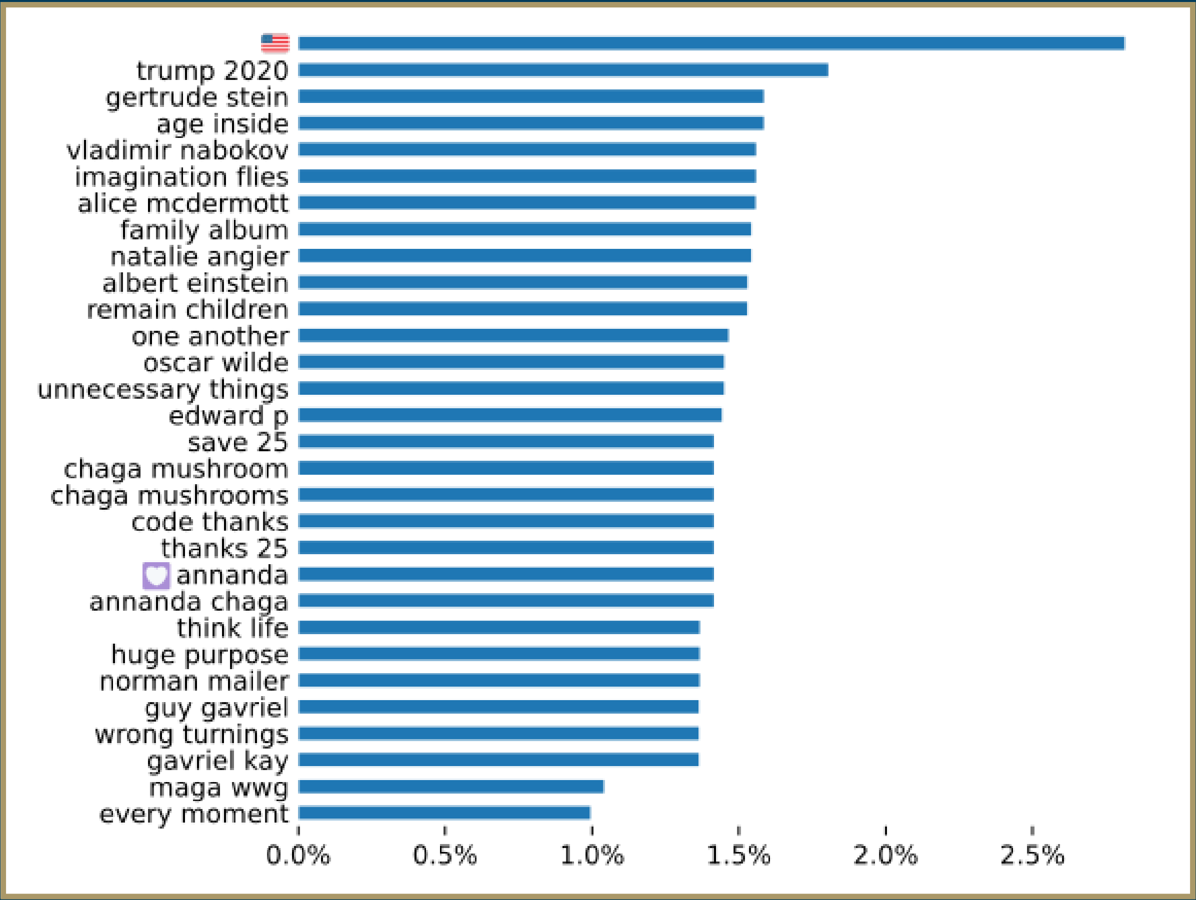


As with the top n-grams in user descriptions, the top unigrams and bigrams from tweet text suggest that support for Donald Trump is an important theme in the dataset, given the presence of "maga", 'trump", "kag", "trump 2020", and "maga wwg". Another common theme is more directly related to QAnon. "q3665" appears in a number of tweets that link to a Fox News segment on QAnon, where host Jesse Watters made favorable statements about the conspiracy theory and its believers; "great awakening" is a common slogan used by QAnon adherents. Likewise, the prevalence of the words “truth” and “children” are indicative of QAnon theorizing regarding an alleged global child sex trafficking ring.

The top bigrams from tweets mostly do not directly relate to QAnon—or even politics or contemporary events as such. A number of prominent novelists and scientists appear here: Gertrude Stein,Vladimir Nabokov, Albert Einstein, Edward P. Jones, and others. Further investigation is necessary to determine whether these names indicate that QAnon adherents are quoting from notable public thinkers and writers, or if these names are invoked as part of an effort to counter QAnon discourses.

Also of interest in the list of top bigrams is a number of references to chaga mushrooms (and Annanda Chaga Mushrooms, a Canadian company that distributes products made from these and other mushrooms). "Alternative" medicine advocates have suggested that chaga mushrooms can be used to effectively treat COVID-19.

Figure 13. Most prevalent bigrams in tweet text



Conclusion

In this report, we have provided an overview of activity related to the QAnon conspiracy theory between July 2018 and October 2020. We find approximately 85 million tweets related to QAnon sent by nearly four million unique users. Of the 85 million tweets, approximately 14% are original tweets (i.e., tweets that are neither retweets nor quote tweets), 79% are retweets, and 7% are quote tweets.

As with most social media datasets, most of the distributions described here follow a power-law distribution. For example, the overwhelming majority of tweets in the dataset were never retweeted; the overwhelming majority (approximately 72%) of tweets that were retweeted at least once were retweeted three times or fewer. On the far side of the distribution, 2% of retweeted tweets were retweeted more than 100 times.

This suggests that most users are lightly engaged, with fewer users engaging much more frequently; most tweets likely receive little attention, with fewer standout tweets receiving large amounts of attention.

Additionally, we found that just over 50% of the tweets in the dataset express support for QAnon (or belonging in the QAnon community); another 10% more broadly express support for Trump but don't explicitly mention QAnon; another 10% explicitly criticize QAnon; 2% are neutral in orientation, either providing information without including any suggestion of support or criticism of QAnon or requesting information in such a manner. We also note that a quarter of tweets are difficult to classify into one of these four categories.

The distribution of the number of tweets in the dataset over time provides a roadmap to several future lines of inquiry. First, Twitter has engaged in several waves of moderation targeting users who promote the QAnon conspiracy theory beginning in mid-2020. The nature of our dataset means that the dataset grows dramatically in the immediate aftermath of this moderation activity, as a range of users comment on the new account suspensions. At this stage, we do not know how much of the spikes after the moderation waves are expressing support for QAnon (or condemnation of Twitter's moderation activity), how much is broadly critical of QAnon (or supportive of the moderation activity), and how much is a neutral sharing of information or requesting information. We will soon build a classifier to label all tweets in our dataset according to the "Affinity" coding scheme described above. This will allow us to see whether the relative levels of support and criticism of QAnon change over time—and in particular, whether the spikes in activity after moderation activity suggest that the moderation activity is effective in reducing the amount of pro-QAnon content on Twitter.

Additionally, we note an overall increase in the volume of tweets about QAnon in 2020. It seems that the baseline level of tweets per day is higher in 2020 than in previous times, and the spikes are dramatically higher than spikes before 2020. We suspect that two things drove much of this increase: the COVID-19 pandemic and the resurgence of Black Lives Matter and anti-police brutality protests after George Floyd's death on May 25, 2020. We intend to conduct case studies of the tweets around important dates related to these two themes to understand whether and how tweets about QAnon were related to COVID-19 and Black Lives Matter.

We chose to use October 31, 2020 as the cutoff for this analysis because we believe that tweets related to QAnon in November 2020 (around the time of the presidential election) might be substantially different. Particularly as Donald Trump drove unfounded conspiracy theories about electoral fraud and the illegitimacy of the outcome, we expect to find tweets that discuss both QAnon and these false accusations.

At the same time, Twitter has engaged in more moderation activity, particularly in the aftermath of the insurrection at the U.S. Capitol on January 6, 2021, where symbols related to QAnon were frequently seen in the crowds surrounding the Capitol and in the crowds that stormed the building. Future work will investigate how this increase in salience and urgency for QAnon supporters interacted with the increased moderation activity on Twitter. This dataset—including the new data that is still coming in—will provide ample opportunities to investigate QAnon in numerous ways. This paper provides an overview of QAnon-related activity on Twitter that will serve as a foundation for future analysis, and we hope that this can be a useful baseline for other researchers examining QAnon on Twitter and other social networking platforms or online activity about other contentious or harmful topics.

Bibliography

Amarasingam, A., & Argentino, M.-A. (2020, July). T*he QAnon Conspiracy Theory: A Security Threat in the Making?* Combating TerrorisCenter at West Point. https://ctc.usma.edu/the-qanon-conspiracy-theory-a-security-threat-in-the-making/

Argentino, M.-A., & Amarasingam, A. (2020, October 29). *Q-Pilled: Conspiracy Theories, Trump, and Election Violence in the United States*. International Centre for Counter-Terrorism - The Hague. https://icct.nl/publication/q-pilled-conspiracy-theories-trump-and-election-violence/

Baer, S. K. (2018, June 17). *An Armed Man Spouting A Bizarre Right-Wing Conspiracy Theory Was Arrested After A Standoff At The Hoover Dam*. BuzzFeed News. https://www.buzzfeednews.com/article/skbaer/qanon-believer-arrested-hoover-dam

Bailey, H., Shammas, B., & Bellware, K. (2020, May 28). Chaotic scene in Minneapolis after second night of protests over death of George Floyd. *Washington Post*. https://www.washingtonpost.com/nation/2020/05/27/george-floyd-minneapolis-reaction/

Barkun, M. (1996). Religion, Militias and Oklahoma City: The Mind of Conspiratorialists. Terrorism and Political Violence, 8(1), 50–64. https://doi.org/10.1080/09546559608427332

Barkun, M. (2016). Conspiracy Theories as Stigmatized Knowledge. *Diogenes*, 1–7. https://doi.org/10.1177/0392192116669288

Bellingcat. (2021, January 8). The Journey of Ashli Babbitt. Bellingcat. https://www.bellingcat.com/news/2021/01/08/the-journey-of-ashli-babbitt/

Berger, J. M. (2018). The Alt-Right Twitter Census. VOX-Pol. https://www.voxpol.eu/download/vox-pol_publication/AltRightTwitterCensus.pdf

Collins, B., & Zadrozny, B. (2021, January 8). Twitter bans Michael Flynn, Sidney Powell in QAnon account purge. NBC News. https://www.nbcnews.com/tech/tech-news/twitter-bans-michael-flynn-sidney-powell-qanon-account-purge-n1253550

DiGrazia, J. (2017). The Social Determinants of Conspiratorial Ideation. Socius, 3, 1–9. https://doi.org/10.1177/2378023116689791

franzke, aline shakti, Bechman, A., Zimmer, M., Ess, C., & the Association of Internet Researchers. (2020). Internet research: Ethical guidelines 3.0. https://aoir.org/reports/ethics3.pdf

Hafner, K. (2006, August 23). Researchers Yearn to Use AOL Logs, but They Hesitate (Published 2006). The New York Times. http://www.nytimes.com/2006/08/23/technology/23search.html

Hemsley, J., Jackson, S., Tanupabrungsun, S., & Ceskavich, B. (2019). bitslabsyr/stack: STACKS 3.1. Zenodo. https://doi.org/10.5281/zenodo.2638848

Introne, J., Iandoli, L., DeCook, J., Yildirim, I. G., & Elzeini, S. (2017). The Collaborative Construction and Evolution of Pseudo-knowledge in Online Conversations. 1–10. https://doi.org/10.1145/3097286.3097297

Jackson, S. (2018, September 26). FlockWatch: Tracking Changes in Language in Text Datasets Over Time. VOX - Pol. https://www.voxpol.eu/flockwatch-tracking-changes-in-language-in-text-datasets-over-time/

Jackson, S. (2019). Sjacks26/FlockWatch 1.0. Zenodo. https://doi.org/10.5281/zenodo.2639607

Kaplan, A. (2020, November 9). Here are the QAnon supporters running for Congress in 2020. Media Matters for America. https://www.mediamatters.org/qanon-conspiracy-theory/here-are-qanon-supporters-running-congress-2020

Manjoo, F. (2021, January 6). With One Presidential Phone Call, QAnon Shows. New York Times. https://www.nytimes.com/2021/01/06/opinion/trump-qanon-georgia-call.html

Marcin, T. (2020, June 5). K-pop fans are flooding QAnon hashtags with memes and fancams. Mashable. https://mashable.com/article/kpop-flood-qanon-hashtags/

Martineau, P. (2017, December 19). The Storm Is the New Pizzagate—Only Worse. New York Magazine. https://nymag.com/intelligencer/2017/12/qanon-4chan-the-storm-conspiracy-explained.html

Notopoulos, K., & Lytvynenko, J. (2021, January 6). QAnon Believers Were Part Of Mob That Stormed The Capitol. Buzzfeed News. https://www.buzzfeednews.com/article/katienotopoulos/qanon-believers-stormed-the-capitol

Ortutay, B. (2020, October 15). YouTube follows Twitter and Facebook with QAnon Crackdown. AP News. https://apnews.com/article/youtube-qanon-conspiracy-theories-ef03a889e68239de6692ce42666d97d8

Pitcavage, M. (2001). Camouflage and Conspiracy: The Militia Movement From Ruby Ridge to Y2K. American Behavioral Scientist, 44(6), 957–981. https://doi.org/10.1177/00027640121956610

QAnon: Twitter bans accounts linked to conspiracy theory. (2020, July 22). BBC News. https://www.bbc.com/news/world-us-canada-53495316

Roose, K. (2020, October 19). What Is QAnon, the Viral Pro-Trump Conspiracy Theory? New York Times. https://www.nytimes.com/article/what-is-qanon.html

Rose, J. (2020, December 30). Even If It's "Bonkers," Poll Finds Many Believe QAnon And Other Conspiracy Theories. NPR. https://www.npr.org/2020/12/30/951095644/even-if-its-bonkers-poll-finds-many-believe-qanon-and-other-conspiracy-theories

Sardarizadeh, S., & Goodman, J. (2020, August 7). Facebook removes QAnon conspiracy group with 200,000 members. BBC News. https://www.bbc.com/news/technology-53692545

Scott, M. (2020, November 6). Europe's QAnon followers embrace US election conspiracy theories. Politico. https://www.politico.eu/article/europe-qanon-us-election/

Sen, A., & Zadrozny, B. (2020, August 10). QAnon groups have millions of members on Facebook, documents show. Nbc News. https://www.nbcnews.com/tech/tech-news/qanon-groups-have-millions-members-facebook-documents-show-n1236317

Smith, A. (2020, December 29). VOAT SHUTDOWN: QANON AND ANTI-VACCINATION REDDIT CLONE SHUT DOWN ON CHRISTMAS DAY. Independent. https://www.independent.co.uk/life-style/gadgets-and-tech/voat-shutdown-christmas-qanon-reddit-b1779890.html

Sommer, W. (2020, April 30). A QAnon Devotee Live-Streamed Her Trip to N.Y. to "Take Out" Joe Biden. Daily Beast. https://www.thedailybeast.com/a-qanon-devotee-live-streamed-her-trip-to-ny-to-take-out-joe-biden

Sunstein, C. R., & Vermeule, A. (2009). Conspiracy Theories: Causes and Cures. Journal of Political Philosophy, 17(2), 202–227. https://doi.org/10.1111/j.1467-9760.2008.00325.x

Swami, V., Chamorro-Premuzic, T., & Furnham, A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. Applied Cognitive Psychology, 24(6), 749–761. https://doi.org/10.1002/acp.1583

Swami, V., Coles, R., Stieger, S., Pietschnig, J., Furnham, A., Rehim, S., & Voracek, M. (2011). Conspiracist ideation in Britain and Austria: Evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories: Conspiracist ideation. British Journal of Psychology, 102(3), 443–463. https://doi.org/10.1111/j.2044-8295.2010.02004.x

Timberg, C. (2020, October 3). Twitter banished the worst QAnon accounts. But more than 93,000 remain on the site, research shows. Washington Post. https://www.washingtonpost.com/technology/2020/10/03/twitter-banished-worst-qanon-accounts-more-than-93000-remain-site-research-shows/

Timberg, C., & Stanley-Becker, I. (2020, October 28). QAnon learns to survive—And even thrive—After Silicon Valley's crackdown. Washington Post. https://www.washingtonpost.com/technology/2020/10/28/qanon-crackdown-election/

Trump, D. (2020, March 13). Proclamation on Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID-19) Outbreak. The White House. https://www.whitehouse.gov/presidential-actions/proclamation-declaring-national-emergency-concerning-novel-coronavirus-disease-covid-19-outbreak/

Twitter Safety. (2020, July 22). We've been clear that we will take strong enforcement action on behavior that has the potential to lead to offline harm. In line with this approach, this week we are taking further action on so-called 'QAnon' activity across the service. [Tweet]. @TwitterSafety. https://twitter.com/TwitterSafety/status/1285726277719199746

van Prooijen, J.-W., Krouwel, A. P. M., & Pollet, T. V. (2015). Political Extremism Predicts Belief in Conspiracy Theories. Social Psychological and Personality Science, 6(5), 570–578. https://doi.org/10.1177/1948550614567356

Watkins, A. (2019, July 21). He Wasn't Seeking to Kill a Mob Boss. He Was Trying to Help Trump, His Lawyer Says. New York Times. https://www.nytimes.com/2019/07/21/nyregion/gambino-shooting-anthony-comello-frank-cali.html

Winter, J. (2019, August 1). Exclusive: FBI document warns conspiracy theories are a new domestic terrorism threat. Yahoo!News. https://news.yahoo.com/fbi-documents-conspiracy-theories-terrorism-160000507.html

Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. Ethics and Information Technology, 12(4), 313–325. http://dx.doi.org.libproxy.albany.edu/10.1007/s10676-010-9227-5

Zimmerman, M. (2020, November 29). QAnon's Rise in Japan Shows Conspiracy Theory's Global Spread. Bloomberg. https://www.bloomberg.com/news/articles/2020-11-29/qanon-s-rise-in-japan-shows-conspiracy-theory-s-global-spread

Appendix A: Collection term information

This appendix contains information about the collection terms used for this project. Each of the items in the "Collection term" column of the table that follows was provided to STACKS as a track term to use in the Twitter API call. Functionally, this is equivalent to having a single term that consists of all of the terms listed below separated by a space.

Additional collection terms were added over time as described in the main text. The "Start date" column indicates the date on which a given collection term was added to the collection criteria.

All but two of these collection terms are still actively used. "we are q" and "q sent" were both removed as active collection terms on November 15, 2019. When they were originally added, the author responsible for data collection (Jackson) did not realize that spaces in the API call function as boolean AND operators rather than being treated as a single phrase.[1] Upon realizing this, these two terms were removed to reduce the number of tweets collected that have nothing to do with QAnon.

qanon	7/9/18
pizzagate	7/12/18
whoisq	7/18/18
sethrich	7/18/18
qisback	7/27/18
wwg	7/27/18
1wga	7/27/18
pedowood	7/27/18

[1] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/guides/basic-stream-parameters#track

qarmy	7/27/18
pedogate	7/27/18
wwg1wga	7/27/18
qisworldwide	8/8/18
qisreal	8/8/18
weareq	8/8/18
pizzagateisreal	8/8/18
qenlightened	8/18/18
qstormers	8/18/18
whoisqanon	8/18/18
qalert	8/18/18
qanondeltaforce	8/29/18
italianqarmy	8/29/18
pedogateisreal	8/29/18
qanonpatriots	8/29/18
qanonarmy	9/4/18

qposts	9/4/18
qanonposts	9/4/18
qarmytrain	9/11/18
qproofs	9/11/18
ww1wga	9/11/18
qsentme	10/17/18
newq	10/17/18
qclearance	10/17/18
qteam	10/17/18
qillbox	10/17/18
q3121anon	10/17/18
we are q*	10/17/18
qdrop	10/17/18
oppedohuntersunited	10/17/18
pedohuntersunited	10/17/18
oppedhunt	10/17/18
qbreakingnewsqyah	10/29/18

qanonwhitehat	10/29/18
qwarriors	11/5/18
q sent*	11/8/18
wherewegoonewegoall	11/8/18
qsentus	11/8/18
qmovement	11/8/18
qanondna	11/14/18
officialqanon	11/30/18
realqanon	11/30/18
qofficial	11/30/18
officialq	11/30/18
qpostsnow	11/30/18
qincidences	12/17/18
1wgaworldwide	1/1/19
wwg1wgaworldwide	1/1/19
qprotege	1/30/19
anonsknew	3/11/19

frazzledrip	3/20/19
qmemes	5/1/19
1wga_worldwide	5/1/19
qanon2019	5/16/19
wwg1wga_worldwide	5/16/19
wwg1gwa	5/28/19
wwg1wwga	5/28/19
pedogatenews	5/28/19
pedovores	5/28/19
cannibalsofpedowood	5/28/19
qanonisreal	5/31/19
1wgaww	6/13/19
wwgiwga	7/3/19
qarmykorea	7/3/19
qarmyjapan	7/3/19
qcommunity	7/17/19
qnavy	7/17/19

followq	7/28/19
opqanon	7/28/19
qbaby	7/28/19
qroof	7/28/19
chantforq	7/28/19
qpatriot	8/14/19
qarmycorea	8/14/19
qdecodes	8/21/19
qarmycanada	8/21/19
qarmychina	8/21/19
qarmyrussia	8/21/19
qrevolution	8/26/19
worldwideqanon	8/26/19
werqww	10/3/19
qworldwide	10/3/19
qnn	10/3/19

qmapjapan	10/3/19
qanons	10/3/19
opfuhq	10/3/19
qnuts	10/3/19
antiqrallytweetstorm	10/3/19
forcetheq	10/10/19
whatisq	10/10/19
qanonmategate	10/10/19
qarmyjapanflynn	10/23/19
qfamily	11/8/19
qunited	11/8/19
qtaskfoce	11/8/19
qstrong	11/8/19
aanon	11/8/19
aanon_qsarmy	11/8/19
forcingtheq	11/8/19
qsarmy	11/15/19

2q2q	1/21/20
qarmytaskforce	3/4/20
qanondeutsch	4/1/20
qanongermany	4/1/20
qanonfrance	4/1/20
qanonworldwide	4/8/20
qanonnederland	4/8/20
adrenochromewithdrawal	4/8/20
qanonuk	4/8/20
qanonitalia	4/8/20
qanondeutsche	4/8/20
adrenachrome	4/8/20
adrenochromeharvesting	4/8/20
adrenochrome	4/8/20
qanonaustrailia	4/15/20
qanoncanada	4/22/20
qanon_uk	5/5/20

qanons_uk	5/5/20
qanonsuk	5/5/20
qproof	5/5/20
quantumq	5/15/20
ukqanon	5/15/20
qarmyworldwide	5/15/20
qnitedstatesofamerica	5/15/20
thankq	5/15/20
qanonfollowthewhiterabbit	5/15/20
qanonfinland	5/21/20
qarmydach	5/21/20
qanondach	5/21/20
qarmygermany	5/21/20
qanonart	5/21/20
qanonspain	5/29/20
qanonmexico	5/29/20
qanonitaly	5/29/20

qanonjapan	5/29/20
qanonisrael	5/29/20
qanonportugal	5/29/20
qanoniran	5/29/20
qanonaustralia	5/29/20
qanonindia	5/29/20
qanonhk	5/29/20
qanonbrasil	5/29/20
qanonpoland	5/29/20
qanonserbia	5/29/20
tg1tga	5/29/20
qanon_worldwide	5/29/20
qanonsworldwide	5/29/20
thinkqanon	6/10/20
pedowoodisreal	6/10/20
qanongermanydach	6/10/20
qanonturkey	6/25/20

qanonkr	6/25/20
qanonsouthafrica	6/25/20
qanonjp	6/25/20
qanaonmorldwide	6/25/20
qanonus	6/25/20
qanonusa	6/25/20
qmapsearch	6/25/20
qanonmap	6/25/20
pizzagateemails	6/25/20
qanon2020	8/18/20
qanoncult	8/27/20
qidiots	9/22/20
qarmyweirdos	9/22/20
qmorons	9/22/20
saveourchildrenfromqanon	9/22/20
qanonsense	9/22/20
qkluxklan	9/22/20
qanondon	10/21/20
qanonsspain	10/21/20

Appendix B: Collection term as account handles

A number of collection terms used to build this dataset correspond to the handles (also known as screen names) used by accounts. In mid-September 2020, we identified which of the active collection terms were currently used as account handles.

We wish to emphasize that many of the accounts listed in this appendix do not tweet about QAnon at all. Even for those accounts where tweets that at-mention the account are deemed relevant, we make no assertion that the account itself is relevant for QAnon. This is apparent for the accounts in this list that have not even been active in the time since QAnon emerged. Instead, this decision is based on our determination of how people who tweet at that handle use the handle.

Because non-verified users can change their account handle at any time, it is possible that some of these terms are no longer used as account handles. It is also possible that some terms that don't appear on this list have been used as account handles during some unknown period of time. And it is also possible that some of the terms in this list have actually been used as account handles at different points in time.

Because we use collection terms as "track" parameters for our stream API call, we collect any tweet that at-mentions any account whose handle corresponds to a collection term. As described in the manuscript, we treat any at-mention of some of these terms as sufficient to know that a tweet is somehow related to QAnon. For those terms-as-account-handles that aren't always used to discuss QAnon (for example, WWG, which is the handle for a verified account that reports on the gaming industry), we subject the tweet to the relevance coding procedure outlined in the manuscript.

We note that some of these terms correspond to accounts that themselves have no relation to QAnon (such as WWG). Indeed, some of these accounts haven't tweeted in years, long before the QAnon conspiracy theory emerged. Despite this, we still determined that at-mentions of some accounts that have no relation to QAnon are still sufficient to indicate that a tweet is about QAnon. For example, @qnavy is an account that was created in 2010 and last tweeted in 2012. We treat any tweet that at-mentions qnavy as relevant for our dataset because we think of qnavy as analogous to qarmy, a widely used term to refer to Q believers.

As of January 2021, many of these accounts have been suspended or have changed their handle.

Account handles that are collection terms	**@mention of this account sufficient to indicate relevance?**
1wga	N
2q2q	Y
AAnon	Y
adrenachrome	Y
adrenochrome	Y
AnonsKnew	Y
FollowQ	N
ForceTheQ	Y
FRAZZLEDRIP	Y
newq	N
OfficialQ	N
OfficialQAnon	Y
opqanon	Y
pedogate	Y
pedogateisreal	Y
pedogatenews	Y
pedovores	Y
pedowood	Y
pizzagate	Y
pizzagateisreal	Y
q3121anon	Y
qalert	Y
qanon	Y
Qanon_uk	Y

qanon2019	Y
qanon2020	Y
qanonarmy	Y
QanonArt	Y
QAnonAustralia	Y
QAnonCanada	Y
qanoncult	Y
QAnonDeutsch	Y
qanonfrance	Y
QAnonGermany	Y
QanonIndia	Y
QANONIRAN	Y
qanonisrael	Y
qanonisreal	Y
qanonitalia	Y
qanonitaly	Y
qanonmexico	Y
QanonPatriots	Y
qanonposts	Y
qanons	Y
qanonspain	Y
QAnonsWorldwide	Y
QanonTurkey	Y
QanonUk	Y
qanonus	Y
qanonusa	Y

qanonwhitehat	Y
qanonworldwide	Y
qarmy	Y
qarmycorea	Y
QArmyGermany	Y
qarmyjapan	Y
QArmyJapanFlynn	Y
qarmykorea	Y
qbaby	N
qclearance	N
qcommunity	N
qdecodes	Y
qdrop	Y
Qenlightened	Y
QFamily	N
qillbox	Y
qisback	N
Qisreal	Y
qisworldwide	Y
qmapjapan	Y
QMemes	N
Qmovement	Y
qnavy	Y
qnn	Y
Qnuts	N
QOfficial	N

QOfficial	N
qpatriot	Y
QPosts	Y
qpostsnow	Y
qProof	Y
qproofs	Y
qprotege	N
qrevolution	N
qRooF	N
QsArmy	Y
qsentme	Y
qsentus	Y
QStrong	N
Qteam	Y
QuantumQ	N
QUnited	N
QWarriors	N
Qworldwide	N
realqanon	Y
sethrich	Y
ThankQ	Y
ThinkQanon	Y
ukqanon	Y
weareq	Y
werqww	N
whatisq	N

whoisQ	Y
whoisqanon	Y
worldwideqanon	Y
ww1wga	Y
WWG	N
WWG1WGA	Y
Wwg1Wwga	Y
wwgiwga	Y
* Many of the accounts listed here are not active in the QAnon space. Even where we determine that an @mention of an account is sufficient to indicate that a tweet is relevant for our dataset, that determination is based on the content of the tweets that @mention the account rather than the activity of the account itself.	