

The Manufacture of Political Echo Chambers by Follow Train Abuse on Twitter

Christopher Torres-Lugo*

Observatory on Social Media
Indiana University, Bloomington, USA

Kai-Cheng Yang

Observatory on Social Media
Indiana University, Bloomington, USA

Filippo Menczer

Observatory on Social Media
Indiana University, Bloomington, USA

ABSTRACT

A growing body of evidence points to critical vulnerabilities of social media, such as the emergence of partisan echo chambers and the viral spread of misinformation. We show that these vulnerabilities are amplified by abusive behaviors associated with so-called “follow trains” on Twitter, in which long lists of like-minded accounts are mentioned for others to follow. This leads to the formation of highly dense and hierarchical echo chambers. We present the first systematic analysis of U.S. political train networks, which involve many thousands of hyper-partisan accounts. These accounts engage in various suspicious behaviors, including some that violate platform policies: we find evidence of inauthentic automated accounts, artificial inflation of friends and followers, and abnormal content deletion. The networks are also responsible for amplifying toxic content from low-credibility and conspiratorial sources. Platforms may be reluctant to curb this kind of abuse for fear of being accused of political bias. As a result, the political echo chambers manufactured by follow trains grow denser and train accounts accumulate influence; even political leaders occasionally engage with them.

CCS CONCEPTS

• **Networks** → **Social media networks**; • **Human-centered computing** → *Social network analysis*.

KEYWORDS

social media integrity, echo chamber, abusive behaviors

ACM Reference Format:

Christopher Torres-Lugo, Kai-Cheng Yang, and Filippo Menczer. 2018. The Manufacture of Political Echo Chambers by Follow Train Abuse on Twitter. In *xxxxx*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

In the past decade, online social media have become an important platform for organizing and participating in social movements regarding various public issues like economic inequality [9, 22], human rights [40], and especially political elections [19]. Social

media features like anonymity, high efficiency, broad coverage, and decentralization greatly facilitate organization efforts [30, 38, 42]. Unfortunately, the same features invite problems like inauthentic behaviors [18, 33], echo chambers [35], and the wide spread of toxic information [2, 23, 36] that pose new challenges to preserve the integrity of the information ecosystem. As the online world becomes closely entangled with the real world, studies of bad actors and their manipulation of online discourse are crucial.

“Follow trains” are a way for social media users to suggest other accounts to their followers. In recent years, we have observed the widespread abuse of follow trains for political manipulation on Twitter and other social media platforms. The characteristic behavior of political follow trains is the publishing of spam-like follow train tweets (“train cars”) that typically contain a list of mentions (accounts), a media object, and possibly a few words. The screenshots in Figure 1 show examples of two follow trains. The organizers (“train conductors”) post the train tweets so that their followers can follow the accounts being mentioned (“train riders”).

The essential goal of conducting follow trains is for the riders to efficiently gain more followers, which is a direct violation of Twitter’s platform manipulation and spam policy.¹ Tweets from train accounts are used to constantly seek attention from government officials and politicians by means of mentioning, retweeting, and replying. Occasionally they get amplified by influential users and reach a much broader audience. For example, President Trump has retweeted train accounts in the past.

Since online movements may have a huge impact on the real world, their abuse must be understood and prevented. Toward this goal, this paper presents, to the best of our knowledge, the first systematic analysis of political follow trains. We focus on the Twitter platform in the context of U.S. politics. We propose an analytical framework to characterize train accounts from multiple perspectives, with a focus on their social networks and questionable behaviors. We build datasets of train accounts, collect their tweets, and contrast their behaviors against those of accounts in baseline datasets to provide meaningful contexts for our analysis.

We find that political follow trains are highly coordinated, persistent, polarized, and fully focused on U.S. politics. They form two distinct, hierarchical, and dense communities that amplify conservative and liberal narratives, respectively. Train accounts are more active than a sample of political Twitter users in terms of posting tweets and establishing social ties. In addition to generating and amplifying a large amount of spam-like train tweets, some of the accounts also abuse the platform through automation and abnormal tweet deletions. By analyzing their tweets, we find that train accounts are actively sharing toxic information, such as

*Corresponding author; email: torresch@indiana.edu

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

xxxx, xxxx, 2020, xxx, xx

© 2018 Association for Computing Machinery.

ACM ISBN XXXX-X-XXXX-XXXX-X/XX/XX...\$15.00

<https://doi.org/10.1145/1122445.1122456>

¹<https://help.twitter.com/en/rules-and-policies/platform-manipulation>

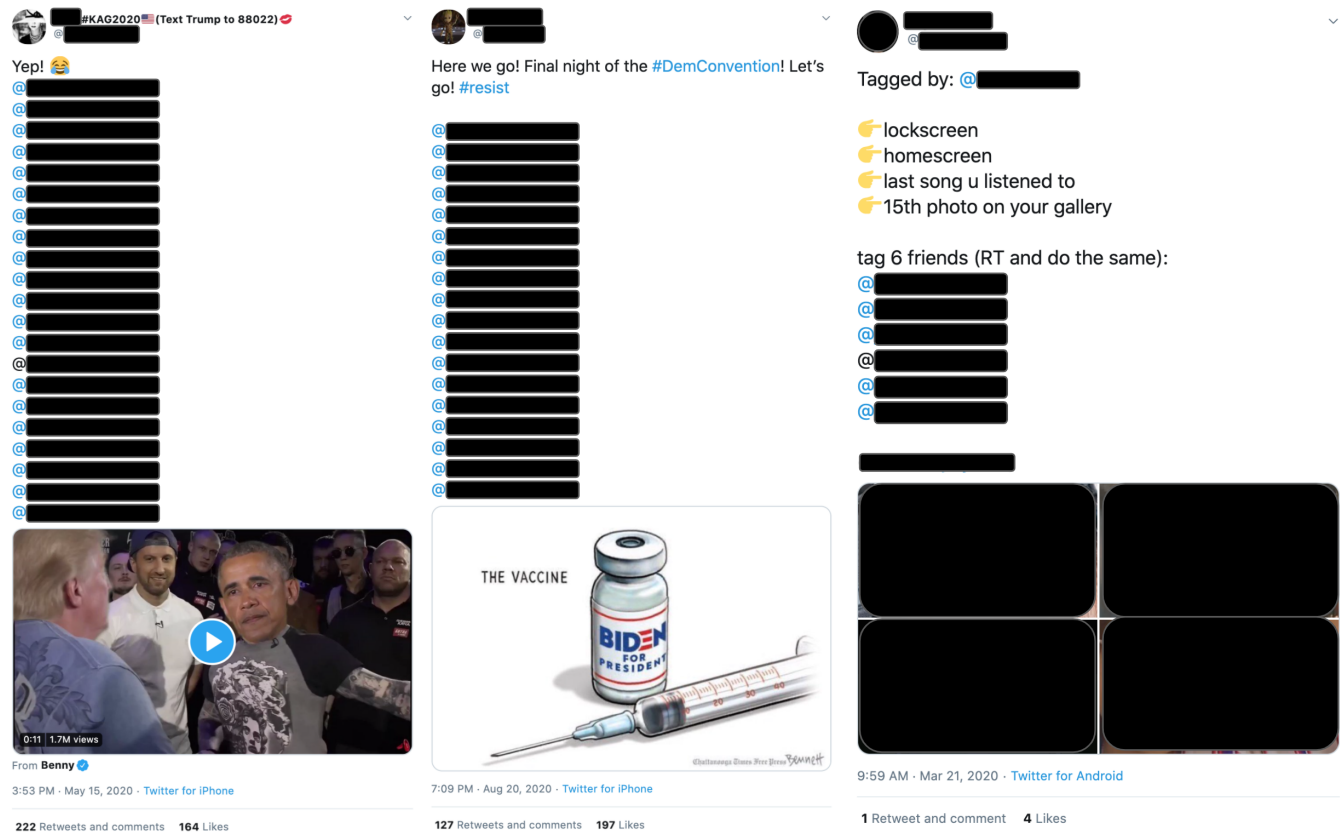


Figure 1: Screenshots of exemplar tweets from (left) conservative follow trains, (center) liberal follow trains, and (right) the tagging game. Some user information is redacted to protect privacy.

low-credibility news and conspiracy theories. Despite their violations of platform policies, political train accounts are subject to few disciplinary actions, like suspension.

2 BACKGROUND

Social media influence stems from having high visibility on a platform. Despite early analysis showing that influence is not exclusively determined by the number of followers [8], it is generally assumed that users need to have many followers to increase their visibility. Some unscrupulous actors therefore resort to follower growth hacking. A well-studied growth hack is to purchase “fake followers,” which often consist of inauthentic or compromised accounts [1, 10]. There are also reports of organized exchanges to turn unpaid customers and volunteers into fake followers [29, 41]. Conducting follow trains is another follower growth hack. However, there is a fundamental difference: as part of influence campaigns, political follow trains are coordinated actions with the ultimate goal of building a well-connected community with maximized real-world impact.

In addition to violating Twitter’s policies, political follow trains may also produce undesirable outcomes. For example, polarized and segregated echo chambers are commonly observed on social media [21, 25, 28], possibly leading to radicalization [7, 44]. Behind

the curtain is the interplay between socio-cognitive biases, like the tendency to establish belief-consistent social ties [14, 24], and social media mechanisms, like friend recommendations and the ease of (un)following [35]. Blindly following train rider recommendations can easily accelerate the formation of polarized echo chambers.

Political follow trains also make the online community more vulnerable to inauthentic actors, who will be blindly followed when recommended. Well-known types of inauthentic accounts include trolls [48] and malicious social bots [18], which have been actively involved in online discussions of elections across various countries [4, 5, 12, 17, 39, 48]. Political trains provide an easy mechanism for an entity to control automated accounts programmed to follow accounts mentioned by a conductor. Even train conductors can be automated. Recently, more attention has been drawn toward a new type of inauthentic actors that act in a coordinated fashion to increase influence and evade detection [32, 33, 37]. Follow trains facilitate the formation of coordinated inauthentic networks.

Another problem regarding follow trains is the spread of toxic information, such as conspiracy theories and misinformation. Concerns about “fake news” on social media have been growing since the 2016 U.S. presidential election [6, 23, 27, 43]. During the 2020 COVID-19 pandemic, misinformation related to the outbreak, also

known as the “infodemic,” has also spread virally [49]. Recent studies show that polarized echo chambers are associated with the diffusion of misinformation [13] and inauthentic actors like malicious bots are responsible for spreading low-credibility information related to politics and the pandemic [36, 45]; this suggests that political follow trains may also exacerbate the misinformation problem — a conjecture we explore in this paper. Due to the potential real-world consequences of health misinformation and conspiracy theories, it’s important to thoroughly investigate the role of political follow train in such abuse.

3 DATA COLLECTION

3.1 Network Analysis

First, we focus on the social (mention and follow) networks of train accounts. Based on our observations, we label a tweet as a *follow train tweet* if it mentions nine or more other accounts, contains media, and has been retweeted at least 40 times. Any account having at least one original follow train tweet is considered a *train conductor*. Accounts mentioned in follow train tweets are labeled as *train riders*. The conductors could also be riders, but the conductor label takes precedence in our dataset. Such an operationalization allows us to distinguish train conductors and riders at scale.

We utilize snowball sampling to crawl the mention network and create a dataset of train conductors and riders. Starting from a train conductor, we query the Twitter’s search API to retrieve all of its original follow train tweets (excluding replies, retweets, and quotes) and extract all the mentioned accounts (riders). The procedure is repeated recursively on the riders until no new accounts emerge. The data collection took place between February 16 and 26, 2020. The resulting mention network, denoted as *train-net*, contains 274 conductor nodes, 11,118 rider nodes, and 29,107 edges. Note that the categorization here only reflects the behaviors of the accounts in the data collection period; as mentioned earlier, riders could act as conductors at a different time.

To gauge the structure of the mention network, we need a suitable baseline. We use a similar method to collect a group of accounts that share similar behaviors but in a non-political context. We take advantage of an online game played by many Twitter users in the early days of the 2020 COVID-19 lock-down. In the game, each tagged user is asked to post a screenshot of their phone and mention six friends to continue the game with the same instruction. The screenshot of an exemplar tagging tweet can be seen in Figure 1. The data collection took place between March 21 and 27, 2020. We searched Twitter for the phrases “*tagged by*,” “*lockscreen*,” “*home-screen*,” and “*last song u listened to*” to find potential participants. Next, we built a mention network among the matched accounts and only kept the largest connected component, with 5,567 nodes and 7,189 edges. We denote this dataset as *tagging-net*.

3.2 Behavioral Analysis

We are also interested in the profiles and behaviors of the accounts, especially the conductors. To this end, we conducted an exhaustive manual annotation of the accounts in *train-net*. The great majority are hyper-partisan. We label them as conservative or liberal accounts based on the clues in their profiles and timelines.

Table 1: Composition of accounts in terms of conductor/rider roles and political alignment from train-recent and train-decahose.

	Conductor	Rider	Total
Conservative	458	1,901	2,308
Liberal	103	362	473
Total	561	2,263	2,824

However, the train-net dataset has some shortcomings for this analysis. First, the number of conductors is relatively small. Second, the dataset includes a minority of accounts whose behavior is not of interest for various reasons: 9% are verified accounts, 3% are suspended or deleted, and 1% are non-political or inactive. We therefore create some new datasets.

To find more train conductors, we leverage the observation that conductor accounts retweet follow train tweets from other conductors. Therefore we collect the 200 most recent tweets of the conductors as of September 22–23, 2020, and we search their retweets for train tweets. The union set of the conductors is then joined with a random sample of 20% of the riders in *train-net*. The minority groups mentioned above are excluded from this sample of riders. If an account in this sample is already present among the newly added conductors, the conductor label takes precedence. The composition of the resulting set of accounts in terms of conductor/rider roles and political alignment is shown in Table 1. Finally, to analyze profiles and behaviors of these accounts, we collect their 200 most recent tweets and denote this dataset of tweets as *train-recent*. We also sample 24,274,474 of their tweets posted between January 1, 2018 and June 30, 2020 from the OSoMe historical archive [11] that offers a 10% sample of the public tweet stream. We denote this dataset as *train-decahose*.

While the train conductors and riders considered are all political, not all political accounts are involved in follow trains. We therefore need a suitable baseline to show whether the characteristics of the train accounts are typical of political users. To build a representative sample of political accounts on Twitter, we first generate a list of 100 hashtags that set apart the conservative and liberal accounts in *train-net* (see details in Figure 5). We then scan the historical archive and find all the accounts that tweeted on June 21, 2020 and that had any of the selected hashtags in their descriptions. Among the matched political accounts, we sample 2,107 conservative and 392 liberal ones as our political baseline, to match the proportion in Table 1. As we did for the train accounts, we collect 200 most recent tweets by the political baseline accounts in the *political-recent* dataset, and their historical tweets from January 1, 2018 to June 30, 2020 in the *political-decahose* dataset.

4 ANATOMY OF FOLLOW TRAIN NETWORKS

4.1 Mention Networks

Let us first analyze the *train-net* and *tagging-net* mention networks, in which each node represents an account and an edge pointing from node A to node B means that account A has mentioned B. Edge weights represent the numbers of mentions.

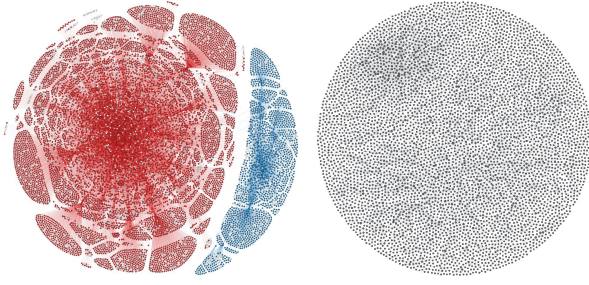


Figure 2: Visualizations of the mention networks among accounts in the (left) train-net dataset and (right) tagging-net baseline. We used Gephi with the Fruchterman-Reingold layout and the following parameters: area = 10,000, gravity = 10. In the train-net network, conservative and liberal accounts are colored in red and blue, respectively; accounts excluded from the behavioral analysis are colored in white.

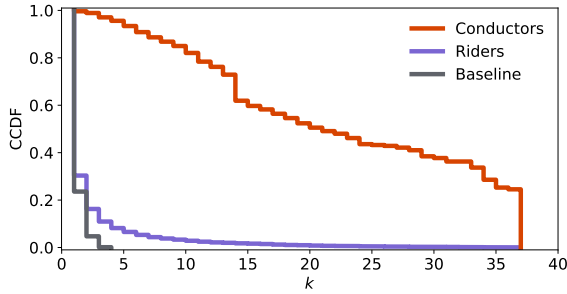


Figure 3: Complementary cumulative distribution of the k -shell indexes for accounts in the train-net network and tagging-net baseline. The categorization of conductors and riders in train-net reflects the account statuses during the data collection period. Mann-Whitney U tests show that the distributions are significantly different ($p < 0.01$) with large effect sizes.

Figure 2 visualizes the networks. The train-net network contains two densely connected communities. Our manual annotation reveals that the larger community mainly consists of conservative accounts while the smaller one mainly consists of liberal accounts. Note that our snowball-sampling crawler started with a conservative account; it was therefore surprising to find liberal accounts in the network. A close examination of the rare edges between the two communities suggests that most of them are likely unintended.

Comparing the two mention networks in Figure 2, we find that each community in the train-net network has a densely connected core and many peripheral nodes, while the tagging-net network displays a more homogeneous structure. To confirm our observation, we calculate the k -shell index [26], which measures node centrality and influence, of the nodes in both networks and plot the distributions in Figure 3. For the tagging-net baseline

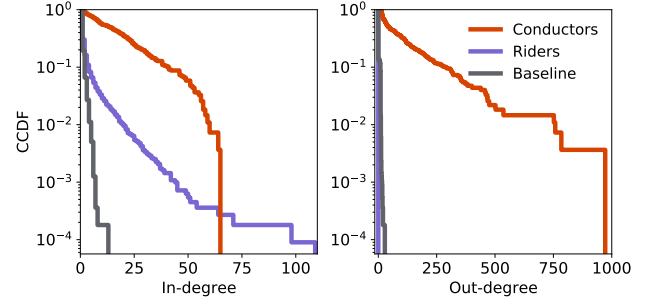


Figure 4: Complementary cumulative distributions of (left) in-degree and (right) out-degree for accounts in the train-net mention network and tagging-net baseline. The categorization of conductors and riders in train-net reflects the account statuses during the data collection period. Mann-Whitney U tests show that the distributions are significantly different ($p < 0.01$) with large effect sizes.

network, the k -shell indexes of all nodes are smaller than five. The train-net network, on the contrary, has a deeply hierarchical structure with a very dense core. Conductors tend to have higher k values than riders, showing that they are situated in the core of the network.

To further characterize the mention network, we plot the in-degree and out-degree distributions of the nodes in Figure 4. High in-degree indicates that an account is mentioned by many others and high out-degree means the account is mentioning many others. We find that the train-net accounts, and especially the riders, tend to have much higher in-degree than the tagging-net baseline accounts. Riders in train-net and accounts in tagging-net both have relatively low out-degree values while conductors have staggeringly high out-degree values, as expected.

We also examine the clustering structure of the networks. While the train-net and tagging-net networks have similar densities (2.2×10^{-4} vs. 2.3×10^{-4}), the former has a much higher average clustering coefficient (0.13 vs. 0.03) [31]. This suggests that some accounts act as both conductors and riders.

The above analyses show that the mention network of political trains is heavily clustered and hierarchically organized around a dense core of highly active conductors. The baseline mention network, on the other hand, has a more homogeneous and decentralized structure that emerges from a lack of hierarchy.

4.2 Follow Networks

We are also interested in the follow networks induced by political trains. Since querying the following relationship between each pair of accounts is not practical due to Twitter’s API rate limit, we adopt a sampling strategy. We split the train-net accounts into the conservative group and the liberal group as they form two distinct communities in the mention network (see Figure 2). Together with the accounts in tagging-net, we have three different groups of accounts. We sample 5,000 account pairs within each group and use Twitter’s friendship API to check whether the accounts in each pair are following each other. An account pair is considered to have

Table 2: Percentages of sampled account pairs with follow edges within and across groups. “Conservative” and “Liberal” refer to the conservative and liberal accounts in the train-net dataset; “tagging” refers to the accounts in the tagging-net dataset.

	Conservative	Liberal	Tagging
Conservative	30.0%	0.1%	0.0%
Liberal		51.1%	0.0%
Tagging			0.9%

a follow edge as long as either account is following the other. The same procedure is also applied to pairs of accounts across groups.

We report the percentages of account pairs with follow edges within and across groups in Table 2. Accounts within the same conservative and liberal groups are much more likely to follow each other than accounts in the baseline group or across groups. The lack of follow edges between the tagging-net group and the train groups is expected because they are collected from different contexts. But it is somewhat surprising to see such few connections between the conservative and liberal groups, considering they are all highly political accounts. This result shows that the follow networks induced by political trains form echo chambers that are not just densely connected, but also highly segregated.

5 PROFILE AND BEHAVIORAL CHARACTERIZATION

5.1 Account Profiles

We start with the account descriptions to provide an intuitive characterization of the political follow train identities. Twitter allows each user to compose a short description (or bio) with up to 160 characters. In addition to plain text, descriptions can include hashtags. Through manual inspection, we find that the train accounts commonly include political hashtags and identity keywords in their descriptions.

Let us first focus on hashtags. The word shift graph [20] in Figure 5 provides an informative comparison between the hashtag usage of the conservative and liberal accounts in the train-net dataset. The graph ranks the hashtags by their contribution to the frequency difference between the two groups. We can see that campaign slogan hashtags dominate both sides. For the conservative accounts, we also note the frequent use of hashtags referring to the QAnon conspiracy theory (including #wwg1wga, short for “where we go one, we go all”). The liberal accounts, on the other hand, tend to be associated with social movements like #MeToo and Black Lives Matter (#BLM).

We also analyze the self-identifications of the train-recent and political-recent accounts. The descriptions contain many references to topics such as law, gun-rights, family, climate change, etc. A previous study suggests that some inauthentic actors adopt this strategy to target specific audiences by using topics that mostly appeal to either right- or left-leaning voters [15]. We manually selected nine common themes and calculated the percentages of accounts in the train-recent and political-recent datasets whose descriptions contain the corresponding keywords (Table 3).

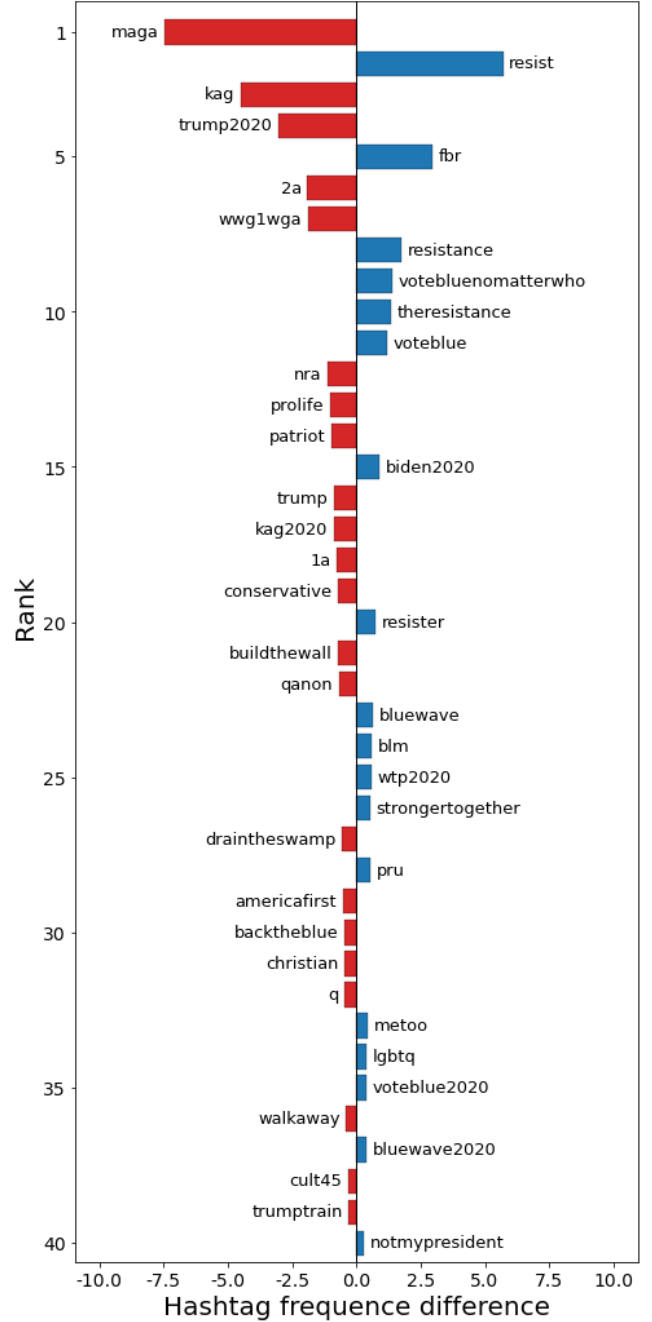


Figure 5: Forty hashtags that are most discriminative between the descriptions of conservative and liberal accounts in train-net, ranked by their contribution. The bars indicate the differences of hashtag frequency — a negative (positive) value means the corresponding hashtag is more frequently used in the conservative (liberal) group. These 40 hashtags contribute about 50% of the difference between the two groups. We use the top 100 hashtags to generate the political-decahose dataset.

Table 3: Percentage of accounts with particular self-identifications in different sub-groups of train-recent accounts and the corresponding political-recent baseline. To match the accounts, we first find accounts whose descriptions contain the keywords we define, then remove the false positive cases manually.

Theme	Train-conservative	Political-conservative	Train-liberal	Political-liberal
Military & Veterans	25.9%	17.8%	9.5%	5.4%
Family-oriented	30.6%	27.0%	22.8%	19.0%
Constitution, Gun-rights, Patriotism	30.3%	28.0%	6.0%	5.9%
God and Religion	29.8%	27.4%	2.2%	4.6%
Pro-life	8.6%	6.4%	0%	0%
BLM	0.03%	0.01%	17.2%	23.7%
QAnon	18.9%	21.4%	0%	0%
Follow	11.6%	9.0%	23.4%	15.9%
No DMs (direct messages)	6.5%	5.3%	15.3%	4.1%

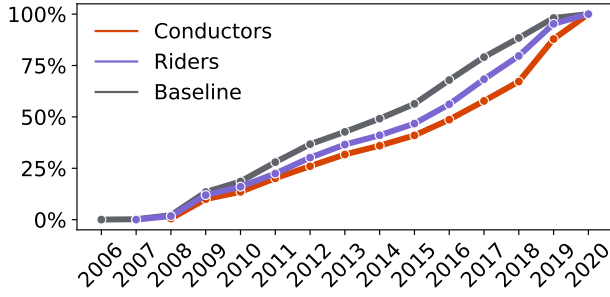


Figure 6: Cumulative distributions of creation years for accounts in the train-recent dataset and political-recent baseline. Mann-Whitney U tests show that the distribution for conductors is significantly different from those of the other two groups ($p < 0.01$) with small effect sizes. The difference between the distributions of riders and baseline is not significant ($p = 0.48$).

We note that the themes of each train group match those of the corresponding political baseline.

Account profiles contain further information that can provide insights about the political follow trains. The creation date distributions in Figure 6 show that conductor accounts tend to be created more recently than riders and baseline accounts in political-recent. We speculate that this could be due to suspension of conductor accounts and/or purchase of fake rider accounts.

5.2 Following Behavior

Let us examine the following behavior of train accounts. In the train-decahose and political-decahose datasets, each tweet contains the profile information that reflects the status of the account at the time the tweet was posted. A series of tweets by an account provides snapshots capturing the temporal evolution of the account’s profile. By examining the difference in friend counts between tweets in consecutive days, we can estimate the account’s daily growth rate of friends. We plot the distribution of the average increase in friends for different accounts in Figure 7(top). The train

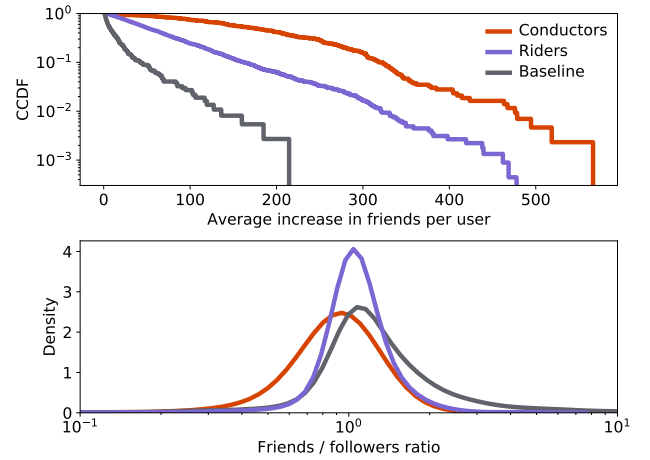


Figure 7: Top: Complementary cumulative distributions of average daily friend growth for accounts in the train-decahose dataset and political-decahose baseline. We excluded a few cases with a daily growth above the maximum allowed by Twitter, which may be due to Twitter API data errors. Bottom: Kernel Density Estimations (KDE) of the distributions of the friend/follower ratios for the same accounts. Mann-Whitney U tests show that the distributions are significantly different ($p < 0.01$) with large effect sizes.

accounts establish social ties much more aggressively than accounts in the political baseline.

It is important to note that when an account has more than 5,000 friends, Twitter imposes constraints on their friend/follower ratio. Although these constraints are not published, it is generally understood that an account must have a friend/follower ratio below a critical value close to one. In other words, one can follow additional accounts only if they have a similar number of followers. As a result, political train accounts must accumulate followers in order to aggressively follow new accounts. This provides a key to interpreting their following behavior. In fact, Figure 7(bottom) shows that rider accounts have a friend/follower ratio that is narrowly distributed

around one. This suggests that the mutual following patterns promoted by political trains are designed to create increasingly dense networks while circumventing Twitter’s policy.

5.3 Accounts Suspension

Since the 2018 U.S. midterm election season, Twitter implemented more aggressive enforcement actions against policy violations and suspended accounts at a higher rate.² Since the accounts involved in follow trains, especially the conductors, are violating platform policies, we are interested in their suspension rates. We check the profile status of all accounts in train-net and tagging-net through Twitter’s API. Note that the political-recent dataset is not a suitable baseline here because suspended accounts are removed from the historical archive. As of August 31, 2020, 996 (0.09%) of the train-net accounts have been suspended and 345 (0.03%) are unavailable, in addition to 1,061 accounts that were inaccessible at the beginning of data collection. This contrasts with the accounts in the tagging-net baseline, 881 of which (0.16%) are suspended and 350 (0.06%) are unavailable. These numbers suggest that Twitter has not taken aggressive actions towards the train accounts, which is consistent with the observation that many train accounts were created years ago.

6 ABUSIVE BEHAVIORS

In this section, we turn our focus to certain automated and abnormal behaviors of train accounts that may flag abuse.

6.1 Automation

As discussed in the Background section, various inauthentic actors might be involved in political follow trains; we focus on bots here. To estimate the prevalence of automation, we adopt BotometerLite,³ a scalable off-the-shelf bot detection tool [47]. For each account, BotometerLite generates a bot score between 0 and 1 with higher values indicating more bot-like behaviors.

The distributions of bot scores in Figure 8 show that most of the accounts in train-recent and political-recent are human-like. However, we do observe a larger number of bot-like accounts among train accounts and especially conductors. This suggests that follow trains may be sustained in part by social bots.

6.2 Abnormal Tweeting Behaviors

An examination of the user timelines of some of the train accounts shows that they publish a significant volume of tweets. In addition to this, while annotating the datasets, we noticed that some of the train accounts routinely delete their tweets in bulk. Users have the freedom to delete their posts and may have legitimate reasons to do so [3]. However, the deletion feature can also be abused. For example, trolls and malicious bots may delete their tweets to conceal their activities and intentions and evade detection [16, 46, 48].

To infer the tweet publishing and deletion events, we use the train-decahose and political-decahose datasets. Each tweet contains an associated user object with information that reflects the status of the account when the tweet was posted. Although the datasets only contain samples of the tweets from each account, these

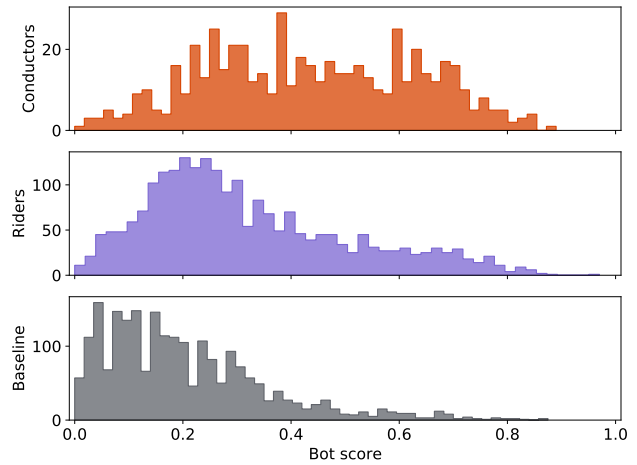


Figure 8: Bot score distributions of (top) conductor, (middle) rider, and (bottom) baseline accounts from the train-recent dataset and political-recent baseline. The percentage of bot-like accounts (bot score above 0.5) is 39.4% for the conductors, 12.0% for the riders, and 5.3% for the baseline. Mann-Whitney U tests show that the distributions are significantly different ($p < 0.01$) with large effect sizes.

can still provide multiple snapshots of an account at different times. A decrease or increase in the tweet counts between consecutive snapshots of an account indicates tweet deletion or new published tweets, respectively. Note that the estimates obtained through this method provide a *lower bound* on the number of new tweets and deletions; the true numbers could be much larger. However, our estimates only consider snapshots when deletions or new tweets are observed; they do not necessarily occur every day.

In a study characterizing tweet deletion behaviors, Almuhiemi et al. [3] show that Twitter users delete 7–11 tweets on average weekly. In contrast, Figure 9 shows the distributions of the estimated numbers of daily tweets deleted and published by accounts in the train-decahose and political-decahose datasets. Train accounts perform tweet deletion at a staggering frequency: on average, 591 for conductors and 343 for riders. These deletion rates are extremely high even in comparison to political baseline accounts (175 on average). And the tails of the distributions highlight accounts with thousands of deleted tweets per day. Although Twitter terms prevent us from inspecting the deleted content, such abnormal behaviors are strongly suggestive of abuse. Train accounts, and especially riders, also produce higher volumes of tweets compared to the baseline.

To further explore these high tweeting volumes, let us examine the rate at which new tweets are produced. In Figure 10 we plot the distributions of seconds between consecutive tweets. The conductor distribution is more skewed to the left, indicating shorter time intervals between tweets. This behavior is less pronounced in the rider and baseline accounts. We measure the fraction of accounts that tweet at least 10 tweets within an interval of 10 seconds. We find that 73% of the conductors meet this threshold, while the percentage decreases to 50% for riders and 41% for the baseline

²<https://transparency.twitter.com/en/reports/rules-enforcement.html>

³<https://botometer.osome.iu.edu/botometerlite>

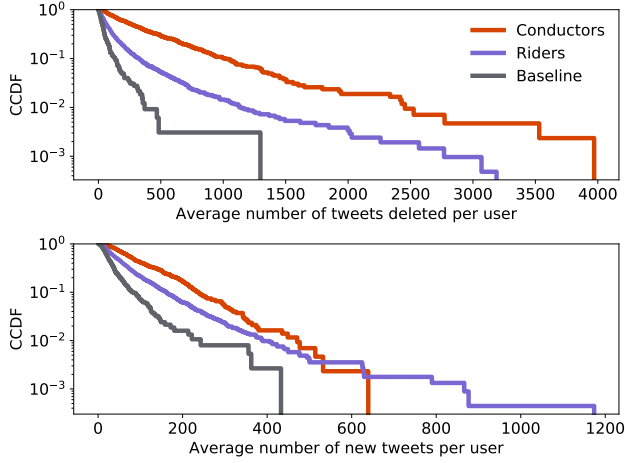


Figure 9: Complementary cumulative distribution of (top) average estimated daily tweet deletions per account and (bottom) average estimated daily new tweets per account in the train-decahose dataset and political-decahose baseline. Mann-Whitney U tests show that the distributions are significantly different ($p < 0.01$) with large effect sizes.

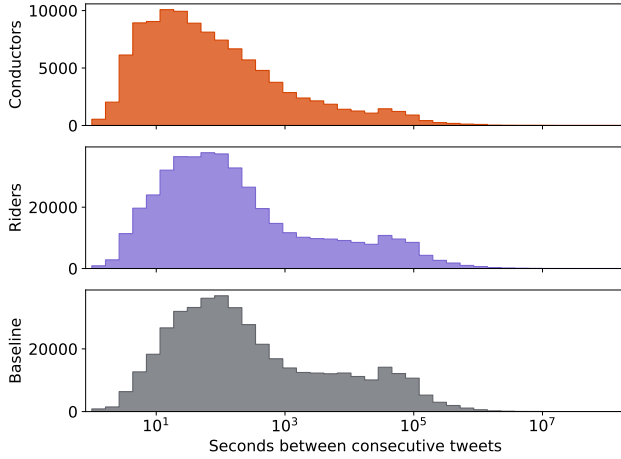


Figure 10: Distribution of burstiness, measured in seconds between consecutive tweets, for (top) conductor, (middle) rider, and (bottom) baseline accounts from the train-recent and political-recent datasets. Mann-Whitney U tests show that the distributions are significantly different ($p < 0.01$) with large effect sizes.

accounts. This finding is suggestive of suspicious tweeting patterns — either using automated tools or carelessly retweeting every tweet in the feed regardless of its content. Twitter recently announced a new mechanism to curb this kind of behavior.⁴

⁴<https://twitter.com/TwitterComms/status/1309178715717369856>

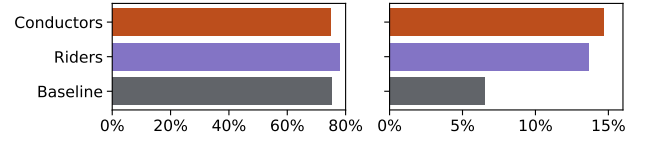


Figure 11: Percentage of users that share links to low-credibility sources above a threshold: (left) $p_{\text{low-cred}} > 0$ and (right) $p_{\text{low-cred}} > 0.25$, among conductor and rider accounts in the train-recent dataset and baseline accounts in the political-recent dataset.

7 SPREADING TOXIC INFORMATION

In this section, we analyze the spread of low-credibility news and conspiracy theories by political train accounts.

7.1 Low-credibility News

We identify low-credibility news by compiling a list of low-credibility domains from the recent literature [6, 23, 34, 36], using the same method as Yang et al. [45]. We then find URLs that link to low-credibility news outlets embedded in the tweets in the train-recent and political-recent datasets. Note that although breitbart.com is labeled as a low-credibility source, it is very popular among conservative Twitter accounts in general. To make sure that our results are not dominated by this single domain, we remove it from the list of low-credibility domains used in the present analysis.

For each account, we compute the ratio $p_{\text{low-cred}}$ of URLs linking to low-credibility domains to the total number of URLs they share (URLs linking to Twitter and other social media sites are disregarded). Figure 11 shows the percentage of users with $p_{\text{low-cred}}$ above different thresholds. In all groups, most of the users share at least one URL linking to low-credibility domains ($p_{\text{low-cred}} > 0$). However, as we increase the threshold to $p_{\text{low-cred}} > 0.25$, the percentages of conductor and rider accounts that meet the criterion are higher than those of the baseline accounts. This finding suggests that although low-credibility information is shared by most political users, it is more dominant in the diet of train accounts.

7.2 Conspiracy Theories

Some conspiracy theories have gained significant momentum and attention in the contexts of the 2020 U.S. presidential election and the COVID-19 pandemic. A particularly notorious and dangerous example is the QAnon conspiracy theory, which was once only accepted by some fringe groups and has not yet drawn much attention from the academic community. As we write this paper, however, multiple media have reported how QAnon has become more mainstream, merged with false narratives about the pandemic, led to violence, and started to affect people’s lives. Popular social media platforms, including Facebook⁵ and Twitter,⁶ recently banned QAnon accounts, pages, and groups.

The results in Figure 5 suggest that at least some of the political train accounts label themselves as QAnon believers. To further

⁵<https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence>

⁶<https://twitter.com/TwitterSafety/status/1285726277719199746>

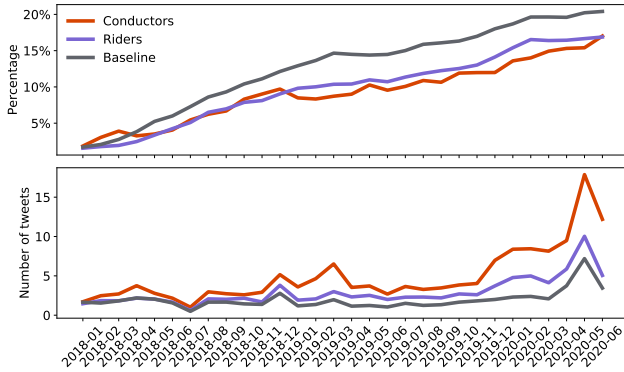


Figure 12: Top: Monthly percentages of accounts having at least one QAnon keyword in their profile descriptions. Bottom: Monthly average numbers of tweets containing QAnon-related keywords. In both plots we show the accounts in the train-decahose dataset and political-decahose baseline, and normalize the values by the number of accounts that were created until that point in time.

quantify the prevalence of such accounts, we manually curate a list of related keywords: “#q”, “qanon”, “WWG1WGA”, “Qarmy”, “QAngel”, “Qcrush”, “GreatAwakening”, “TheStormIsUponUs”, “qtuesday”, “taketheoath”, “redpill”, “thestormisuponus”, “thestormishere”, “qanonarmy”, “savethechildren”, “questioneverything”, “truthseeker”, “17Anon”, “QDrop”, “WeAreQ”, “WeAreAllQ”, “wwg”, “wga”, “TheStorm”, “TheStormIsComing”, “QCrushMAGARoller”, “QSentMe”, and “flynn”.⁷ We match these keywords using a case-insensitive regular expression.

To show the involvement levels of the users, we match the keywords against the descriptions of accounts in the train-decahose and political-decahose datasets to find those who associate themselves with QAnon. The time evolution of the percentages of QAnon supporters among different groups is shown in Figure 12(top). We observe that the proportion of QAnon accounts increases continuously in all groups over time, which is consistent with the reports that the conspiracy theory is getting more and more popular. Note that the comparison between train and baseline groups is not meaningful here because of baseline sampling bias (QAnon hashtags were included in the selection of baseline accounts).

We also check the rate at which QAnon-related content is generated. To do so, the keywords are matched against the tweets in the train-decahose and political-decahose datasets. We then calculate the monthly average number of QAnon-related tweets for each group and show its time evolution in Figure 12(bottom). We observe that train accounts, especially the conductors, produce QAnon content at higher rates than the baseline accounts, despite the sampling bias mentioned above.

⁷Lt. Gen. Michael Flynn, a former National Security Advisor to President Trump, is included due to the prevalence of references to him among QAnon supporters and his association to the conspiracy theory.

The forced vaccination statistis are using COVID19 to push for mandatory vaccines, and will use it to limit your pursuit of life, liberty, and happiness, all under the guise of safety. #1A #4A
21stcenturywire.com/2019/12/23/bil... via @21WIRE

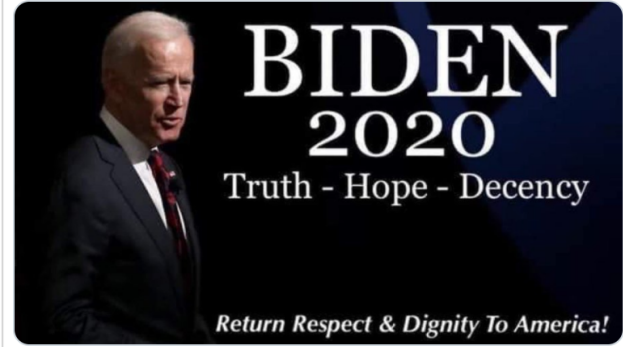


Bill Gates, MIT Develop New 'Tattoo ID' to Check For Va...
21WIRE | New 'wearable tech' will help government and corporations to streamline the administration of mass ...
21stcenturywire.com

8:30 PM · May 1, 2020 · Twitter for iPhone

252 Retweets 30 Quote Tweets 187 Likes

It could come out that tRump faked the entire diagnosis for political gain and his supporters would still think that was okay! Vote friends!! Fake or Real we have an election! #Biden2020 #VoteBlue #COVID #Resist #BlueWave #LGBTQ #BlackLivesMatter 🍌
#BidenHarris2020ToSaveAmerica



9:59 PM · Oct 4, 2020 · Twitter for iPhone

27 Retweets 42 Likes

Figure 13: Top: Screenshot of a tweet containing vaccine-related conspiracy, published by a conservative train account. Bottom: Screenshot of a tweet claiming that President Trump faked his COVID-19 diagnosis, published by a liberal train account.

The conspiratorial content generated and amplified by train accounts is not limited to QAnon. They are also responsible for promoting COVID-19 conspiracy theories, which range from COVID-19 being fake to vaccine misinformation and China being responsible for the virus, as well as glorifying acts of resistance to public health directives, among others. The screenshots of two exemplar tweets are shown in Figure 13.

8 DISCUSSION

In this paper, we provide an in-depth analysis of accounts involved in U.S. political follow trains on Twitter. These accounts are highly

polarized and segregated into two distinct echo chambers pushing conservative and liberal narratives, respectively. Each echo chamber is a dense, clustered, and hierarchical social network organized around a small core of conductor accounts. These echo chambers are manufactured by following behaviors that circumvent platform rules. Although train accounts have similar descriptions to those of “ordinary” political accounts, they are more likely to display inauthentic and abusive behaviors, such as abnormal tweet deletions and high-frequency posting. They are also responsible for spreading low-credibility news as well as conspiracy theories. These abusive behaviors may lead to real-world harm.

The political train phenomenon poses new challenges to social media platforms. Moderation is needed to mitigate the undesirable outcomes of political trains. Although Twitter has stepped up their efforts to maintain a healthy online discussion around critical issues like elections and public health, our finding that the train accounts have a similar suspension rate as the baseline group suggests that aggressive actions have not yet been taken to curb the abusive behaviors of political trains. Political trains also negatively affect the online experience of ordinary social media users who are exposed to false, inflammatory, and hyper-political information amplified by train accounts. Learning the characteristics of such accounts can help the public recognize and ignore, block, or report trains.

Although the present study focuses on Twitter in the U.S., political trains also exist on other platforms like Facebook and Instagram, in other countries, and different languages. This calls for an extension of the present analysis to different platforms, contexts, and languages in the future. Our framework can be easily applied to different platforms and contexts, provided that data from such platforms is available.

REFERENCES

- [1] Anupama Aggarwal and Ponnurangam Kumaraguru. 2015. What they do in shadows: Twitter underground follower market. In *2015 13th Annual Conference on Privacy, Security and Trust (PST)*. IEEE, 93–100.
- [2] Wasim Ahmed, Josep Vidal-Alaball, Joseph Downing, and Francesc López Seguí. 2020. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *Journal of Medical Internet Research* 22, 5 (2020), e19458.
- [3] Hazim Almuhammedi, Shomir Wilson, Bin Liu, Norman Sadeh, and Alessandro Acquisti. 2013. Tweets are forever: a large-scale quantitative analysis of deleted tweets. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 897–908.
- [4] Adam Badawy, Emilio Ferrara, and Kristina Lerman. 2018. Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 258–265.
- [5] Alessandro Bessi and Emilio Ferrara. 2016. Social bots distort the 2016 US Presidential election online discussion. *First Monday* 21, 11 (2016).
- [6] Alexandre Bovet and Hernán A Makse. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature communications* 10, 1 (2019), 1–14.
- [7] Jonathan Bright. 2018. Explaining the Emergence of Political Fragmentation on Social Media: The Role of Ideology and Extremism. *Journal of Computer-Mediated Communication* 23, 1 (2018), 17–33. <https://doi.org/10.1093/jcmc/zmx002> arXiv:<https://academic.oup.com/jcmc/article-pdf/23/1/17/23822774/zmx002.pdf>
- [8] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, P Krishna Gummadi, et al. 2010. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- [9] Michael D Conover, Emilio Ferrara, Filippo Menczer, and Alessandro Flammini. 2013. The digital evolution of occupy wall street. *PloS one* 8, 5 (2013), e64679.
- [10] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2015. Fame for sale: efficient detection of fake Twitter followers. *Decis. Support Syst.* 80 (2015), 56–71.
- [11] C A Davis, G L Ciampaglia, L M Aiello, K Chung, M D Conover, E Ferrara, A Flammini, G C Fox, X Gao, B Gonçalves, P A Grabowicz, K Hong, P Hui, S McCaulay, K McKelvey, M R Meiss, S Patil, C Peli Kankanamale, V Pentchev, J Qiu, J Ratkiewicz, A Rudnick, B Serrette, P Shiralkar, O Varol, L Weng, T Wu, A J Younge, and F Menczer. 2016. OSOme: The IUNI Observatory on Social Media. *PeerJ Computer Science* 2 (2016), e87. <https://doi.org/10.7717/peerj-cs.87>
- [12] Ashok Deb, Luca Luceri, Adam Badaway, and Emilio Ferrara. 2019. Perils and Challenges of Social Media and Election Manipulation Analysis: The 2018 US Midterms. In *Companion Proc. WWW Conf.* 237–247. <https://doi.org/10.1145/3308560.3316486>
- [13] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113, 3 (2016), 554–559.
- [14] Michela Del Vicario, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2017. Modeling confirmation bias and polarization. *Scientific reports* 7 (2017), 40391.
- [15] Renee DiResta, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. 2019. The tactics & tropes of the Internet Research Agency. (2019).
- [16] Tugrulcan Elmas, Rebekah Overdorf, Ahmed Furkan Özkalay, and Karl Aberer. 2019. Lateral Astroturfing Attacks on Twitter Trending Topics. *arXiv preprint arXiv:1910.07783* (2019).
- [17] Emilio Ferrara. 2017. Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election. *First Monday* 22, 8 (2017).
- [18] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.
- [19] Claudia Flores-Saviaga, Brian C Keegan, and Saiph Savage. 2018. Mobilizing the trump train: Understanding collective action in a political trolling community. *arXiv preprint arXiv:1806.00429* (2018).
- [20] Ryan J Gallagher, Morgan R Frank, Lewis Mitchell, Aaron J Schwartz, Andrew J Reagan, Christopher M Danforth, and Peter Sheridan Dodds. 2020. Generalized Word Shift Graphs: A Method for Visualizing and Explaining Pairwise Comparisons Between Texts. *arXiv preprint arXiv:2008.02250* (2020).
- [21] R Kelly Garrett. 2009. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication* 14, 2 (2009), 265–285.
- [22] Benjamin Gleason. 2013. # Occupy Wall Street: Exploring informal learning about a social movement on Twitter. *American Behavioral Scientist* 57, 7 (2013), 966–982.
- [23] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 US presidential election. *Science* 363, 6425 (2019), 374–378.
- [24] Thomas T. Hills. 2019. The Dark Side of Information Proliferation. *Perspectives on Psychological Science* 14, 3 (2019), 323–330. <https://doi.org/10.1177/1745691618803647>
- [25] K.H. Jamieson and J.N. Cappella. 2008. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford University Press.
- [26] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. 2010. Identification of influential spreaders in complex networks. *Nature physics* 6, 11 (2010), 888–893.
- [27] David Lazer, Matthew Baum, Yochai Benkler, Adam Berinsky, Kelly Greenhill, Filippo Menczer, Miriam Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven Sloman, Cass Sunstein, Emily Thorson, Duncan Watts, and Jonathan Zittrain. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- [28] Jae Kook Lee, Jiyoung Choi, Cheonsoo Kim, and Yonghwan Kim. 2014. Social media, network heterogeneity, and opinion polarization. *Journal of communication* 64, 4 (2014), 702–722.
- [29] Yuli Liu, Yiqun Liu, Min Zhang, and Shaoping Ma. 2016. Pay Me and I’ll Follow You: Detection of Crowdturfing Following Activities in Microblog Environment.. In *IJCAI*. 3789–3796.
- [30] Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, et al. 2011. The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International journal of communication* 5 (2011), 31.
- [31] Filippo Menczer, Santo Fortunato, and Clayton A Davis. 2020. *A First Course in Network Science*. Cambridge University Press.
- [32] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, and Maurizio Tesconi. 2020. Coordinated Behavior on Social Media in 2019 UK General Election. *arXiv preprint arXiv:2008.08370* (2020).
- [33] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. 2020. Uncovering Coordinated Networks on Social Media. *arXiv preprint arXiv:2001.05658* (2020).
- [34] Gordon Pennycook and David G Rand. 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116, 7 (2019), 2521–2526.
- [35] Kazutoshi Sasahara, Wen Chen, Hao Peng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2020. Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science* (2020).

- <https://doi.org/10.1007/s42001-020-00084-7>
- [36] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature communications* 9, 1 (2018), 1–9.
 - [37] Karishma Sharma, Emilio Ferrara, and Yan Liu. 2020. Identifying Coordinated Accounts in Disinformation Campaigns. *arXiv preprint arXiv:2008.11308* (2020).
 - [38] Kate Starbird and Leysia Palen. 2012. (How) will the revolution be retweeted? Information diffusion and the 2011 Egyptian uprising. In *Proceedings of the acm 2012 conference on computer supported cooperative work*. 7–16.
 - [39] Massimo Stella, Emilio Ferrara, and Manlio De Domenico. 2018. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences* 115, 49 (2018), 12435–12440.
 - [40] Leo Graiden Stewart, Ahmer Arif, A Conrad Nied, Emma S Spiro, and Kate Starbird. 2017. Drawing the lines of contention: Networked frame contests within# BlackLivesMatter discourse. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–23.
 - [41] Gianluca Stringhini, Gang Wang, Manuel Egele, Christopher Kruegel, Giovanni Vigna, Haitao Zheng, and Ben Y Zhao. 2013. Follow the green: growth and dynamics in twitter follower markets. In *Proceedings of the 2013 conference on Internet measurement conference*. 163–176.
 - [42] Zeynep Tufekci and Christopher Wilson. 2012. Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of communication* 62, 2 (2012), 363–379.
 - [43] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
 - [44] Magdalena Wojcieszak. 2010. 'Don't talk to me': effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism. *New Media & Society* 12, 4 (2010), 637–655.
 - [45] Kai-Cheng Yang, Christopher Torres-Lugo, and Filippo Menczer. 2020. Prevalence of Low-Credibility Information on Twitter During the COVID-19 Outbreak. In *Proc. ICWSM Intl. Workshop on Cyber Social Threats (CySoc)* (Virtual). <https://doi.org/10.36190/2020.6>
 - [46] Kai-Cheng Yang, Onur Varol, Clayton A Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. Arming the public with artificial intelligence to counter social bots. *Hum. Behav. and Emerg. Tech.* 1, 1 (2019), 48–61.
 - [47] Kai-Cheng Yang, Onur Varol, Pik-Mai Hui, and Filippo Menczer. 2020. Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1096–1103.
 - [48] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019. Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. In *Companion Proceedings of The 2019 World Wide Web Conference*. 218–226.
 - [49] John Zarocostas. 2020. How to fight an infodemic. *The Lancet* 395, 10225 (2020), 676.