Measuring the Influence of Trollbots in Twitter during the 2018 US Midterm Election

Abstract

The use of troll bots in social media and its consequences has been a subject of great public interest. In this study, we measure the influence of trollbots during the 2018 US Midterm Elections. We selected 2.94M tweets and 870k profile data posted between October 31, 2018, and November 7 using nine keywords. On December 10, 2019, we rescraped Twitter to find suspended and removed accounts. We identified the suspended accounts as Trollbots and used Machine Learning on the removed accounts to find other trollbots. We found that the top 1% of the Trollbots, influenced the creation of 7% of all the tweets. In Twitter, Tweet or a retweet made by another user can be retweeted. We estimate the number of times the tweet or the retweet gets retweeted by other users to calculate the influence of a Tweet. We added the influence of all Tweets made by a user to calculate user wise influence.

1 Introduction

Twitter is a social networking service where users interact with each other using short messages called Tweet. Tweets spread through retweets or replies. For example, when Alice replies/retweets, a tweet made by Bob, Alice and Bob's followers see the tweet.

People use social media, at least in part, to form an opinion about lifestyle, health, politics, and purchase (Varol et al. 2017). To reach many audience, various ethical and unethical techniques have been devised. One such unethical technique is the use of trollbots. Here we defined Trollbots as the account which violates Twitter's terms of service. The most common reason for the suspension is spam and abusive behaviour (Twitter).

In this study, we rely on Twitter for most of the bot detection to focus on measuring the influence.

1.1 Summary and Findings

- 13.6% of the users in the datasets no longer existed. 4.1% had been suspended. Of the remaining 9.5%, 22% were bots. Overall, 6.19% of the users were bots. They made 12% of all the Tweets.
- On Average, trollbots were 53% more influential than humans
- Top 1% of the trollbots influenced the creation of 7% of all the retweets/replies.
- On average, a single URL was shared 26.48 times in 1.98 cascades. When a Tweet mentions a URL. Any further retweet belongs to the same cascade.
- For URLs that were shared more than ten times, the influence percentage of bots was 70% greater than their size.
- 74.18% of all interaction (retweets and replies) took place between humans, and 1.78% took place between bots. 12.04% was bot-human, and 12% was human-bot interaction.

2 Related Work

2.1 Social Media, Trolls and Bots

People are spending more and more time on social media. Globally, people spend 2 hours and 23 minutes on average in Social Media. 40% people use social media to stay up to date with news and current events (GlobalWebIndex 2019). 68% of American adults get their news from social media while 42% find news on social media to be mostly accurate (PewResearchCenter 2018).

12% of US adults get news from Twitter(PewResearchCenter 2018). Twitter is the number one platform for government leaders. 97% of all UN members have an official presence in Twitter (Twiplomacy 2018). Furthermore, Twitter API provides easy access to data. Consequently, Twitter has been widely studied (Rizoiu et al. 2018; Grinberg et al. 2019; Bovet and Makse 2019; Morstatter et al. 2018; Munger 2017; Gruzd and Roy 2014; Zannettou et al. 2019a; Howard, Kollanyi, and Woolley 2016).

Social Media have been used for positives like democratizing online discussion, organizing civil movements (González-Bailón, Borge-Holthoefer, and Moreno 2013), augumenting public health (Dredze 2012), forecasting (Asur and Huberman 2010; Nguyen, Shirai, and Velcin 2015; Liu et al. 2016), forming social connections (Ellison, Steinfield, and Lampe 2007), and for other greater goods (Moorhead et al. 2013; Househ, Borycki, and Kushniruk 2014). However, recently, more focus is being put on the negatives with high focus on disinformation and bots (Forelle et al. 2015; Bradshaw and Howard 2017; Marwick and Lewis 2017).

Different types of bots exist in social media. A simple bot will post predetermined messages at predetermined intervals (Haustein et al. 2016). Bots are also used to increase followers and make an account appear popular (Cresci et al. 2015). Here, Botnet refers to a network of social media bots working together to make an influence. Many botnets use a hybrid human/automation approach (Grimme, Assenmacher, and Adam 2018). Botnets have been used to promote spam (Ferrara 2018), which seem to shifted to social media due to the effectiveness of spam filter (Gao et al. 2010; Chu, Widjaja, and Wang 2012; Ferrara 2018), manipulate stocks (Ferrara 2015), manipulate elections (Morstatter et al. 2018), and for various other purposes (Abokhodair, Yoo, and McDonald 2015). Many of the suspended users are a part of Botnets. Botnets were detected as early as 2010 US Midterm Election (Mitter, Wagner, and Strohmaier 2014). Over time, their use and study have increased. Studies have detected and analyzed bots in the 2017 German Federal Election (Morstatter et al. 2018) ,2017 French Presidential Election (Ferrara 2017) and during various US Elections (Mitter, Wagner, and Strohmaier 2014; Bovet and Makse 2019; Rizoiu et al. 2018; Bessi and Ferrara 2016; Howard, Woolley, and Calo 2018; Howard, Kollanyi, and Woolley 2016; Deb et al. 2019). In Elections, they have been used to support candidates (Luceri et al. 2019), attack people (Mueller 2019) and spread fake news (Vosoughi, Roy, and Aral 2018; Grinberg et al. 2019). Political bots are active beyond the election time. (Stewart, Arif, and Starbird 2018) found that Russian bots infiltrated both right and left-leaning communities and spread different narratives (Mueller 2019).

In Twitter, the trollbots share content in multiple channels (Paul and Matthews 2016). This activity is in line with literature, which shows that information from multiple sources appears more trustworthy than a single one (Harkins and Petty 1981). Trollbots attack people. Attacking trustworthiness has been shown to diminish the credibility of the Original Poster (Pornpitakpan 2004). Bots are continuous, repetitive, but most of the time, they are also far from reality (Paul and Matthews 2016). The high volume seems to ensure early exposure and (Petty et al. 1994) shows that early first impression is more likely to be accepted by the brain. Due to a phenomenon called Sleeper Effect, where information gets disassociated from the source while remembering (Underwood and Pezdek 1998; Paul and Matthews 2016) low credibility sources can have the persuasive power to unbiased or neutral people. As a boost to sleeper effect, people are 31% more likely to remember what they see on twitter, compared to the normal web (Park 2018).

2.2 Trollbot Detection

Most studies (Rizoiu et al. 2018; Yang et al. 2019; Shao et al. 2018) use Botometer(Davis et al. 2016; Yang et al. 2019) for bot detection. Botometer's Machine Learning algorithm provides an account-level bot classification. Botometer classifies account partially and completely automatized as bots. Botometer uses more than a thousand features created from temporal activity, network structure, content analysis, sentiment analysis, and user profile data to determine a bot score (Davis et al. 2016; Yang et al. 2019). (Bessi and Ferrara 2016) illustrated that profile customization, geographical metadata, and activity statistics provided the strongest signals for bot detection. However, Botometer has high false positive and detects organizational accounts as bots (Varol et al. 2017) (Botometer 2019).

(Ferrara 2017) and (Kudugunta and Ferrara 2018) had success detecting bots using fewer features. (Ferrara 2017) obtained 93% accuracy with an AUC-ROC score of 92% in their best model using Random Forest classifier. (Kudugunta and Ferrara 2018) used 3,000 labelled examples to train a system with an AUC greater than 99% using Adaboost Classifier with Over Sampling and Undersampling of data using the SMOTENN algorithm.

2.3 Influence Detection

(Zannettou et al. 2019b) uses Hawkes Process to determine the influence Iranian and Russian bots had on pushing URLs in 4 social media platforms. They found that Russian trolls were extraordinarily influential and efficient in spreading URLs. (Shao et al. 2018) found that bots amplify URLs in early moments before an article goes viral, while (Ferrara 2018) found the same for spam. Bots target users with many followers through replies and manipulation. This method was very efficient.

(Shao et al. 2018) found that articles spread mostly through tweets and retweets. Their study showed that people do not discriminate between resources shared by humans and bots. (Shao et al. 2018) and (Grinberg et al. 2019) found super spreaders. Moreover, (Grinberg et al. 2019) found that 1% of individuals accounted for 80% of fake news exposure. (Varol et al. 2017) also found that 2% user accounts were responsible for 60% of the conversation.

(Rizoiu et al. 2018) found that social bots were 2.5 times more influential than humans. They used Botometer to label the bots and introduced a scalable algorithm for estimating user influence in retweet cascades. For each tweet in an information cascade, it uses time of post and the number of followers to determine the probability of whether a tweet is a retweet of another. They then find influence which can be used to find out how many users the tweet possibly influenced to retweet. It was tested successfully in artificial social media. Other studies have attempted to measure influence before this. (Weng et al. 2010) used eigenvector centrality of the connection to measure influence, but it is not scalable into big cascades. There are other methods like (Rodriguez, Balduzzi, and Schölkopf 2011; Cho et al. 2013; Linderman and Adams 2014), but they either have scaling issues or require a full diffusion graph, which Twitter does not provide.

3 Methods

3.1 Data Collection

Tweets containing nine manually identified keywords were selected from collected live data from Twitter. The keywords are:

- 2018Midterms
- Election

- Election2018
- 2018Senate
- WinBlue
- WinRed
- BlueWave
- RedWave
- Republican
- Democrat

Twitter provides Streaming API, Search API, and a premium Firehose API to give access to tweets. Streaming API provides limited access to live data, while Firehose API provides complete data. However, due to the associated costs, Firehose API was not an option in this study. (Morstatter et al. 2013) compares the Streaming API with Firehose API. They found that Streaming API was nearly as good as a random sample of Firehose API when the dataset was large enough. Although the 1% API was not as good as a 1% random sample from the Firehose API in all of their tests, it estimated the top hashtags correctly when the data was large enough.

In this study, we use Streaming API. While collecting data, information about the tweeter and the tweet was collected.

We collected following information about a tweet:

1 **Timestamp:** UTC Timestamp in which the post was made

2 **ID:** Post ID provided by twitter

3 **Text:** Text in the tweet and the parent tweet if the tweet is a reply

4 **User:** Username of the tweeter/retweeter/replier

5 **Replies:** The number of replies the tweet has received. During live collection, this value is zero. We collect them during recollection of top tweets.

6 **Retweets:** The number of retweets the tweet has received. During live collection, this value is zero. We collect them during recollection of top tweets.

7 **Likes:** The number of likes the tweet has received. During live collection, this value is zero. We collect them during recollection of top tweets.

8 **Reply To ID:** The ID of the parents tweet if this is retweet or a reply.

9 **Response Type:** Either Tweet or Retweet or Reply

We collected the following profile information from the Twitter API:

1 **Username:** Username of the user who posted the tweet

2 **Location:** Binary if geolocation is enabled

3 **Is Verified:** A binary if the profile has been verified

4 **Total Tweets:** Total number of tweets created by the user

5 **Total Following:** Total accounts the user is following

6 **Total Followers:** Total Followers

7 **Total Listed:** Number of times a user has been listed

8 **Total Status:** Total Status

9 **Total Likes:** Total likes the user has received

10 **Has Background:** Binary if an account has a background

11 **Is Protected:** Binary if an account is protected

12 **Profile Modified:** Binary if a profile has been modified

(Kudugunta and Ferrara 2018) and (Ferrara 2017) use the same parameters in their machine learning architecture.

We collected Tweets from October 31, 2018 (1 AM UTC) to November 7, 2018 (midnight UTC). On July 18, 2019, we used Twitter API to rescrape the top 100k most retweeted Tweets which did not exist in our dataset.

3.2 Data Processing

We used SentiStrength in the Tweets to compare the sentiment. SentiStrength (Thelwall et al. 2010) is used to annotate short, informal tweet like texts (Bessi and Ferrara 2016). It can capture positive and negative emotions at an accuracy of 60.6% and 72.8%. SentiStrength gives a positive and negative score between 0 and 4. The positive sentiment is subtracted from the negative like in (Bessi and Ferrara 2016) to get a whole number. SentiStrength is used instead of another popular solution VADER because it performed better in manual analysis.

3.3 Trollbot Detection

A year after the data collection, we checked twitter to find accounts that no longer existed. 117650 of the 863539 accounts did not. Of them, the 32323 suspended accounts were verified to be troll bots. The remaining 85327 could or could not be bots as they also include accounts deactivated by users. So, we used Machine Learning in this data. Initially, we had used Machine Learning on the complete dataset, but we received bad accuracy using that mechanism.

We selected 95% of suspended users and an equal amount of random non suspended users, which we conveniently refer to as humans to train a SMOTENN network using the profile features detailed above. It had an AUC of 0.95 in the training set and 0.93 in the test set.

3.4 Cascading and Influence Detection

A tweet cascade starts when a Tweet is made. Any retweet or response which involve that Tweet belongs to that cascade. If Bob retweets Alice's tweet, and Eve retweets Bob's retweet, the tweets made by Alice, Bob, and Eve will belong in the same cascade. The Twitter API will show Eve and Bob as a direct descendant of Alice without any mention that Eve retweeted Bob's retweet.

(Du et al. 2013) defines influence as the average number of users who get in contact with the content created by a user u. However, Twitter does not provide the diffusion graph and the number of people reached. (Rizoiu et al. 2018) define the influence of a user over a retweet cascade as "the expected number of time the tweet is retweeted – direct retweets or descendants in a diffusion scenario – over all possible diffusion scenarios associated." They define the influence of a user as the sum of the influence of tweets authored by a user. We use their algorithm and definition. (Rizoiu et al. 2018) uses time and number of followers to estimate a user influence. Using this mechanism, a high influence score can be provided to highly connected users who never start diffusions and to active retweeters with little followership. (Rizoiu et al. 2018) tested their algorithm on an artificial social network with 1000 users and found that the influence calculated had a Spearman correlation coefficient of 0.88 with the actual influence.

In (Rizoiu et al. 2018)'s algorithm, for each tweet in a cascade, the probability of it being a descendant of each previous tweet is calculated using a softmax function. Mathematically, probability that a j_{th} tweet is a retweet of i_{th} tweet is measured by :

$$p_{ij} = \frac{m_i e^{-r(t_j - t_i)}}{\sum_{k=1}^{j-1} m_k e^{-r(t_j - t_k)}}$$

where,
t_j-t_i is used as exponential decay between the timing of original tweet and that of the retweet,
r is a hyperparmeter which they found to be 6.8×10^{-4},
m is the number of followers

Then for every tweet in a cascade, pairwise influence is calculated as:

$$m_{ij} = \begin{cases} \sum_{k=i}^{j-1} m_{ik} p_{kj}^2 & ,\text{i} < \text{j} \\ 1 & ,\text{i} = \text{j} \\ 0 & ,\text{i} > \text{j} \end{cases} \quad (1)$$

Then the total influence of a node is the sum of the pairwise influence score m_{ij} over all subsequent nodes. For a derivation of this, the original study (Rizoiu et al. 2018) and its citations should be referred.

(Rizoiu et al. 2018) tested the validation of their algorithm when they had access to full data. As we used streaming API, we did not. So random cascades were synthesized using the following algorithm to test the validity:

- An aggregate database D was created. In the database D, values of D_u was set to all possible usernames and $D_u{}^a$ and $D_u{}^s$ set to 0.
- For 500000 cascades, following steps was repeated:
  - Length of cascade, n was randomly selected from the all possible size of cascades we captured. Then n random values were selected from our cascade that included username, time and followers count.
  - Influence was calculated for the selected n values. 1% of values were randomly selected from the n, and influence was calculated for it too.
  - The total sample influence, $D_u{}^s$ and actual influence $D_u{}^a$ in the dataset D was updated for a user u, by adding the calculated value with the current values of $D_u{}^s$ and $D_u{}^a$.

Then we converted the raw influence score to a user wise percentage. A correlation of 0.96 existed between the full dataset and the 1% sample. We used this method instead of (Rizoiu et al. 2018)'s original one due to limitations in computing power.

4 Data Analysis

4.1 Exploratory Data Analysis

Our dataset consisted of 2.94M Tweets by 870k users. After Machine Learning, 6.19% of these accounts were found to be bots. They made 15.6% of the Tweets. The mean sentiment of bots was -0.19, slightly more negative than the average non-bot at -0.16. We removed stopwords to compare the words used by bots and humans (used for convenience to refer to users who are not Trollbots). We selected words that appear more than 0.03% times in the dataset (which was 41386 times for humans and 6148 for bots). The dominant words are in 2 and 1. From Table 2, we can see that most of the trollbots seem to be alt-right or republican. Qanon is an alt-right conspiracy theory while kag refers to "Keep America Great". Other bot dominant words are also popular among the republicans. Among humans, one incident seems to have stood out. A Florida State University (FSU) student had been arrested for throwing chocolate milk and the word related to it dominate the human side.

Word	Percentage Difference
chocolate	78.885052
milk	76.540881
fsu	76.08353
throwing	75.873016
woman	59.52
arrested	58.247903
men	56.640345
jewish	55.259654
nazis	49.34877
name	46.192053
put	44.897959
away	43.264503
something	41.046607
angry	40.981013
think	40.675676

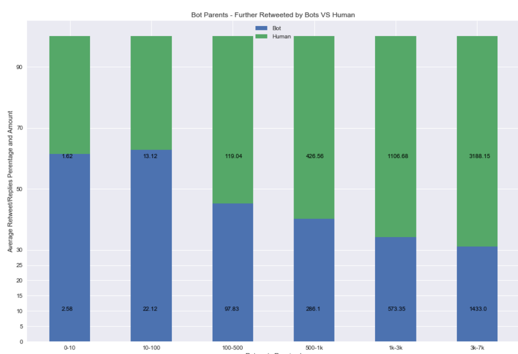
Table 1: Words used more often by humans

Then we analyzed the 141436 cascades containing more than ten tweets to calculate human-bot interaction. We made the following observations:

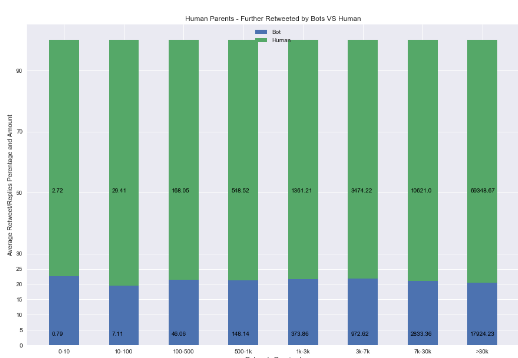
- 1.78% of all retweets were made from bot Tweets by bots.
- 12.04% of all retweets were made from bot Tweets by humans.
- 12% of all retweets were made from humans Tweets by bots.
- 74.18% of all retweets were made from human Tweets by humans.

Word	percentage_difference
qanon	32.315522
kag	31.95122
walkaway	28.093948
usa	27.511962
patriots	26.732673
racism	26.315789
lie	25.85034
americafirst	25.561798
wwg1wga	25.519288
voteredtosaveamerica	25.504152
low	24.262295
truth	24.08377
borders	23.773006
destroy	23.054755
lying	22.685185

Table 2: Words used more often by bots



(a) Bots Parents



(b) Humans Parents

Figure 1: Retweet cascades for tweets made by humans and bots are grouped by their size and visualized. Here, we see that bots dominate cascades started by bots in the early retweets. As the size gets bigger, more humans see it. For human started tweets, their percentage is similar in all sizes

Next, we compared the cascades started by bots and humans. Figure 1 can be zoomed for detail. In Figure 1a we can see that the percentage of bots was high for smaller cascades started by bots. As the size of the cascade increases, the percentage of bots decreases. We made the opposite observation for humans in Figure 1b. If a tweet receives few audiences, they are the followers. Bots are mostly followed by bots and humans by humans. This finding goes in line with those made by earlier research which shows that bots and humans are mostly followed by their type.

4.2 Retweet Influence Analysis

The average influence of a human was 4.12, while that of a trollbot was 6.34. This is lesser than (Rizoiu et al. 2018)'s finding who found that bots were 2.5 times more influential than humans. But they used Botometer to determine bots while we relied mostly on Twitter.

Figure 2 shows the most influential users.

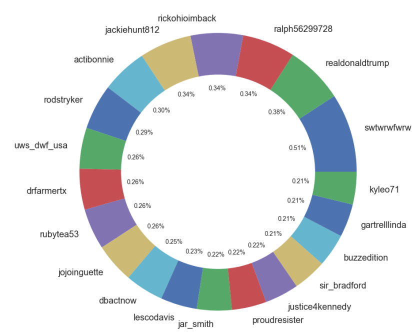


Figure 2: Most influential users. Users with many followers and users who made many tweets were found to be the most influential ones. Most interesting is user @swtwrwfwrw who made over 500 tweets during this timeframe and was found to be more influential than @DonaldJTrump

Then, we measure made by the percentage of users. To find this out, we first calculated the user wise influence of all users in our dataset as a sum of their influence in all their tweets. Then, we used our label bots and not bots, referred conveniently as humans. We calculated the Influence percentage by using the total influence of all users. Figure 3, shows that that in both categories top 1% of the users account for around 70% of their total influence

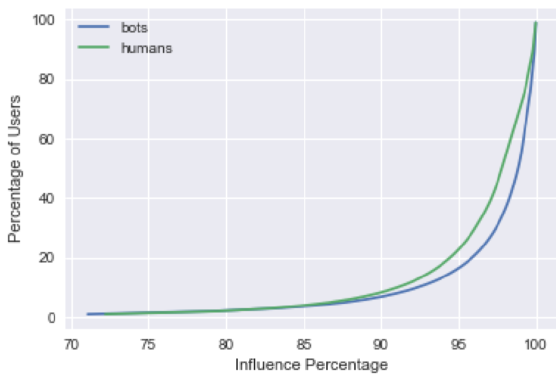


Figure 3: Percentage of top users and their influence percentage.

Bots are responsible for 10% of the total influence, while the top 1% is responsible for 7%. Out of 2.94M data in our cascade, this might imply that bots were responsible for the creation of 180k human tweets. However, we cannot make this conclusion without further research. When calculating influence, we could have calculated the influence of a bot or a human in the diffusion graph multiple times. We use the graph in 4 to illustrate the problem in making this assumption.

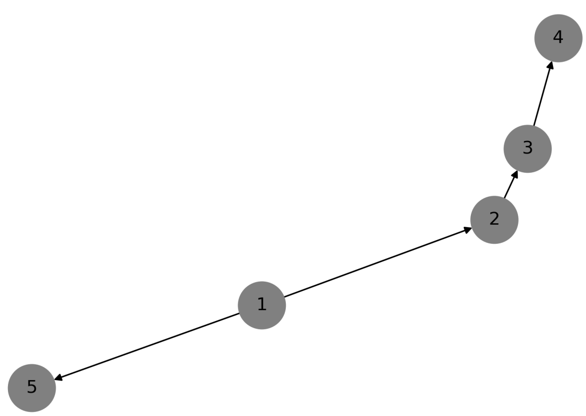


Figure 4: Diffusion Graph

In Figure 4, the influence score of node 1 will be 4, node 2 will be 2, of node 3 will be 1, and of node 4 and node 5 will be 0. If 1 and 2 are bots, the total bot influence score will be 6. However,1 created 2. Although this disparity might cancel out in 2 large groups across a large dataset when taken in percentage, we still refrain from making that conclusion. Further research is needed to make this verification. If we can make the verification, we can measure the impact of bots in terms of monetary value by comparing it with the advertising cost.

4.3 URL Influence Analysis

Before performing URL influence calculation, we resolved short URLs and removed twitter.com links. We kept the URLs which we could not resolve as they were. 217,350 unique URLs belonging to 53,378 unique domains were shared in our dataset by 210,102 unique users 0.6 Million times. Ten most popular domains in our dataset with their frequency are in Table 3. The top 10 domains for humans and bots, apart from t.co are in Table 4.

Domain Name	**Frequency**
t.co	127717
vote.gop	96127
thehill.com	57476
cnn.com	54129
youtube.com	51347
wtxl.com	38124
nytimes.com	35863
foxnews.com	35277
instagram.com	31634
breitbart.com	30290
pscp.tv	27702

Table 3: Most Common URLs

Human	**Bots**
vote.gop	vote.gop
thehill.com	youtube.com
cnn.com	foxnews.com
youtube.com	breitbart.com
wtxl.com	thegatewaypundit.com
nytimes.com	pscp.tv
foxnews.com	cnn.com
instagram.com	instagram.com
breitbart.com	thehill.com
pscp.tv	facebook.com

Table 4: Most Common Human and Bots URL

A single URL was shared 26.48 times in 1.98 cascades on average.

We found that the in URL cascades, for URLs that were shared more than ten times, the influ- ence percentage of bots was 70% greater than their size.

5 Conclusion, Limitations and Further Work

In this study, we show that bots have been highly influential and that most of the influence comes from a few top bots. In our dataset, 6.19% of the users were bots. They created 12% of the Tweets are their average influence was 53% more than that of a human. Overall they accounted for 10% of the total influence, and the top 1% were responsible for 7% of the influence. In line with previous research, we have also shown that bots and humans form a cluster around each other and join when a tweet gets successful. This knowledge can help to point out a possible direction in the fight against bots.

- We performed this study in a limited dataset. It would have been better to perform it in the complete data. However, we did not have access to it. Twitter Firehose API or Search API can be used in the future to have complete access to the data. Apart from Firehose and Search API, a mechanism for finding Twitter posts using Twitter's snowflake algorithm looks promising (Baumgartner 2019)
- The current analysis was done by considering all bots as a single entity. Although easy, this distinction has many flaws. In futures, attempts should be made to divide different types of bots and study their agenda
- As Twitter attracts some demographics more than others, using random Twitter data is not representative of the nation or the voting population. Careful clustering can be done to create a representative data to get representative influence in the future.
- The validation of the influence detection algorithm can be done in a better way. Even Machine Learning techniques can be used to make a better influence detection algorithm to predict the diffusion edges.
- Reddit can be analyzed along with Twitter to create a better spread graph. A bot detection system can be made for Reddit too. Combining multiple sources, a long term study can be conducted to measure the success/failure of bots in changing people's opinion.

References

Abokhodair, N.; Yoo, D.; and McDonald, D. W. 2015. Dissecting a social botnet: Growth, content and influence in twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 839–851. ACM.

Asur, S., and Huberman, B. A. 2010. Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, 492–499. IEEE Computer Society.

Baumgartner, J. 2019. Reconstructing twitter's firehose.

Bessi, A., and Ferrara, E. 2016. Social bots distort the 2016 us presidential election online discussion. *First Monday* 21(11-7).

Botometer. 2019.

Bovet, A., and Makse, H. A. 2019. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications* 10(1):7.

Bradshaw, S., and Howard, P. 2017. Troops, trolls and troublemakers: A global inventory of organized social media manipulation.

Cho, Y.-S.; Galstyan, A.; Brantingham, P. J.; and Tita, G. 2013. Latent self-exciting point process model for spatial-temporal networks. *arXiv preprint arXiv:1302.2671*.

Chu, Z.; Widjaja, I.; and Wang, H. 2012. Detecting social spam campaigns on twitter. In *International Conference on Applied Cryptography and Network Security*, 455–472. Springer.

Cresci, S.; Di Pietro, R.; Petrocchi, M.; Spognardi, A.; and Tesconi, M. 2015. Fame for sale: Efficient detection of fake twitter followers. *Decision Support Systems* 80:56–71.

Davis, C. A.; Varol, O.; Ferrara, E.; Flammini, A.; and Menczer, F. 2016. Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 273–274. International World Wide Web Conferences Steering Committee.

Deb, A.; Luceri, L.; Badaway, A.; and Ferrara, E. 2019. Perils and challenges of social media and election manipulation analysis: The 2018 us midterms. In *Companion Proceedings of The 2019 World Wide Web Conference*, 237–247. ACM.

Dredze, M. 2012. How social media will change public health. *IEEE Intelligent Systems* 27(4):81–84.

Du, N.; Song, L.; Rodriguez, M. G.; and Zha, H. 2013. Scalable influence estimation in continuous-time diffusion networks. In *Advances in neural information processing systems*, 3147–3155.

Ellison, N. B.; Steinfield, C.; and Lampe, C. 2007. The benefits of facebook "friends:" social capital and college students' use of online social network sites. *Journal of computer-mediated communication* 12(4):1143–1168.

Ferrara, E. 2015. Manipulation and abuse on social media by emilio ferrara with ching-man au yeung as coordinator. *ACM SIGWEB Newsletter* (Spring):4.

Ferrara, E. 2017. Disinformation and social bot operations in the run up to the 2017 french presidential election. *First Monday* 22(8).

Ferrara, E. 2018. Measuring social spam and the effect of bots on information diffusion in social media. In *Complex Spreading Phenomena in Social Systems*. Springer. 229–255.

Forelle, M.; Howard, P.; Monroy-Hernández, A.; and Savage, S. 2015. Political bots and the manipulation of public opinion in venezuela. *arXiv preprint arXiv:1507.07109*.

Gao, H.; Hu, J.; Wilson, C.; Li, Z.; Chen, Y.; and Zhao, B. Y. 2010. Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, 35–47. ACM.

GlobalWebIndex. 2019. 2019 q1 social flagship report.

González-Bailón, S.; Borge-Holthoefer, J.; and Moreno, Y. 2013. Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist* 57(7):943–965.

Grimme, C.; Assenmacher, D.; and Adam, L. 2018. Changing perspectives: Is it sufficient to detect social bots? In *International Conference on Social Computing and Social Media*, 445–461. Springer.

Grinberg, N.; Joseph, K.; Friedland, L.; Swire-Thompson, B.; and Lazer, D. 2019. Fake news on twitter during the 2016 us presidential election. *Science* 363(6425):374–378.

Gruzd, A., and Roy, J. 2014. Investigating political polarization on twitter: A canadian perspective. *Policy & Internet* 6(1):28–45.

Harkins, S. G., and Petty, R. E. 1981. The multiple source effect in persuasion: The effects of distraction. *Personality and Social Psychology Bulletin* 7(4):627–635.

Haustein, S.; Bowman, T. D.; Holmberg, K.; Tsou, A.; Sugimoto, C. R.; and Larivière, V. 2016. Tweets as impact indicators: Examining the implications of automated "bot" accounts on t witter. *Journal of the Association for Information Science and Technology* 67(1):232–238.

Househ, M.; Borycki, E.; and Kushniruk, A. 2014. Empowering patients through social media: the benefits and challenges. *Health informatics journal* 20(1):50–58.

Howard, P. N.; Kollanyi, B.; and Woolley, S. 2016. Bots and automation over twitter during the us election. *Computational Propaganda Project: Working Paper Series*.

Howard, P. N.; Woolley, S.; and Calo, R. 2018. Algorithms, bots, and political communication in the us 2016 election: The challenge of automated political communication for election law and administration. *Journal of information technology & politics* 15(2):81–93.

Kudugunta, S., and Ferrara, E. 2018. Deep neural networks for bot detection. *Information Sciences* 467:312–322.

Linderman, S., and Adams, R. 2014. Discovering latent network structure in point process data. In *International Conference on Machine Learning*, 1413–1421.

Liu, T.; Ding, X.; Chen, Y.; Chen, H.; and Guo, M. 2016. Predicting movie box-office revenues by exploiting large-scale social media content. *Multimedia Tools and Applications* 75(3):1509–1528.

Luceri, L.; Deb, A.; Badawy, A.; and Ferrara, E. 2019. Red bots do it better: Comparative analysis of social bot partisan behavior. In *Companion Proceedings of The 2019 World Wide Web Conference*, 1007–1012. ACM.

Marwick, A., and Lewis, R. 2017. Media manipulation and disinformation online. *New York: Data & Society Research Institute*.

Mitter, S.; Wagner, C.; and Strohmaier, M. 2014. A categorization scheme for socialbot attacks in online social networks. *arXiv preprint arXiv:1402.6288*.

Moorhead, S. A.; Hazlett, D. E.; Harrison, L.; Carroll, J. K.; Irwin, A.; and Hoving, C. 2013. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of medical Internet research* 15(4):e85.

Morstatter, F.; Pfeffer, J.; Liu, H.; and Carley, K. M. 2013. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In *Seventh international AAAI conference on weblogs and social media*.

Morstatter, F.; Shao, Y.; Galstyan, A.; and Karunasekera, S. 2018. From alt-right to alt-rechts: Twitter analysis of the 2017 german federal election. In *Companion Proceedings of the The Web Conference 2018*, 621–628. International World Wide Web Conferences Steering Committee.

Mueller, R. S. 2019. Report on the investigation into russian interference in the 2016 presidential election.

Munger, K. 2017. Don't@ me: Experimentally reducing partisan incivility on twitter. *Unpublished manuscript*.

Nguyen, T. H.; Shirai, K.; and Velcin, J. 2015. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* 42(24):9603–9611.

Park, M. 2018. Expanding in-stream video ads to more advertisers globally.

Paul, C., and Matthews, M. 2016. The russian "firehose of falsehood" propaganda model. *Rand Corporation* 2–7.

Petty, R.; Cacioppo, J.; Strathman, A.; and Priester, J. 1994. To think or not to think? exploring two routes to persuasion in: Shavitt s, brock tc, editors. persuasion: Psychological insights and perspectives.

PewResearchCenter. 2018. News use across social media platforms 2018.

Pornpitakpan, C. 2004. The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of applied social psychology* 34(2):243–281.

Rizoiu, M.-A.; Graham, T.; Zhang, R.; Zhang, Y.; Ackland, R.; and Xie, L. 2018. # debatenight: The role and influence of socialbots on twitter during the 1st 2016 us presidential debate. In *Twelfth International AAAI Conference on Web and Social Media*.

Rodriguez, M. G.; Balduzzi, D.; and Schölkopf, B. 2011. Uncovering the temporal dynamics of diffusion networks. *arXiv preprint arXiv:1105.0697*.

Shao, C.; Ciampaglia, G. L.; Varol, O.; Yang, K.-C.; Flammini, A.; and Menczer, F. 2018. The spread of low-credibility content by social bots. *Nature communications* 9(1):4787.

Stewart, L. G.; Arif, A.; and Starbird, K. 2018. Examining trolls and polarization with a retweet network. In *Proc. ACM WSDM, Workshop on Misinformation and Misbehavior Mining on the Web*.

Thelwall, M.; Buckley, K.; Paltoglou, G.; Cai, D.; and Kappas, A. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61(12):2544–2558.

Twiplomacy. 2018. Twiplomacy study 2018.

Twitter. About suspended accounts.

Underwood, J., and Pezdek, K. 1998. Memory suggestibility as an example of the sleeper effect. *Psychonomic Bulletin & Review* 5(3):449–453.

Varol, O.; Ferrara, E.; Menczer, F.; and Flammini, A. 2017. Early detection of promoted campaigns on social media. *EPJ Data Science* 6(1):13.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science* 359(6380):1146–1151.

Weng, J.; Lim, E.-P.; Jiang, J.; and He, Q. 2010. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, 261–270. ACM.

Yang, K.-C.; Varol, O.; Davis, C. A.; Ferrara, E.; Flammini, A.; and Menczer, F. 2019. Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies* 1(1):48–61.

Zannettou, S.; Bradlyn, B.; De Cristofaro, E.; Stringhini, G.; and Blackburn, J. 2019a. Characterizing the use of images by state-sponsored troll accounts on twitter. *arXiv preprint arXiv:1901.05997*.

Zannettou, S.; Caulfield, T.; Setzer, W.; Sirivianos, M.; Stringhini, G.; and Blackburn, J. 2019b. Who let the trolls out?: Towards understanding state-sponsored trolls. In *Proceedings of the 10th ACM Conference on Web Science*, 353–362. ACM.