Bots, Disinformation, and the First Trump Impeachment

Michael Rossetti,¹ ² Tauhid Zaman³

¹Georgetown University, 38th and O Street NW, Washington, 20057, DC, United States
²New York University, 44 West 4th Street, New York, 10012, NY, United States
³Yale University, 165 Whitney Ave, New Haven, 06511, CT, United States

Correspondence: mjr300@georgetown.edu, mjr14@nyu.edu; tauhid.zaman@yale.edu;

Automated social media accounts, known as bots, have been shown to spread disinformation and manipulate online discussions. We study the behavior of bots on Twitter during the first impeachment of U.S. President Donald Trump. We collect over 67.7 million impeachment related tweets from 3.6 million users, along with their 53.6 million edge follower network. We find although bots represent 1% of all users, they generate over 31% of all impeachment related tweets. Among supporters of the Qanon conspiracy theory, a popular disinformation campaign, bots have a prevalence near 10%. We quantify bot impact using the generalized harmonic influence centrality measure. There are more pro-Trump bots, but on a per bot basis, anti-Trump and pro-Trump bots have similar impact, while Qanon bots have less impact. This lower impact is due to the homophily of the Qanon follower network, suggesting this disinformation is spread mostly within online echo-chambers.
Introduction

On December 18, 2019, the United States House of Representatives voted to approve articles of impeachment against President Donald Trump. The resulting trial in the Senate concluded on February 5, 2020 when the Senate voted to acquit the president. During this period, online social media platforms became a battlefield for information warfare between supporters and opponents of the president (1). While much of the activity originated from users engaging in genuine political debate, a significant proportion came from accounts known as bots. Bots are automated social media accounts programmed to share certain content and interact with other users. Due to their automation, bots have the potential to manipulate social media discussions (2–4). The use of bots to spread targeted political messages online, known as computational propaganda (5), has been growing in recent years. Studies have chronicled efforts by bots to manipulate online discussions surrounding U.S. elections and political events since 2016 (6–12).

Recent trends show bots amplifying and spreading false or misleading news stories and conspiracy theories (collectively known as disinformation). The phenomenon of bots spreading disinformation has been observed in online discussions ranging from U.S. elections (12) to public health issues such as vaccines (13, 14) and the COVID-19 pandemic (15). The introduction of disinformation makes the risk posed by bots much greater, as it allows bots to create false narratives that take hold with a large population, resulting in dangerous outcomes. For example: on December 4, 2016, an armed believer of the “Pizzagate” conspiracy theory fired his weapon in a D.C. pizza parlor (16–19); and on January 6, 2021 a mob of Trump supporters influenced by the “Q-Anon” conspiracy theory would storm the U.S. Capitol building in an effort to forcibly overturn the results of the 2020 U.S. presidential election (20, 21).

Due to the unique threat bots pose, it is important to be able to quantify their impact in spreading disinformation. Existing studies provide simple statistics such as the number of bots
or volume of content they share (10, 11). However, these statistics do not incorporate the interaction of social network structure, user activity levels, and user sentiment. Some studies have looked individually at the positioning of bots within the social network (22) or the bot retweet response time (2), but not the interaction of these factors. Recent work has presented a novel network centrality measure known as generalized harmonic influence centrality that combines all of these factors to assess bot impact in online political discussions (23, 24). This centrality measure provides a better measure of bot impact than individually examining each factor.

In this study, we focus on bots discussing the first impeachment of U.S. President Donald Trump on Twitter. We find that bots are 66 times more active than normal human users, producing nearly one third of all impeachment-related content, despite representing less than 1% of all accounts discussing the impeachment. Bots tend to share news from lower quality sources (including disinformation) than their co-partisan human counterparts. Bots have an unusually high prevalence among Qanon conspiracy supporters, indicating that efforts are being made to artificially amplify this disinformation. Using generalized harmonic influence centrality, we show that pro-Trump and anti-Trump bots have a similar level of impact per bot, but Qanon bots have a lower per bot impact. Analysis of the Qanon follower network suggests this lower impact is due to Qanon users existing in an online echo-chamber with a high amount of ideological homogeneity.

**Results**

**Data Collection**

We constructed a list of keywords and hashtags related to the first impeachment of President Donald Trump. This list included partisan terms advocating for each side in the discussion, and non-partisan terms related to developments in the impeachment (see Methods). From December 12, 2019 to March 24, 2020, we collected tweets which contained at least one of these keywords.
This collection process resulted in a dataset of 67.7 million tweets posted by 3.6 million unique Twitter accounts. We also collected the user profiles of these accounts and their 53.6 million edge follower network (see Methods). We provide summary statistics of this dataset in Table 1.

**Account Types**

We classified the Twitter accounts into different groups which characterized different aspects of their preferences and behavior. These included political partisanship, support for the Qanon movement, and whether or not the account was an automated bot. Each group label was assigned using a different method. A summary of the group statistics is provided in Table 1.

Table 1: Statistics of account types and tweets in Twitter dataset for the first impeachment of President Donald Trump.

| Partisanship | Bot status | Qanon status | Accounts | Tweets |
|--------------|------------|--------------|----------|--------|
| Anti-Trump   | Human      | Normal       | 2,273,831| 25,103,340 |
| Pro-Trump    | Human      | Normal       | 1,279,638| 18,717,915 |
| Pro-Trump    | Human      | Q-anon       | 22,926   | 2,845,521  |
| Pro-Trump    | Bot        | Normal       | 11,571   | 9,880,481  |
| Anti-Trump   | Bot        | Normal       | 10,145   | 9,220,258  |
| Pro-Trump    | Bot        | Q-anon       | 2,434    | 1,899,042  |

|               |             |             | **3,600,545** | **67,666,557** |

First, we considered the political partisanship of the accounts. We trained a bidirectional encoder representations from transformers (BERT) model to measure the political partisanship of any given tweet text (25). The model was trained using a subset of the impeachment tweets for which we were able to assign ground-truth labels (see Methods). A label of zero represents strong anti-Trump sentiment and a label of one represents strong pro-Trump sentiment. We used the trained model to measure the sentiment of all tweets in our dataset. Then, we assigned a partisanship score to each Twitter account equal to the mean value of the partisanship score of their tweets. The accounts were labeled anti- (pro-) Trump if their partisanship score was less...
than or equal to (greater than) 0.5. In total, our dataset had 2.3 million anti-Trump accounts and 1.3 million pro-Trump accounts. This left leaning bias in the Twitter conversation aligns with findings from previous studies (23).

We next identified accounts who were supporters of the Qanon movement. This was done by using the profile descriptions of the accounts. We utilized a list of terms commonly used by Qanon supporters (26). Any pro-Trump account that included at least one of these terms in their profile description was labeled as Qanon. Using this process, we identified 25,360 Qanon accounts.

To identify bots we used an algorithm based on a factor graph model (23). This algorithm simultaneously detects multiple bots among a set of Twitter accounts, using only their retweet network (see Methods). We chose this algorithm because it had minimal data requirements compared to other algorithms (27) (as we are easily able to construct the retweet network from the collected tweets), and also because it exhibited performance similar to or better than other algorithms (23). We applied the bot detection algorithm to daily retweet networks constructed from tweets posted each day. This allowed us to identify daily sets of active bots. Our final set of bots consisted of the union of these daily bot sets. In total we found 24,150 bots, of which 10,145 were anti-Trump and 14,005 were pro-Trump. This ratio of bots to humans aligns with other studies of U.S. politics on Twitter (12, 23).

We summarize the prevalence of bots in different groups in Figure 1. We see that bots have a slightly higher prevalence among pro-Trump accounts than anti-Trump accounts (p-value < 10^-6). However, the bot prevalence among Qanon supporters is nearly an order of magnitude larger than it is among normal accounts (p-value < 10^-6). This suggests malicious actors may be attempting to use artificial accounts to amplify Qanon content. On one hand, this may be reassuring, as it suggests there aren’t as many real Qanon supporters on social media as there may appear. On the other hand, significant bot presence means that Qanon content is
being spread at a higher rate and with a potentially higher reach than could be achieved with humans alone. This high bot fraction is even more concerning when looking at the tweet rate of the accounts in Figure 1, as Qanon human supporters tweet approximately ten times more frequently than regular humans (p-value $< 10^{-6}$), and bots tweet approximately one hundred times more frequently than regular humans (p-value $< 10^{-6}$). Having a large number of Qanon bots can lead to an unusually large amount of Qanon content being spread through Twitter. This potentially enhanced reach increases the risk posed by an already dangerous ideology.

| Account Type | Tweet rate [tweets/day] |
|--------------|-------------------------|
| Qanon        |                         |
| pro-Trump    |                         |
| anti-Trump   |                         |

Figure 1: (left) Fraction of bots and (right) average tweet rate of each account type category.

**Content**

There are differences in the nature of content posted by the various account types. We first considered the quality of news stories shared by the accounts. We leveraged a previously published set of 60 news sites (20 mainstream, 20 hyper-partisan, and 20 fake news, with liberal and conservative leaning sites in each category) whose trustworthiness had been rated by eight professional fact-checkers (28). We followed the approach used in prior work (29, 30) and calculated a media quality score for each user by averaging the trustworthiness ratings of any of their impeachment related tweets that contained links to any of those 60 sites. The media quality
score ranges from one to five, with higher values indicating higher trustworthiness. Like other researchers in this space (29–32), we use source trustworthiness as a proxy for article accuracy, because it is not as feasible to rate the accuracy of every shared link.

Figure 2 shows the average media quality score of the different account types. The first striking observation is that the media quality score is much higher for anti-Trump accounts than pro-Trump accounts. This is consistent with past work which found that pro-Trump (Republican) users were much more likely to share news from untrustworthy news sites than anti-Trump (Democratic) users (32). We also find that within each partisan group, the media quality score of bots is lower than that of humans (p-value $< 10^{-6}$ for bots versus humans in each partisan group). Qanon humans have a lower media quality score than both normal pro-Trump humans (p-value $< 10^{-6}$) and normal pro-Trump bots (p-value = 0.0002). However, we find no statistical difference between the average media quality score for Qanon humans and Qanon bots (p-value = 0.29).

In addition to what news media the accounts are sharing, we also investigate the tone and sentiment of the tweets they post. One measure of this is known as toxicity, which captures how harmful or unpleasant a tweet is. We measured the toxicity of all tweets in our dataset using the Detoxify model, a state of the art toxicity detector which has achieved excellent performance on multiple benchmarks (33). The model produces toxicity scores ranging from zero to one, with higher values indicating higher toxicity. The average toxicity score for each account type is shown in Figure 2. One striking observation here is that anti-Trump humans have the highest toxicity score by a wide margin. The next highest toxicity group are pro-Trump humans. Qanon humans and bots have lower toxicity than normal humans (p-value $< 10^{-6}$). Bots of each partisanship group have lower toxicity levels than their co-partisan human counterparts (p-value $< 10^{-6}$). The high toxicity levels of anti-Trump humans may be due to their outrage about Trump’s actions, and their disagreement with his acquittal. Conversely, low toxicity levels
among pro-Trump users may reflect their attempts to post positive content in a show of support for the president.

When we focus on the bot accounts, we find two general patterns. Within each partisan group, bots share lower quality media than humans and post less toxic content than humans. From this we can deduce that bots mainly share low quality news stories (relative to their human counterparts within their partisan group), but they tend not to amplify negative messages which use aggressive language. This suggests that the bots are more focused on spreading information and not on agitating users with toxic posts.

![Figure 2: (left) Average media quality score and (right) average toxicity score of tweets posted by different account types.](image)

Finally, we study bot retweet patterns. A retweet occurs when one account re-posts the content of another, thus sharing the original content with all of their followers. High retweet counts indicate high levels of popularity for the original tweet's content and author. Figure 3 shows the accounts most retweeted by anti- and pro-Trump bot accounts, respectively, and the number of retweets each account received from bots. This allows us to examine which accounts benefited most from bot activity.
Here we see a clear difference in the structure of the bot retweet distributions. On the anti-Trump side, the bot retweets are distributed in a rather uniform manner over the top ten accounts, with retweet counts ranging from 240,336 to 92,233. However, bot retweet counts on the pro-Trump side range from 765,512 to 113,261, with a very concentrated distribution in which Donald Trump receives the most bot retweets by far, earning more than twice the amount of retweets than the second most retweeted account. This suggests Donald Trump is a singular figure in terms of bot retweets.

It is also interesting to note that the top ten bot retweeted accounts on the anti-Trump side are all political pundits. There are no elected or government officials. In contrast, among the top ten accounts on the pro-Trump side, there are three elected officials (President Donald Trump, Congressman Matt Gaetz, Congressman Andy Biggs) and one cabinet official (Chief of Staff Mark Meadows). Finally, we note that the official Twitter account of the republican party (GOP) is among the top ten accounts retweeted by bots, while the account of the democrat party (DNC) is not. This analysis suggests that pro-Trump bots are more actively amplifying officials with political or government power than anti-Trump bots.

**Network Structure**

Bots and Qanon supporters have networks that exhibit distinct properties. We first consider the bot follower network. Figure 4 shows the number of unique accounts who follow at least one bot, and the partisanship of the bots they follow. We find that pro-Trump bots have more followers than anti-Trump bots, which may be due to the pro-Trump bots’ greater numbers. Less than 6% of the bot followers follow bots in both partisan groups, suggesting that the bot followers’ network is highly polarized. This polarization is very visible in the bots’ follower network, as shown in Figure 5. Here we see that the two partisan groups of bots are almost totally disconnected. To obtain a more quantitative measure of this polarization, we calculate
the fraction of followers of each bot type who are co-partisan (i.e. those who share the same political affiliation). A higher value for this measure indicates greater ideological homogeneity in the followers of the bots. Figure 4 shows the co-partisan fractions for each bot type. We find the values are quite high, being above 0.87 for all bot types. There is a 1% difference in the co-partisan follower fraction between the anti-Trump and non-Qanon pro-Trump bots (p-value < $10^{-6}$). However, the Qanon bots have have a co-partisan fraction that is 6% to 7% greater than the non-Qanon bots (p-value < $10^{-6}$). This suggests that Qanon bots have an audience that is even more partisan than a standard bot.

The Qanon follower network, comprised of Qanon bots and humans, has an interesting structure. We find that this network contains a core of bots which are connected to each other, surrounded by a periphery of humans. Interestingly, these humans are only connected to the
bots, and not to each other. We show this network in Figure 5, where the nodes are laid out with bots in the center to highlight this structural property. The connectivity among bots and lack of connectivity among humans suggests there is a hierarchical structure within the Qanon community. The bots act as sources of content for the humans. The humans appear not to engage with each other, but rather mainly consume content from the bots. We will see later that this network structure has implications for the impact of Qanons.

Figure 4: (left) Venn diagram of users who follow anti-Trump and pro-Trump bots. (right) Fraction of co-partisan followers for different bot types.

Figure 5: (left) Follower network of bot accounts colored by partisanship, where anti-Trump bots are blue, and pro-Trump bots are red. (right) Follower network of Qanon accounts, where bots are colored red and humans are colored green.
Impact

We next provide a quantitative measure of the impact the bots have on the impeachment discussion. Thus far we have presented a detailed analysis of the activity, sentiment, and network structure of the bots. Impact is a combination of all of these factors. We would expect more active bots to have more impact, as they post more content. Bots with larger network reach should also have greater impact. Finally, bots whose followers are not strong co-partisans would have more impact because they can persuade these followers to their side. In contrast, strong co-partisan followers who already agree with the bots likely cannot be persuaded further. A measure that combines these factors is known as harmonic influence centrality (34). This is a network centrality based on a classic model for opinion dynamics (35) which incorporates stubborn users with immutable opinions (36). Harmonic influence centrality measures how much a set of nodes shifts the average equilibrium opinion in the network. The centrality naturally incorporates activity, network reach, and opinions to measure the impact of nodes in a network because it is based on an opinion dynamics model which utilizes these factors. In its original form, harmonic influence centrality used only the network structure and activity level. A more recent version, known as generalized harmonic influence centrality (GHIC), incorporates additional data, such as the node sentiment, making it more appropriate for real social networks. GHIC has been used to quantify the impact of bots in Twitter networks discussing various geopolitical events (23). Because our data is quite similar in nature, we use this generalized version of the measure to quantify the impact of the bots.

To apply GHIC, one first must define the network. We want to calculate a daily impact measure for the bots to see how their impact evolves over time. Therefore, the networks we use are daily active follower networks. For a given day, the daily active follower network is the sub-network of the entire Twitter follower network induced by accounts which are active (post at least one tweet) that day. We follow the approach in (23) to calculate the GHIC of different
groups of bots on these networks (see Methods).

We first look at the GHIC of all bots. The daily GHIC of all bots is shown in Figure 6, where positive values indicate a shift toward pro-Trump opinions, and negative values indicate a shift toward anti-Trump opinions. This shows on a given day which partisan side of bots had greater impact. We observe that for most days, the pro-Trump bots have greater impact. This is likely due to their greater number and reach. However, there are days when the anti-Trump bots have greater impact. Upon closer investigation, we find that on these days there was a news story which excited anti-Trump users on Twitter. For instance, on February 11, 2020 Donald Trump suggested that his associate Roger Stone should be pardoned, which likely angered his opponents (37). We see that on this day, the GHIC is negative, indicating that the anti-Trump bots have greater impact. Another interesting day is December 29, 2019. On this day Donald Trump posted a controversial tweet referring to U.S. Speaker of the House of Representatives Nancy Pelosi as “Crazy Nancy”. We see that the GHIC had a large positive value that day, indicating that the pro-Trump bots had greater impact. In general, we find that GHIC achieves a large magnitude on days with polarizing events. This suggests that bots may have greater impact on days when partisan tensions are highest.

In addition to the daily GHIC, we are also interested in the efficiency of each group of bots. By efficiency, we mean the GHIC per bot for a group of bots. This measure would identify which group of bots are more impactful, on average. We consider three groups of bots: anti-Trump, pro-Trump non-Qanon, and pro-Trump Qanon. For each group, we calculate their daily GHIC per bot for all days. The boxplot in Figure 6 shows the distribution of the daily GHIC per bot for each group. We see that both groups of non-Qanon bots have very similar GHIC per bot distributions. Statistical tests indicate that the means of these two groups are not different (p-value =0.49). However, we find that Qanon bots have a lower mean GHIC per bot than non-Qanon bots (p-value < 10^{-6}). To understand why Qanon bots are less efficient, we look
at Figure 4. We saw there that Qanon bots have a higher fraction of co-partisan followers than non-Qanon bots, meaning that Qanon bots are more ideologically aligned with their followers. Because of this, these bots cannot impact their followers’ opinions as much, which lowers their GHIC efficiency. Our finding suggests that Qanon bots are not as effective at persuasion. Rather, they likely preach to a converted audience, which diminishes their impact compared to non-Qanon bots.

Figure 6: (top) Daily generalized harmonic influence centrality (GHIC) score for all bots versus date. (bottom) Boxplot of the daily GHIC per bot for different bot types.

Figure 6: (top) Daily generalized harmonic influence centrality (GHIC) score for all bots versus date. (bottom) Boxplot of the daily GHIC per bot for different bot types.
Discussion

Our study of the Twitter discussion surrounding the first impeachment of Donald Trump found that a small number of bots generated a disproportionately large amount of content. In addition, the combined follower reach of these bots is extensive. A primary bot activity is to spread news, and we found that bots generally spread lower quality news than humans. However, the language used in bot tweets is generally less toxic than that of humans. Using the GHIC measure we are able to quantify the daily impact of the bots. We found that bot impact is highest on days with politically charged events. Overall, pro-Trump bots have a greater impact due to their larger numbers and reach, but their per bot impact is similar to anti-Trump bots. Qanon bots have a lower per bot impact than the other bots. This is likely due to their high fraction of co-partisan followers, which limits their persuasive ability.

The excessive reach and activity level of bots, combined with their propensity to share news from low quality sources, are cause for concern. A small number of bots can amplify certain stories or narratives, causing them to reach a large audience. Bots seem to have the greatest impact on days when there is a large amount of partisan agitation, suggesting that bots may be increasing online polarization. However, one encouraging finding is that the Qanon bots, who spread a particularly dangerous form of disinformation, exist within strong echo-chambers, and as a result have less impact than normal bots.

Materials and Methods

Data Collection

The keywords and hashtags used as search criterion for tweets related to the first impeachment of President Donald Trump are shown in Table 2. Some of these words were added to the collection list as news stories developed. We include the date each term was added to the
collection in the table. The Twitter Search API was used to collect in real time any tweets containing at least one of these terms. After an initial trial run from December 12 to 18, 2019, we then ran this collection process continuously from December 20, 2019 to March 24, 2020. Over this entire time period, we were able to collect 67.6 million tweets, posted by 3.6 million unique Twitter users.

Table 2: Keywords used to collect impeachment related tweets, and the date each was added to the collection.

| Keyword                        | Date added     |
|--------------------------------|----------------|
| #FactsMatter                   | 2019-12-12     |
| #IGHearing                     | 2019-12-12     |
| #IGReport                      | 2019-12-12     |
| #ImpeachAndConvict             | 2019-12-12     |
| #ImpeachAndConvictTrump        | 2019-12-12     |
| #SenateHearing                 | 2019-12-12     |
| #TrumpImpeachment              | 2019-12-12     |
| impeach                         | 2019-12-12     |
| impeached                       | 2019-12-12     |
| impeachment                     | 2019-12-12     |
| Trump to Pelosi                | 2019-12-12     |
| #25thAmendmentNow              | 2019-12-18     |
| #ImpeachAndRemove              | 2019-12-18     |
| #ImpeachmentEve                | 2019-12-18     |
| #ImpeachmentRally              | 2019-12-18     |
| #NotAboveTheLaw                | 2019-12-18     |
| trumpletter                    | 2019-12-18     |
| #GOPCoverup                    | 2020-01-22     |
| #ShamTrial                     | 2020-01-22     |
| #AquittedForever               | 2020-02-06     |
| #CountryOverParty              | 2020-02-06     |
| #CoverUpGOP                    | 2020-02-06     |
| #MitchMcCoverup                | 2020-02-06     |
| #MoscowMitch                   | 2020-02-06     |

As a part of this tweet collection process, we also collected the Twitter profile of each user. The profile included information such as the name of the user, location (if provided), and a short
description provided by the user.

We collected user follower networks using a separate process. Our network convention was to have follower edges point from a user to the person that followed them. This way the follower edges point in the direction of information flow, as tweets from a user appear in the timeline of their followers. To build the follower network for the users in our dataset, we used a customized web crawler to collect a list of followings for each user (i.e. the users they follow). We chose to collect the followings rather than the followers for each user because it reduced our data collection burden. We observed that the follower count can be much larger than the following count for a Twitter user, especially for the more popular users. Therefore, to more easily collect all edges in the follower network, we collected the users followings. To be able to collect the follower network in reasonable time, we collected a maximum of 2,000 followings per user. This value was sufficient for our data collection purposes as 85% of the following counts were below this value. In total we obtained 53.4 million edges in this follower network.

**Partisanship Classification Model**

The partisanship classifier we used is a bidirectional encoder representations from transformers (BERT) language representation model (25). Transformers are a neural network architecture that have shown incredible success in language modeling (38). BERT is a bidirectional transformer pre-trained using a combination of a masked language modeling objective and a next sentence prediction objective on the Toronto Book Corpus (800 million words) (39) and Wikipedia (2.5 million words). The BERT model provides a sentence embedding that can be used for many natural language processing tasks such as sentiment classification. We show the architecture of our BERT sentiment classifier in Figure 7. The tweet text is fed into the BERT model and mapped into an embedding representation. We use the base BERT model which produces a 768 dimensional embedding. This representation is then fed to a fully connected single
layer of 768 neurons with linear activation. The linear layer has two outputs which correspond to pro-Trump and anti-Trump sentiment. To obtain the sentiment of the tweet we use the value from the pro-Trump output so strong anti-Trump and pro-Trump sentiment are equal to zero and one, respectively.

We created training data for the model using strongly partisan users. These users were identified by the content of their Twitter profile descriptions. If a user’s description contained any of the words in Table 3 and none of the keywords in Table 4, the user was given a label of zero, indicating anti-Trump sentiment. The opposite was done to identify pro-Trump users, who were given a label of one. The labels of these strongly partisan users were assigned to their impeachment related tweets in our dataset. This process created a labeled training set of over 14 million tweets.

We used 800,000 of the labeled tweets to train the BERT sentiment classifier The tweets were chosen so that 50% were anti-Trump and 50% were pro-Trump, creating a balanced set of labels. We did not use all of the labeled tweets because the size of this dataset made the training process very slow. We found that using 800,000 tweets resulted in a much faster training process while still producing a highly accurate classifier. To prevent over-fitting during training we use a dropout layer on the BERT output with a dropout probability of 0.3. The classifier was trained for ten epochs over the data using the Adam optimizer. Of the 800,000 labeled tweets, we used 80% for training, 10% for validation, and 10% for testing.

The trained classifier is quite effective at measuring opinions about the impeachment. On the held out testing data it achieved a 96.3% accuracy score. We provide some random samples of tweets from each end of the sentiment spectrum and their partisan sentiment measured by the classifier in Table 5.
Table 3: Keywords used to identify anti-Trump users.

| Hashtag          | Description                                      |
|------------------|--------------------------------------------------|
| #BIDEN2020       |                                                 |
| #BLM             | ‘Black Lives Matter’ - a movement for racial equality |
| #BLUEWAVE        |                                                 |
| #BLUEWAVE2020    |                                                 |
| #DEMCAST         | A left-leaning media outlet                      |
| #FBR             | ‘Follow Black Resistance’                        |
| #IMPEACH         |                                                 |
| #IMPEACHANDREMOVE |                                               |
| #IMPEACHTRUMP    |                                                 |
| #IMPEACHTRUMPNOW |                                                 |
| #IMPOTUS         | ‘Impeached POTUS’                                |
| #METOO           | A movement for gender equality                   |
| #NOTMYPRESIDENT  |                                                 |
| #RESIST          |                                                 |
| #RESISTANCE      |                                                 |
| #RESISTER        |                                                 |
| #THERESISTANCE   |                                                 |
| #VOTEBLUE        |                                                 |
| #VOTEBLUE2020    |                                                 |
| #VOTEBLUENOMATTERWHO |                                         |
| #WTP2020         | ‘We The People 2020’                             |

Bot Detection

Bots exhibit certain traits and behaviors that allow them to be identified. Some of these include excessive retweeting and never posting original tweets. These behaviors are likely due to the bots’ automated nature. Many algorithms have been developed for bot detection in Twitter and each has its own strengths. The algorithm of (23) uses a factor graph model that allows for the simultaneous detection of multiple bots based on the collective retweeting behavior of users discussing a specific topic. The data required by this algorithm is the set of tweets, or more precisely, the set of retweets, about the topic. Another popular algorithm is the machine learning based Botometer (formerly BotOrNot) (27) which utilizes a large amount of data about
Table 4: Keywords used to identify pro-Trump users.

| Hashtag               | Description                                                                 |
|-----------------------|-----------------------------------------------------------------------------|
| #1A                   | The First Amendment                                                         |
| #2A                   | The Second Amendment                                                        |
| #AMERICAFIRST         | A Trump campaign slogan                                                     |
| #BUILDKATESWALL       | A Trump campaign slogan                                                     |
| #BUILDTHEWALL         | A Trump campaign slogan                                                     |
| #CODEOFPATRIOT        | A Trump campaign slogan                                                     |
| #CONSERVATIVE         | A Trump campaign slogan                                                     |
| #DEPLORABLE           | Refers to a Hillary Clinton quote from the 2016 election                     |
| #DRAINTHESWAMP        | A Trump campaign slogan                                                     |
| #KAG                  | ‘Keep America Great’ - a Trump campaign slogan                              |
| #MAGA                 | ‘Make America Great Again’ - a Trump campaign slogan                        |
| #NRA                  | The National Rifle Association                                              |
| #POTUS45              | 45th President (Trump)                                                      |
| #QANON                | Related to Qanon conspiracy theory                                          |
| #THEGREATAWAKENING    | Related to Qanon conspiracy theory                                          |
| #TRUMP                |                                                                             |
| #TRUMP2020            |                                                                             |
| #TRUMPTRAIN           |                                                                             |
| #VETERAN              |                                                                             |
| #WALKAWAY             |                                                                             |
| #WWG1WGA              | Related to Qanon conspiracy theory                                          |

an individual account, including followers, friends, tweets, and profile, in order to determine if it is a bot. Botometer is a good algorithm to use when one wants to determine if an individual account is a bot. However, when one wants to find bots among a large set of users discussing a topic, as is the case in our impeachment analysis, the factor graph algorithm is more convenient. We can identify bots with this algorithm without having to collect any additional data, which is important given how large our daily active user sets are. In addition, it has been found that the factor graph algorithm has slightly better performance than Botometer (23).

The factor graph algorithm first constructs a retweet network based on the retweets in a
Table 5: Tweets from testing dataset and their opinion scores assigned by the BERT sentiment classifier. The tweets are ordered by opinion score. An opinion of zero is anti-Trump and an opinion of one is pro-Trump.

| Tweet text                                                                 | Opinion |
|---------------------------------------------------------------------------|---------|
| Mark my words... Trump is starting a war to distract from the Impeachment. | 0.105   |
| LA Times joins growing list of papers calling for Trump’s impeachment      | 0.17    |
| Flynn sentencing - Jan 28 State of the Union - Feb 4 Stone sentencing - Feb 6 Impeachment - Forever | 0.26    |
| Americans have officially lost any belief in the Democrats’ partisan impeachment sham | 0.96    |
| It’s time for the Senate to end this partisan,impeachment sham once and for all. | 0.97    |
| Pelosi and Dems blew it! Impeachment Sham is Backfiring!                   | 0.98    |

Given set of tweets. In this retweet network the nodes are the users who tweet and an edge $(u, v)$ pointing from user $u$ to user $v$ means that $v$ retweets $u$. The edge $(u, v)$ is given a weight $w_{uv}$ which equals the number of times $v$ retweets $u$. Like with the follower network, the edge direction in the retweet network indicates the flow of information. The algorithm uses the retweet network structure to calculate a bot probability for each user. This probability is based on the empirical observations that bots are likely to retweet humans, but unlikely to retweet bots, and humans are likely to retweet humans and less likely to retweet bots (23). This homophily for the humans and heterophily for the bots is utilized by the algorithm to determine bot probabilities from the structure of the retweet network.

We used the factor graph algorithm to identify bots active each day. To apply the algorithm, we first constructed a retweet network from the tweets posted on a given day. Once the retweet network is constructed, we apply the algorithm to the network using the parameters specified in (23) to simultaneously obtain the joint bot probability for all users. We show one example of the bot probability distribution for a single day in Figure 8. As can be seen, the bulk of users have bot probabilities near 0.5, which is the algorithm saying it cannot determine one way or
another what the user is. In the upper probability range we see that the histogram decreases up to approximately 0.8, and then increases afterwards. This suggests there is a cluster of nodes with bot probabilities in the interval $[0.8, 1.0]$. We used the lower probability bound of this cluster as the bot probability threshold so that this cluster of nodes is identified as bots.

We found a similar behavior in the bot probability distribution across all days. Therefore, we used the same bot probability threshold of 0.8 for each day. Any user who had a bot probability over the threshold in at least one day in our dataset was declared to be a bot. Across all days we found 24,150 bots among 3.6 million active accounts.

![Figure 7: Histogram of the bot probabilities calculated by the factor graph algorithm (23) based on the impeachment retweet network on February 1, 2020 (an example day).](image)

**Generalized Harmonic Influence Centrality Calculation**

To calculate the generalized harmonic influence centrality (GHIC) of a set of nodes we need several pieces of information. First, we need the follower network of the nodes. We use the daily active follower network for this purpose, which is the follower network induced by users who tweet about the impeachment on a given day.
Second, we need to characterize the activity level of the users. We define this to be the tweet rate of the users. We measure the tweet rate as the total number of impeachment tweets posted divided by the duration of our data collection. This gives a stable measure of the posting rate of a user. We use this same posting rate for each daily GHIC calculation as it provides an accurate measure of the general activity level of a user.

Third, we need to know the political sentiment of each user. We obtain this by taking the average sentiment of a user’s tweets as measured by our BERT partisanship classifier.

Fourth, we need to identify a subset of users in the daily active follower network as stubborn. GHIC assumes that the stubborn users are not persuadable and their sentiment does not change. We first set all bots to be stubborn, as they are automated accounts that do not respond to persuasion. We identify stubborn users among the humans based on their political sentiment or opinion. Studies have shown that stubborn users have extreme opinions (41). To operationalize this principle, we follow the approach used in (23,24) to define extreme opinions as those below the 10th percentile and above the 90th percentile of the opinions of all users in our dataset. It has been shown that the GHIC is robust to the precise value of these thresholds (23).

Once the requisite information has been obtained, the GHIC of a set of nodes can be calculated. We present the steps for GHIC calculation here, drawing from the presentation in (23). We are given a follower network $G = (V, E)$ with node set $V$ (the Twitter users) and edge set $E$ (the follower edges). We assume each node follows a set of users that we define as their following. Each node $v \in V$ posts content (tweets) at a rate $\lambda_v$. We define the stubborn users as the set $V_0 \subset V$ and the non-stubborn users as $V_1 \subset V$. We define $\Psi$ as the vector of stubborn opinions and $\theta$ as the vector of non-stubborn opinions. We are given the stubborn opinions $\Psi$ and we obtain the non-stubborn opinions $\theta$ by solving

$$G\theta = F\Psi,$$  \hspace{1cm} (1)
where the matrix $G$ is given by the equation:

$$
G_{ij} = \begin{cases} 
-\sum_{k \in \text{following of } i} \lambda_k & i = j, i \in V_1 \\
\lambda_j & i \neq j, (j, i) \in E, i, j \in V_1 \\
0 & \text{else},
\end{cases}
$$

and the matrix $F$ is given by

$$
F_{ij} = \begin{cases} 
-\lambda_j & (j, i) \in E, i \in V_1, j \in V_0 \\
0 & \text{else}.
\end{cases}
$$

Equation (1) is the equilibrium opinions in the opinion dynamics model upon which GHIC is based.

Next we choose a set of nodes $S \subset V$ for which we want to calculate the GHIC. We define a new network $G' = (V', E')$ where $V' = V/S$ and $G'$ is the sub-network of $G$ induced by $V'$. $G'$ is the network $G$ but with the nodes in $S$ removed. Let $\theta$ and $\theta'$ be the solution of equation (1) assuming the underlying network is $G$ and $G'$, respectively. The GHIC of $U$ is then given by

$$
\text{GHIC}(U) = \frac{1}{|V_1/S|} \sum_{i \in V_1/S} \theta_i - \theta'_i.
$$

We see from this that the GHIC of $S$ is the change in mean non-stubborn equilibrium opinion caused by the presence of the $S$ nodes in the network.

**References**

1. K. Roose, *The New York Times* (2019). URL:https://www.nytimes.com/2019/09/26/technology/trump-impeachment-disinformation.html.

2. C. Shao, *et al.*, *Nature Communications* 9 (2018).
3. E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, *Communications of the ACM* 59, 96 (2016).

4. O. Varol, E. Ferrara, C. Davis, F. Menczer, A. Flammini, Online human-bot interactions: Detection, estimation, and characterization (2017).

5. S. C. Woolley, P. N. Howard, *Computational propaganda: political parties, politicians, and political manipulation on social media* (Oxford University Press, 2018).

6. A. Parlapiano, J. C. Lee, *The New York Times* (2018).

7. S. Shane, *The New York Times* 7 (2017).

8. D. Guilbeault, S. Woolley, *The Atlantic* 1 (2016).

9. N. Byrnes, *Technology Rev.* (2016).

10. A. Bessi, E. Ferrara, *First Monday* (2017).

11. A. Bessi, E. Ferrara, *First Monday* 21 (2016).

12. E. Ferrara, H. Chang, E. Chen, G. Muric, J. Patel, *First Monday* (2020).

13. D. Walter, Y. Ophir, K. H. Jamieson, *American Journal of Public Health* 110, 718 (2020).

14. D. A. Broniatowski, *et al.*, *American journal of public health* 108, 1378 (2018).

15. E. Ferrara, *First Monday* (2020).

16. E. Zuckerman, *Journal of Design and Science* 6, 1 (2019).

17. M. Miller, *Washington Post* (2021).

18. A. Robb, *Rolling Stone* (2017).
19. E. Welch, *New York Times* (2016).

20. K. Roose, *New York Times* (2021).

21. J. Tollefson, *Nature* **590**, 192 (2021).

22. C. Shao, *et al.*, *Plos one* **13**, e0196087 (2018).

23. N. G. d. Mesnards, D. S. Hunter, Z. el Hjouji, T. Zaman, *Operations Research* (2021).

24. D. S. Hunter, T. Zaman, *arXiv preprint arXiv:1806.11253* (2018).

25. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, *arXiv preprint arXiv:1810.04805* (2018).

26. I. Strauch, QAnon #Hashtags Megapost, [https://wt.social/post/fighting-misinformation/nvrqyhu5325591624484/history](https://wt.social/post/fighting-misinformation/nvrqyhu5325591624484/history) Accessed: 2020-02-16.

27. C. A. Davis, Botornot: A system to evaluate social bots (2016).

28. G. Pennycook, D. G. Rand, *Proceedings of the National Academy of Sciences* **116**, 2521 (2019).

29. G. Pennycook, *et al.*, *Nature* **592**, 590 (2021).

30. M. Mosleh, C. Martel, D. Eckles, D. Rand, *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–13.

31. A. Guess, J. Nagler, J. Tucker, *Science advances* **5**, eaau4586 (2019).

32. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, *Science* **363**, 374 (2019).
33. L. Hanu, contributors, Detoxify, https://github.com/unitaryai/detoxify (2020).

34. L. Vassio, F. Fagnani, P. Frasca, A. Ozdaglar, IEEE transactions on control of network systems 1, 109 (2014).

35. M. H. DeGroot, Journal of the American Statistical Association 69, 118 (1974).

36. M. Mobilia, Physical review letters 91, 028701 (2003).

37. T. Bella, The Washington Post (2020).

38. A. Vaswani, et al., Attention is all you need (2017).

39. Y. Zhu, et al., Aligning books and movies: Towards story-like visual explanations by watching movies and reading books (2015).

40. D. P. Kingma, J. Ba, arXiv preprint arXiv:1412.6980 (2014).

41. M. Moussaïd, J. E. Kämmer, P. P. Analytis, H. Neth, PloS one 8, e78433 (2013).