# Evolution of bot and human behavior during elections

Luca Luceri[1,2,3], Ashok Deb[1], Silvia Giordano[2] and Emilio Ferrara[1*]

[1] University of Southern California, Information Sciences Institute, Marina del Rey, California

[2] University of Applied Sciences and Arts of Southern Switzerland, Manno, Switzerland

[3] University of Bern, Bern, Switzerland

luca.luceri@supsi.ch, ashok@isi.edu, silvia.giordano@supsi.ch, emiliofe@usc.edu

*To whom correspondence should be addressed; E-mail: emiliofe@usc.edu

## Abstract

Online social media have become the main communication medium for political discussion. The online ecosystem, however, does not only include human users but has given a space to an increasing number of automated accounts, referred to as bots, extensively used to spread messages and manipulate the narratives others are exposed to. Although social media service providers put increasing efforts to protect their platforms, malicious bot accounts continuously evolve to escape detection. In this work, we monitored the activity of almost 245K accounts engaged in the Twitter political discussion during the last two US voting events. We identified approximately 31K bots and characterized their activity in contrast with humans. We show that, in the 2018 midterms, bots changed the volume and the temporal dynamics of their online activity to better mimic humans and avoid detection. Our findings highlight the mutable nature of bots and illustrate the challenges to forecast their evolution.

## Introduction

The complexity of today's online ecosystem captures the facets of modern information society: Accessing the news, sharing opinions, and entertaining social connections are just a few examples of the variety of engagements that individuals regularly perform online. Political communication is no exception, as it has moved from the world to the digital one. However, various risks (still largely unquantified) may associate with this change of norms, which in turn may seriously impact the real world. Malicious actors, e.g., foreign agents or scammers (Ferrara, 2019), may embed themselves in online social systems and interact with social network users with the objective of deceiving them and manipulating the public opinion.

In the political context, the 2016 Brexit referendum and the 2016 US Presidential election represent recent remarkable examples of social media political manipulation (Bessi & Ferrara, 2016; Howard *et al.*, 2016; Del Vicario *et al.*, 2017; Howard & Kollanyi, 2017; Allcott & Gentzkow, 2017; Woolley & Guilbeault, 2017; Badawy *et al.*, 2018a; Addawood *et al.,* 2019; Badawy *et al.*, 2019). Since then, service providers have been increasing their efforts to suspend malicious actors and maintain a healthy conversation on their platforms. However, nefarious activity on social media has not entirely stopped: social media bots (in short, *bots*), automated and software-controlled accounts (Ferrara *et al.*, 2016a), and trolls, human operators often associated with foreign entities, are still active online (Badawy *et al.*, 2018a; Badawy *et al.*, 2019; Luceri *et al.*, 2019; Im *et al.*, 2019). These play a pivotal role in information and disinformation campaigns globally (Ratkiewicz *et al.*, 2011; Metaxas and Mustafaraj, 2012; Ferrara, 2017; Howard *et al.*, 2017; Shu *et al.*, 2017; Badawy

*et al.*, 2018b; Vosoughi *et al.*, 2018; Stella *et al.*, 2018; Bovet and Makse, 2019; Gringberg *et al.*, 2019; Scheufele & Krause, 2019; Stella *et al.*, 2019, *Ruck et al.*, 2019). Detection of coordinated campaigns is an open challenge for the research community (Ferrara *et al*., 2016b; Varol *et al*., 2017a; Chen & Subramanian, 2018).

In particular, bots, due to their scalable nature, represent a major concern in the fight against media manipulation (Varol *et al.*, 2017b; Monsted *et al.*, 2017; Boichak *et al.*, 2018; Shao *et al.*, 2018; Yang *et al.*, 2019), as further demonstrated by the recent suspension of millions of compromised accounts by Facebook [1] and Twitter [2]. On July 1st, 2019, California became the first State to attempt regulating the usage of bots by requiring a self-disclosure of automated accounts that *intended to impersonate or replicate human activity on social media* (cf., *SB-1001 Bots: disclosure* [3]). In such a scenario, understanding how human users deal with these automated accounts and manipulation attempts is of paramount importance. On the other hand, detecting and keeping the pace of increasingly sophisticated malicious accounts is needed to build and adapt effective countermeasures.

Here, we show how bots' and humans' online activity has mutated over the last two US voting events, i.e., the 2016 Presidential election and the 2018 midterms. To this end, we captured the associated online political discussion on Twitter and monitored a set of about 245K accounts that were substantially active in both the events. Approximately 31K accounts were labeled as likely bots by *Botometer* (Davis *et al.*, 2016; Varol *et al.*, 2017b), a machine learning tool for bot detection on Twitter. We consider the overlapping set of users present in both the circumstances to perform a

comparative analysis between the two election periods. The rationale is to analyze the evolution of human and bot activity, behavior, and interplay.

We compared the temporal activity of bots and human users, recognizing differences in their trends during the 2016 election and similar patterns during the 2018 midterms, which might suggest that bots have been refined to better emulate humans and to avoid detection. We analyzed the volume of each sharing activity (i.e., tweet, retweet, reply, and mention) in the two election periods. We noticed a significant drop in the volume of retweets generated from both human and bot accounts in the 2018 election. On the one hand, humans doubled the usage of replies in the midterms with respect to the 2016 Presidential election. On the other hand, bots' propensity to repeatedly share the same content diminished during the 2018 midterms, fostering a coordinated (multi-bot) strategy, possibly to create an illusion of a consensus. While during the 2016 election bots tried to sow division around various issues (Bessi & Ferrara, 2016), here we show that in the 2018 midterms, bots aimed at surveying human preferences and intents.

Our work conveys two findings: On the one hand, the growing usage of replies, along with the decreasing (low-cost) content re-sharing, indicates an increasing propensity of human users to discuss their ideas instead of only re-sharing others' content. This represents an encouraging step forward in the quest against misinformation spread. On the other hand, the evolving nature of bots and their mutable intents pose novel computational challenges in regard to their modeling and detection.

## Methodology

*Data collection*: We captured the political discussion on Twitter by gathering election related posts (tweets) during the two election periods. We employed the Python module Twython to collect tweets through the Twitter API using a set of keywords as a filter. Keywords were selected ad hoc per each election, as described next. For the 2016 US Presidential Election, 23 keywords (see Supplement for the list of keywords) were used to collect tweets from September 16, 2016 to October 21, 2016, as detailed in (Bessi & Ferrara, 2016). Overall, 42.1 million tweets generated from 5.9 million users were gathered. For the 2018 US Midterms, tweets were collected from October 6, 2018 to November 19, 2018 using the following keywords: *2018midtermelections*, *2018midterms*, *elections*, *midterm*, and *midtermelections*. As a result, we obtained 2.6 million tweets from 997,406 users.

*Data processing*: In this study, we take into account only the users present in both datasets. Thereby, we consider the 278,181 accounts that tweeted both in 2016 and 2018. This subset of users represent a continuum between the two election conversations, other than the core of the online discussion. In fact, these users were involved (as authors of tweets, or as retweeted/replied/mentioned users) in 54 percent and 65 percent of the tweets collected in 2016 and 2018, respectively.
To examine the same time window for the two voting events, data from the two datasets have been filtered considering only tweets ranging from the month prior to the day following the election. The 2016 US Presidential election occurred on November 8, 2016. The 2018 midterms occurred on November 6, 2018. The filtering

results in 8,383,611 tweets from 2016, and 660,296 from 2018, originating in total from 245K users.

*Bot Detection*: One of the most important tasks for the uncovering and understanding of social media manipulation is the identification of automated accounts, i.e., bots. Researchers brought to the table a variety of solutions for the detection of bots (Chavoshi *et al.*, 2016; Subrahmanian *et al.*, 2016; Chen & Subramanian, 2018). While increasingly sophisticated techniques keep emerging (Kudugunta & Ferrara, 2018), in this study, we rely upon *Botometer* [4], a publicly-available machine learning-based tool maintained by Indiana University (Davis *et al.*, 2016; Varol *et al.*, 2017b; Yang *et al*., 2019) to detect automated accounts on Twitter. It is based on an ensemble classifier that aims to provide an indicator, namely *bot score*, used to classify an account either as a bot or as a human. To feed the classifier, the Botometer API extracts about 1,200 features related to the Twitter account under analysis. These features fall in six broad categories and characterize the account's profile, friends, social network, temporal activity patterns, language, and sentiment.

Botometer outputs a bot score: the lower the score, the higher the probability that the user is human. Prior studies used the 0.5 threshold to separate humans from bots. However, according to the re-calibration introduced in Botometer *v3* (Yang *et al*., 2019), along with the emergence of increasingly more sophisticated bots, we here used a bot score threshold to 0.3 (i.e., a user is labeled as a bot if the score is above 0.3). This threshold corresponds to the same level of algorithmic sensitivity of a score equal to 0.5 in prior versions of Botometer (cf. Fig. 5 from (Yang *et al*., 2019)).

We scored the 244,699 accounts on January 2019. The bot score distribution is displayed in Figure 1. Most of the accounts (about 80 percent) are below the knee of the curve, which is approximately around the bot score value of 0.2. This suggests that the value of 0.3, chosen as bot score threshold, represents a conservative choice as no significant difference would occur in the classification by further increasing the threshold. As a result, over the scored accounts, 12.6 percent were classified as bots, 86.1 percent as humans, while the remaining 1.3 percent of the accounts were not found on Twitter (indicating users that have deleted their account, have been suspended for violation of the Twitter rules, or have been quarantined by Twitter for further verification). The percentage of discovered bots (12.6 percent) represent a consistent result with respect to previous studies, e.g., the analysis of the 2016 US Presidential election (Bessi & Ferrara, 2016).

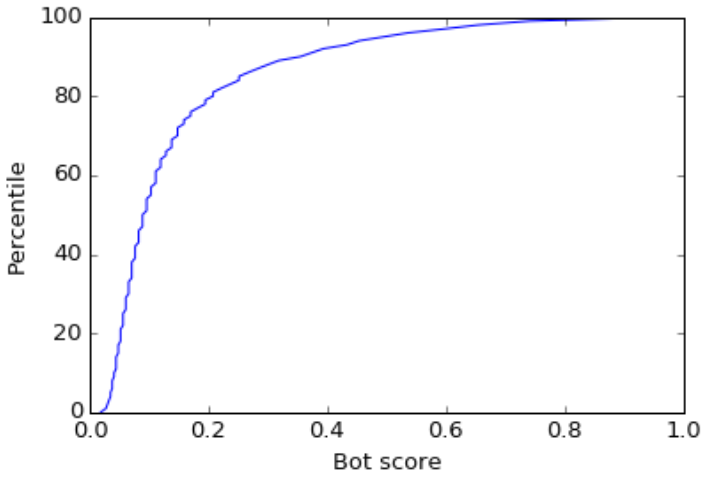


***Figure 1:*** *Bot score distribution of the 244,699 accounts scored with Botometer [4]*

*Sentiment Analysis*: To characterize the emotional content generated by both humans and bots, we rely upon sentiment analysis. More specifically, we employ

SentiStrength (Thelwall & Paltoglou, 2010) to map each tweet to the sentiment it expresses. SentiStrength is a lexicon-based approach that is conceived for social media text analysis. Lexicon-based algorithms are based on sentiment lexicons, dictionary of emotions where words are attributed a given sentiment strength. SentiStrength attributes a numerical score of sentiment intensity. In particular, it returns separately both a positive and negative score, which range from 1 to 5 (with 5 being the greatest strength). Overall, we are interested in the total sentiment, thus, we subtract the negative sentiment from the positive one. Thereby, the final score ranges from -4 (most negative) to 4 (most positive).

*Granger Causality*: To investigate the interplay between humans and bots, we evaluate whether human interaction with bots can be predicted by leveraging the volume of bots activity. To this end, we use the Granger causality test (Granger, 1969) on the time series representing the volume of shared content of these two classes of users.

More in general, Granger causality is used to determine whether time series X can be used to predict time series Y. Granger causality postulates that X "Granger-causes" Y if the predictions of future values of Y based on the combination of the past values of X with the past values of Y are better than the predictions of Y based only on the past values of Y. This holds true unless also the reverse (Y Granger-causes X) is verified. In such a case, no conclusion can be drawn. We further apply a differentiation to remove seasonal effects and, then, we tested the stationary time series. The autoregression of Y is augmented by lagged values of X

and those individually significant (t-statistic) that increase the explanatory power of the regression (F-test).

## Results

In this section we present our analysis of temporal dynamics of human and bot activity, human and bot behavior, and causal modeling of human-bot interplay dynamics.

*Temporal dynamics*: We explore bot and human dynamics by measuring the time lag between consecutive sharing activities (tweets). Figure 2 displays the inter-time tweet distribution comparing bot and human users in the 2016 and 2018 elections. Notably, in the 2016 election, the distribution of bots' tweet activity largely differs from the humans' distribution [5]. The discrepancy is particularly relevant in the time range between 10 minutes and 3 hours, consistent with other findings (Pozzana & Ferrara, 2018): in 2016, bots shared content at a higher rate with respect to human users.

On the other hand, in the 2018 midterms, inter-time distributions are similar, suggesting the possibility that bots have been refined to emulate human timing.

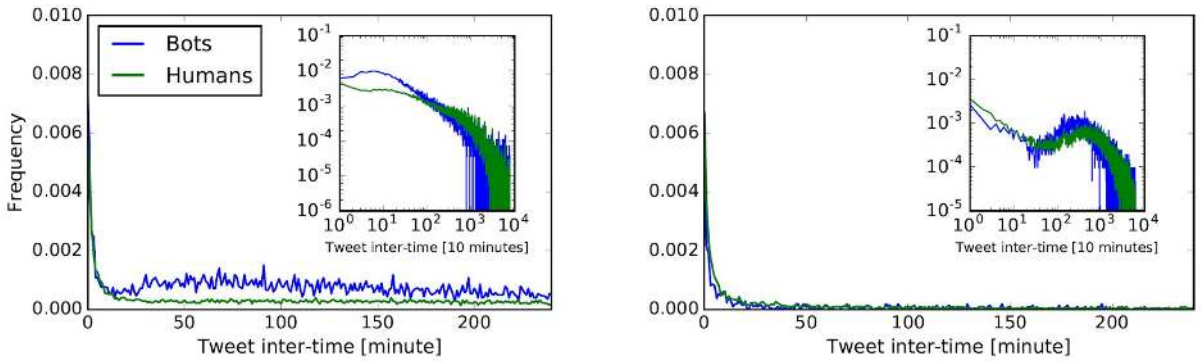


***Figure 2:*** *Posting inter-time for bots and humans in 2016 (left) and 2018 (right).*

To better understand the impact of each type of sharing activity on this result, we disentangle Twitter posts in tweets (original content generated by users), retweets (re-share of the original content generated by other users), replies (response to a post), and mentions (involvement of other users in a post). Figure 3 depicts the inter-time distribution of the above sharing activities in the 2016 election. Notice that we do not show the corresponding plots for the 2018 midterms as no relevant discrepancy can be appreciated between each of these sharing activities and the overall distribution in Figure 2 (right). From Figure 3, it is noticeable that the inter-time distributions of tweet and retweet present the two principal gaps between humans and bots in 2016. This finding is in line with the established bots' strategy consisting in overwhelming online platforms with high volume of tweets and retweets, as shown in previous studies (Ferrara *et al.*, 2016a; Bessi & Ferrara, 2016).

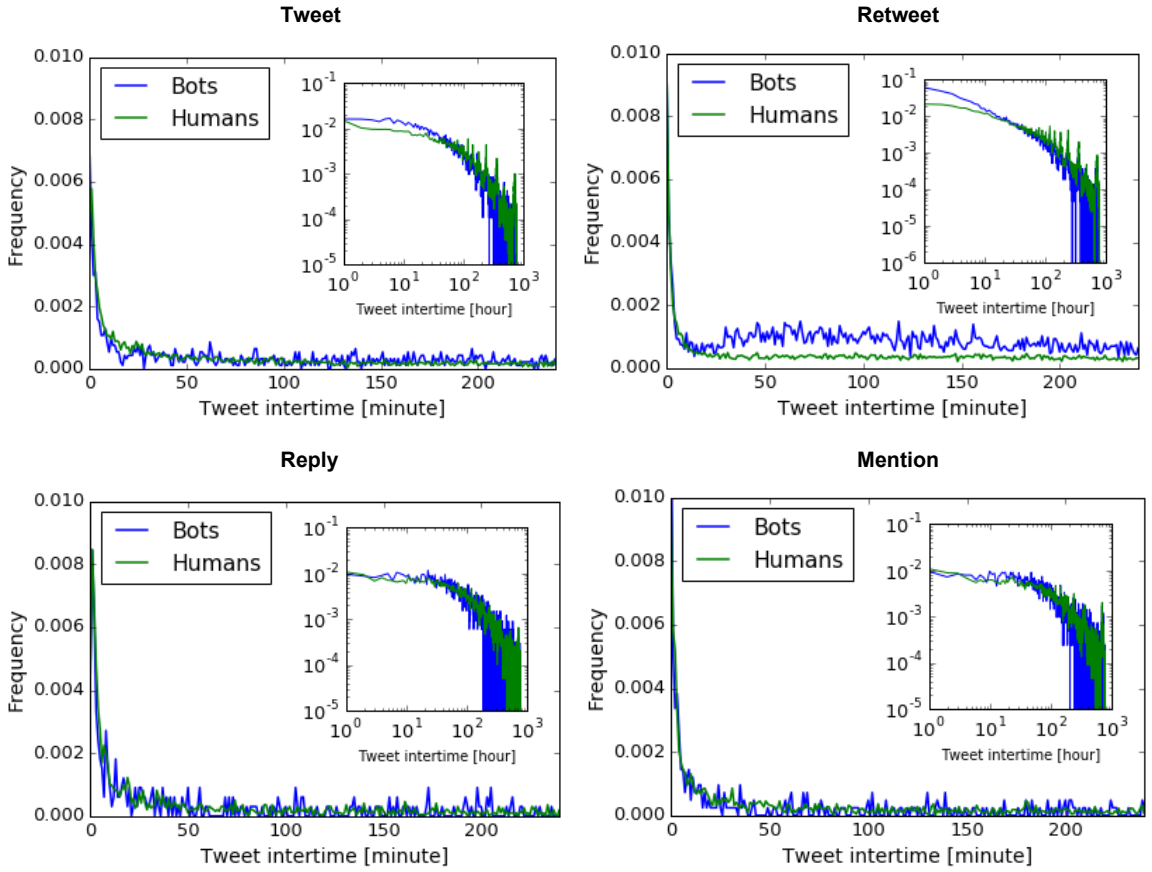


***Figure 3:*** *Consecutive tweet (top left), retweet (top right), reply (bottom left), and mention (bottom right) inter-time for bots and humans in 2016.*

*Activity Volume:* To further investigate the nature of the difference in the temporal dynamics, we measure the volume of each sharing activity in the two election periods. In Table 1, we show the percentage of each activity over all the posts for bots and humans. It can be noticed that both humans and bots significantly diminished the amount of retweets in the 2018 Midterm (t-test results: $t(5{,}840{,}537)=64.6$, $p<.001$ for humans and $t(3{,}083{,}630)=42.4$, $p<.001$ for bots), while the percent of mentions does not exhibit a noticeable variation in both groups of account. Additionally, humans have doubled the amount of replies (t-test: $t(5{,}840{,}537)=152.3$, $p<.001$) in 2018 and bots generated more original tweets (t-test: $t(3{,}083{,}630)=29.8$, $p<.001$) in 2018 than in 2016.

The growing propensity of humans to discuss a post (either positively or negatively) instead of simply re-sharing the content generated by other accounts, represents an encouraging finding. This insight acquires even more significance considering the cost related to each form of interaction. While retweeting is a one-click operation, with a relatively small human cost in terms of time and effort, a reply requires a larger undertaking. From the perspective of bots, a retweet can be programmatically executed with one command line; however, programmatically composing a meaningful reply requires the use of sophisticated natural language models, such as those based on deep neural networks (Radford *et al.,* 2019), which often require significant computing resources for training.

Interestingly, although bots increased the amount of tweets (with respect to the 2016 election) and, comparably to humans, reduced the amount of retweets, the gap in the inter-time distribution in 2018 appears to be reduced, suggesting a more cautious broadcasting strategy adopted by bots.

**Humans**

| Sharing activity | 2016 | 2018 |
|---|---|---|
| Tweet | 0.16 | 0.15 |
| Retweet | 0.76 | 0.72 |
| Reply | 0.04 | 0.09 |
| Mention | 0.04 | 0.04 |

**Bots**

| Sharing activity | 2016 | 2018 |
|---|---|---|
| Tweet | 0.12 | 0.15 |
| Retweet | 0.83 | 0.79 |
| Reply | 0.02 | 0.03 |
| Mention | 0.03 | 0.03 |

***Table 1**. Volume (in percent) of sharing activities of humans (left) and bots (right) in the two election periods*

To shed light on this result and, more in general, to explore the evolution of the actions of both humans and bots on Twitter, we now analyze separately the three sharing activities that showed more variation between the two election periods, i.e., tweet, retweet, and reply.

*Tweet Activity*: As far as original tweets are concerned, we found an anomalous pattern in the sharing activity of bots. Both in 2016 and 2018, bots tweeted multiple times the content they generated. While in 2016 this repeated activity was mainly performed by each bot separately, i.e., each message was published multiple times by the same bot, in 2018 this repeated activity was also shared between bots, i.e., multiple bots shared the same message once. More specifically, during the 2016 (resp. 2018) election period, 5 percent (resp. 2.4 percent) of the tweets generated by bots were published more than once by the same author. The significant drop (t-test: $t(388,196)=17.4$, $p<.001$) in this repeated activity during the 2018 election is, however, replaced by an increasing amount of single sharing operations (of the same content) performed by multiple bots, possibly in the context of a coordinated effort.

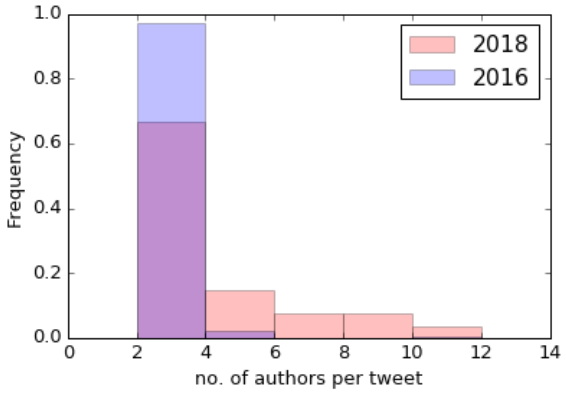


***Figure 4**: Probability distribution of multiple authors sharing the same content during the 2016 and 2018 election periods.*

In Figure 4, we investigate this scenario by showing the probability distribution related to the number of multiple authors sharing the same content in 2016 and 2018. Although the majority of content is shared by a few bots (from 2 to 3) in both periods under consideration, it is noticeable how in the 2018 election the number of authors sharing the same content increased. To evaluate the statistical difference between the sets of multiple authors in 2016 and 2018, we employ the Mann-Whitney rank test, which in turn corroborates our intuition ($p\text{-value} < 0.001$). We hypothesize that the distributed activity of bots can be a strategy to elicit the illusion of a consensus and, possibly, to avoid detection.

*Reply Activity*: We investigate whether the same repeated activity is also recognizable in the replies provided by bots to other accounts' posts. We found some instances of multiple replies in both the periods under investigation, but in both cases with a limited extent (around 1.1 percent of replies were repeated).

To characterize the emotional content of the replies generated by bots and humans, we rely upon sentiment analysis and in particular on SentiStrength. In Figure 5, we show the sentiment distribution of bots' (top) and humans' (bottom) replies (left plots), in contrast with the sentiment of their original tweet (right plots).

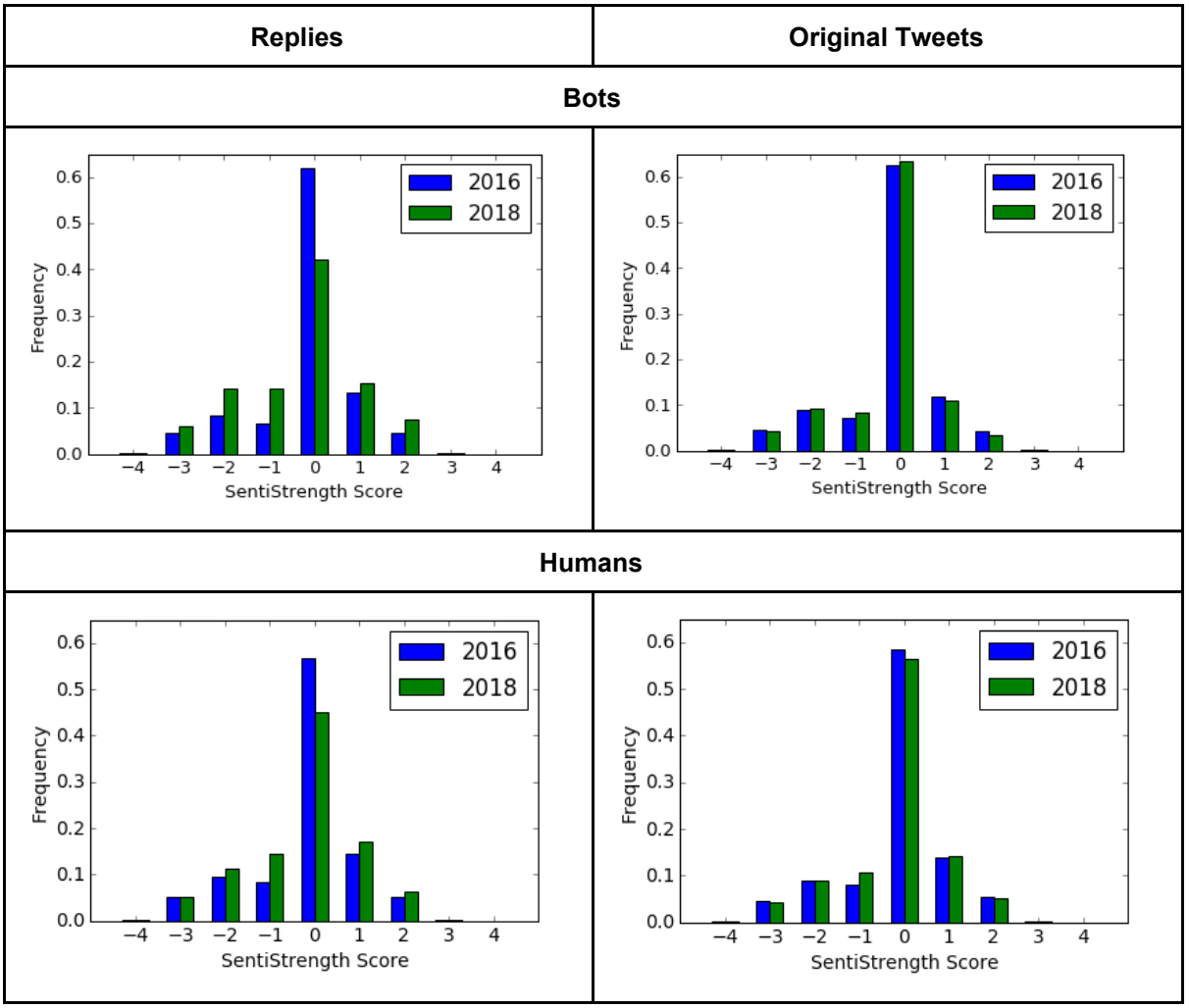


***Figure 5**: Sentiment of bots' (top row) and humans' (bottom row) replies (left column) and original tweets (right column)*

Two facts are worth noting. First, both humans and bots shared less neutral (SentiStrength score = 0) replies in 2018. The sentiment gap between the two election periods is more evident for bots, especially for negative replies, with the

2016 period exhibiting significantly more negative replies than 2018. With respect to original tweets, no remarkable difference exists between 2016 and 2018 regardless of the source (human or bot). Overall, the difference between the sentiment of replies and original tweets is only noticeable in the 2018 election period, while the distributions are similar for the 2016 election.

*Retweet Activity*: To examine the re-sharing activity in the two voting periods, we initially consider the top retweeted posts created by bots and shared by humans. By analyzing the top 10 retweets in this subset, we notice an interesting difference between the two periods. In the 2016 period prior to the election, the majority of the retweeted posts were in support of candidate Trump, and in opposition to candidate Clinton. On the other hand, in the midterms, the most retweeted posts of bot-generated content are polls on the Twitter platform. From now on, we refer to these tweets to as *poll-tweets*. In Table 2, we show the poll-tweets in the top 10 of human retweets of bot-generated content. The objective of these tweets seems to have a quantitative understanding of voter turnout, political leaning, and preferences. This exploratory attitude of bots in the midterms election appears in contrast with the behavior towards the candidates shown in the 2016 Presidential election.

Although poll-tweets seem harmless and aimed only at surveying human opinion, their turnout might impact the human perception on the polled issues. To understand whether bots fostered the spread of poll-tweets, we analyzed the most retweeted content by bots. Results show that the most re-shared post is a poll-tweet (cf., Table 2), and 4 additional poll-tweets are in the top 10 retweeted content by bots. In Figure

6, we depict the timeline of the hourly volume of retweets shared by bots and humans related to the poll-tweets listed in Table 2. Notice that we do not display the volume of retweets in the days after October 21 as the amount of these tweets during that period is negligible. Interestingly, human users participated in these polls to a larger extent with respect to bots, which in turn were responsible for the creation and sharing of these posts.

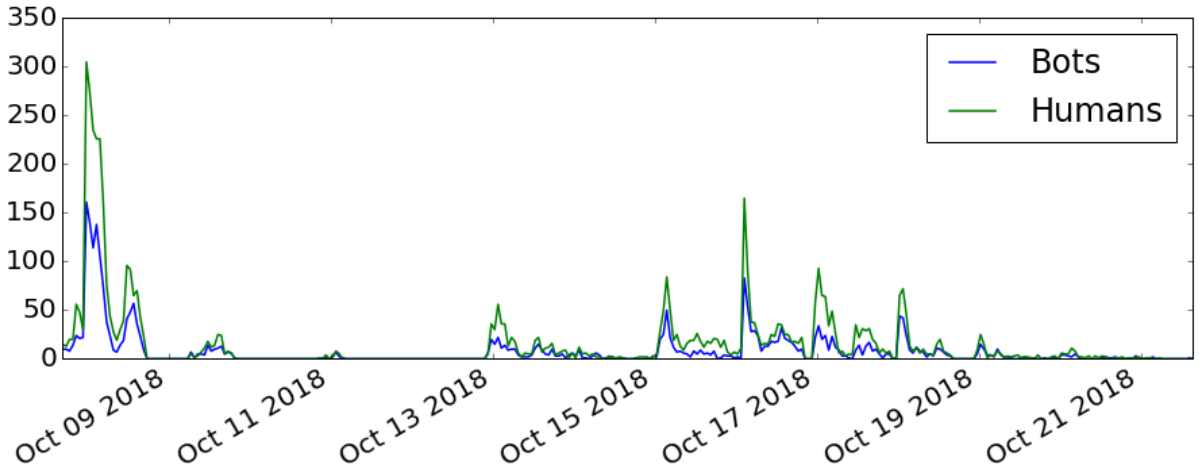


***Figure 6**: Timeline of the retweet volume of the poll-tweets shared by bots and humans*

Additionally, to investigate whether bots' retweet activity predicts humans participation in these polls, we evaluate whether the hourly volume of poll-tweets shared by bots "Granger-causes" the hourly volume of poll-tweets shared by humans. Results of the Granger-causality test, displayed in Table 2 along with the poll-tweet content, corroborate this hypothesis. In fact, for each poll-tweet under investigation, the volume of retweets from bot is Granger-cause of the volume of human retweet.

| Poll-tweets | F-test | p-value | steps |
|---|---|---|---|
| @kwilli1046: *Which Party Do You Plan To Vote For In The 2018 Midterm Election? Please vote and retweet for bigger sample size* | 22.8 | <0.001 | 1 |
| @The_Trump_Train: *RT if you agree: We need ICE agents at every polling station during elections.* | 15.2 | <0.001 | 1 |
| @???: *With all what's going on, if elections is today, would you vote for @realDonaldTrump? Please vote and Retweet.* | 15.1 | <0.001 | 3 |
| @kwilli1046: *Should voters in federal elections be required to show ID at the polls? Please vote and retweet for bigger sample size* | 2.9 | 0.01 | 6 |
| @???: *IF, AND ONLY IF, YOU ARE VOTING ON NOV 6, 2018, please answer this poll.* | 14.3 | <0.001 | 2 |
| @Golfman072: *As of Sept. 30th-Oct 5th how likely are you to vote in the MID-TERM elections? #Redwavepolls* | 20.7 | <0.001 | 1 |
| @Golfman072: *As of Oct 7th-13th how likely are you to vote in the MID-TERM elections? #Redwavepolls* | 4.5 | 0.03 | 1 |

***Table 2**. The 7 poll-tweets within the top 10 human retweets of bot-generated content. The first column displays the content of the poll-tweet; the second and third column show the value of F-test for significant causality and the corresponding p-value, respectively; the fourth column depicts the temporal lag (in hours) for which the causality has the highest F-test value. Note: we show the user names of suspended accounts; however, for privacy reasons, we anonymized (using @???) the user names of active accounts associated with the originators of these examples.*

Interestingly, three accounts that created the most successful polls (@kwilli1046, @The_Trump_Train, and @Golfman072) have been suspended by Twitter.

Additionally, another suspended account classified as human (@MikeTokes) published a poll-tweet ("*NATIONAL POLL: you are voting in the November 6, 2018 elections, what party are you voting for and why?*") that received a large amount of retweets from bots (top 3 of bots-retweet from human generated content). This may indicate a combined approach leveraging both human and bot activities.

To further explore the retweet interactions from bots to humans (i.e., bots retweeting human content), we measure to what extent bots targeted humans within the social network. Prior work shows that bots target the most connected humans (Stella *et al.*, 2018). In particular, the number of incoming edges (e.g., followers) is often associated with the influence and the centrality that each user has in the social network. To this end, here, we compute the in-degree centrality [6] of the retweet network. The rationale is to understand whether bots interacted mainly with the most retweeted (influential) humans, whose endorsement can be beneficial to spread information across the social network. This, in turn, might indicate a strategy for increasing the resonance of bots' content. The nodes of the retweet network are the users, which are connected by a direct link representing a retweet. In this context, the in-degree centrality measures the incoming interactions of each user over all the interactions. Hence, accounts with high in-degree centrality indicate users whose content has been largely re-shared and, thus, represent highly influential nodes in the social network.

In Figure 7, we display the distribution of the in-degree centrality related to the humans targeted by bots in the two voting periods. Whilst in the 2018 midterms most

of the probability mass is in the range between 0 and 0.01 (low centrality score), in the 2016 election bots also targeted a considerable set of humans with larger centrality scores. On average, we find that the humans targeted by bots in 2016 has in-degree centrality scores two times larger with respect to the humans targeted in 2018 ($1.7 \cdot 10^{-2}$ vs. $8 \cdot 10^{-3}$, *p-value* < .001). We conclude that in 2016 bots supporting the presidential candidates were interested in attracting the attention of the most influential human users in the social network, while during the midterms they interacted with any targets regardless of their position and relevance in the social network.

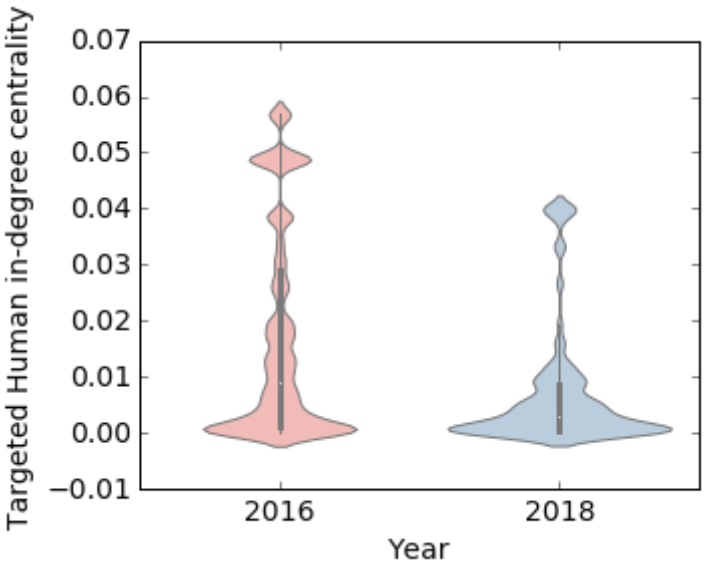


***Figure 7****: Distribution of in-degree centrality of humans targeted by bots*

To further investigate the human-bot interplay, we evaluate the causality between bot and human interactions. We measure the Granger causality (Granger, 1969) of the daily volume of retweets between humans and bots. Our results show that, in the 2018 midterms, retweets from bots to humans (i.e., bots retweeting human

generated content) were Granger-cause (*F-test* = 3.51, *p-value* = *0.02*, steps = 5 days) of the retweets from humans to bots, while there was no signal of Granger causality revealed in the 2016 election.

In the 2016 Presidential election, users likely re-shared content disregarding the authenticity of the information and its source (Bessi & Ferrara, 2016; Allcott & Gentzkow, 2017; Shao *et al.,* 2018; Vosoughi *et al.,* 2018; Bovet & Makse, 2019; Grinberg *et al.,* 2019; Scheufele & Krause, 2019). On the other hand, in 2018 humans likely engaged with bots in part due to bots prior interaction with them. Furthermore, in 2018, the volume of retweets from bots to humans was also Granger-cause (*F-test* = 4.46, *p-value* = 0.01, steps = 5 days) of the retweets from bots to bots, suggesting the possibility that some bots strategically coordinated among each other.

## Discussion and conclusions

Online social media have been dealing with considerable issues of abuse and manipulation. Various studies showed that bots have been widely employed to affect public opinion. In this work, we examined the activity of a set of approximately 245K accounts active on Twitter during the last two US voting events: the 2016 Presidential election and the 2018 midterms. Among this set of accounts, about 31K were labeled as likely bots by using Botometer.

We analyzed the online behavior of humans and bots by comparing the volume and the temporal dynamics of their sharing activities. We showed that, while in 2016 bots

and humans tweeted at a different rate, in 2018, bots better aligned with humans' activity trends, suggesting the hypothesis that some bots have grown more sophisticated. We also noticed a relevant reduction in the usage of retweets, both from human and bot accounts. Human users significantly increased the volume of replies, which denotes a growing propensity of humans in discussing (either positively and negatively) their ideas instead of simply re-sharing content generated by other users. This is a positive sign, since the spread of low-credibility content during the 2016 US presidential election has been often associated with indiscriminate re-sharing (Bessi & Ferrara, 2016; Shao *et al.,* 2018; Vosoughi *et al.,* 2018). Bots, however, exhibited behaviors aimed at polling human opinions and preferences.

While the increase in usage of replies, along with the reduction of retweets, may represent an encouraging step forward in the fight against misinformation, this intuition should be contextualized considering social media history and development over the last few years: Prior to the investigations into the 2016 US presidential election, most social media users may have not been aware of the existence of malicious and/or automated accounts. This may have changed over the course of the last few years. Although the observed change exhibited by human users in 2018 provides an optimistic perspective, there exists an inevitable interplay between the behavior of human users and bots. The mutable nature of bots, coupled with their continuous online presence, should be cause of concern when considering the integrity of the online information ecosystem, especially with respect to online political discussions concerning voting events all over the world.

This set of open problems poses numerous challenges in the fight against social media abuse and motivates further research for a better understanding of nefarious actors' behavior, strategies, and their evolution over time.

## Acknowledgements

EF gratefully acknowledges support by the Air Force Office of Scientific Research (AFOSR award FA9550-17-1-0327). LL and SG are supported by the Swiss National Science Foundation via the CHIST-ERA project UPRISE IoT. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFOSR or the U.S. Government.

## Notes

[1] https://edition.cnn.com/2019/05/23/tech/facebook-transparency-report/index.html

[2] https://www.bbc.com/news/technology-44682354

[3] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001

[4] https://botometer.iuni.iu.edu/

[5] Note that, to validate our findings, we repeated this evaluation at varying bot score threshold (from 0.3 to 0.7) with no significant changes on the results.

[6] The in-degree centrality is a network analysis measure that assigns a score to every node in a network based only on the number of inbound links held by each node. Formally, the in-degree centrality of a generic node *u* is the number of its

incoming edges normalized by the maximal possible in-degree in the network, i.e., *n-1* in a network with n nodes.

## References

A. Addawood, A. Badawy, K. Lerman, E. Ferrara, 2019. “Linguistic Cues to Deception: Identifying Political Trolls on Social Media,” *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 13, No. 01, pp. 15-25).

H. Allcott, M. Gentzkow, 2017. “Social media and fake news in the 2016 election,” *Journal of Economic Perspectives*, 31(2):211–36.

A. Badawy, E. Ferrara, K. Lerman, 2018a. “Who falls for online political manipulation?,” *Companion Proceedings of the 2019 World Wide Web Conference*, 162-168.

A. Badawy, E. Ferrara, K. Lerman, 2018b. “Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign,” *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.

A. Badawy, A. Addawood, K. Lerman, E. Ferrara, 2019. “Characterizing the 2016 Russian IRA influence campaign,” *Social Network Analysis and Mining*, 9(1), 31.

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

O. Boichak, S. Jackson, J. Hemsley, S. Tanupabrungsun, 2018. “Automated diffusion? bots and their influence during the 2016 us presidential election,” International Conference on Information, pp. 17–26.

A. Bovet, H.A. Makse, 2019. “Influence of fake news in Twitter during the 2016 us presidential election,” *Nature Communications,* 10(1):7.

C. Cadwalladr, 2017. “The great British Brexit robbery: how our democracy was hijacked,” The Guardian, 7.

N. Chavoshi, H. Hamooni, A. Mueen, 2016. “DeBot: Twitter Bot Detection via Warped Correlation,” *ICDM*, 817–822.

Z. Chen, D. Subramanian, 2018. “An Unsupervised Approach to Detect Spam Campaigns that Use Botnets on Twitter,” *arXiv preprint arXiv:1804.05232*.

S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, M. Tesconi, 2017. “The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race,” *WWW '17 Companion Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 963-972.

C.A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, 2016. “Botornot: A system to evaluate social bots,” *Proceedings of the 25th International Conference Companion on World Wide Web.*

M. Del Vicario, F. Zollo, G. Caldarelli, A. Scala, W. Quattrociocchi, 2017. “Mapping social dynamics on facebook: The brexit debate,” *Social Networks* 50:6–16.

E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, 2016a. “The rise of social bots,” *Communications of the ACM*, 59(7):96–104.

E. Ferrara, O. Varol, F. Menczer, A. Flammini, 2016b. "Detection of promoted social media campaigns," *10th Int AAAI Conf on Web and Social Media*, pp. 553-556.

E. Ferrara, 2019. “This History of Digital Spam,” *Communications of the ACM*, 62(8).

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

C.W.J. Granger CW, 1969. “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica*, volume 37, number 3, pp. 424–438.

N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, 2019. “Fake news on Twitter during the 2016 U.S. presidential election,” *Science*, 363(6425):374–378.

P.N. Howard, B. Kollanyi, 2016. “Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum,” *SSRN*: https://ssrn.com/abstract=2798311 or http://dx.doi.org/10.2139/ssrn.2798311

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

P.N. Howard, G. Bolsover, B. Kollanyi, S. Bradshaw, L.M. Neudert, 2017. “Junk news and bots during the US election: What were michigan voters sharing over Twitter,” *CompProp*, OII, Data Memo.

J. Im, E. Chandrasekharan, J. Sargent, P. Lighthammer, T. Denby, A. Bhargava, L. Hemphill, D. Jurgens, E. Gilbert, 2019. “Still out there: Modeling and identifying russian troll accounts on Twitter,” *arXiv preprint arXiv:1901.11162.*

S. Kudugunta, E. Ferrara, 2018. “Deep Neural Networks for Bot Detection,” *Information Sciences,* 467, 312–322.

L. Luceri, A. Deb, A. Badawy, E. Ferrara, 2019. “Red bots do it better: comparative analysis of social bot partisan behavior,” *Companion Proceedings of the 2019 World Wide Web Conference*, 1007-1012.

P.T. Metaxas, E. Mustafaraj, 2012. “Social media and the elections,” *Science,* 338(6106):472–473.

B. Mønsted, P. Sapiezynski, E. Ferrara, S. Lehmann, 2017. “Evidence of complex contagion of information in social media: An experiment using twitter bots,” *Plos One*, 12(9):e0184148

I. Pozzana, E. Ferrara, 2018. “Measuring bot and human behavioral dynamics,” *arXiv preprint*, arXiv:1802.04286.

A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, 2019. "Language models are unsupervised multitask learners," OpenAI Blog 1(8).

J. Ratkiewicz, M. D. Conover, M. Meiss, B. Gonçalves, A. Flammini, F. Menczer, 2011. “Detecting and tracking political abuse in social media,” *Fifth Int AAAI Conf on Weblogs and Social Media*, 11:297–304.

D. J. Ruck, N. M. Rice, J. Borycz, R. A. Bentley, 2019. “Internet Research Agency Twitter activity predicted 2016 US election polls,” *First Monday*, *24*(7).

D.A. Scheufele, N.M. Krause, 2019. “Science audiences, misinformation, and fake news,” *PNAS* p. 201805871.

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

K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, 2017. “Fake news detection on social media: A data mining perspective,” *ACM SIGKDD Explorations Newsletter,* 19(1):22–36.

M. Stella, E. Ferrara, M. De Domenico, 2018. “Bots increase exposure to negative and inflammatory content in online social systems,” *Proceedings of the National Academy of Sciences,* 115(49):12435–12440.

M. Stella, M. Cristoforetti, M. De Domenico, 2019. “Influence of augmented humans in online interactions during voting events,” *PloS one*, 14(5):e0214210.

V.S. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, 2016. “The DARPA Twitter Bot Challenge,” *Computer* 49, 6.

M.B.K. Thelwall, G. Paltoglou G, 2010. “Heart and soul: sentiment strength detection in the social web with sentistrength,” *Journal of Language and Social Psychology,* pp. 24–54.

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

O. Varol, E. Ferrara, C.A. Davis, F. Menczer, A. Flammini, 2017b. “Online human-bot interactions: Detection, estimation, and characterization,” *Int. AAAI Conference on Web and Social Media,* pp. 280–289.

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

S.C. Woolley, D.R. Guilbeault, 2017. “Computational propaganda in the united states of america: Manufacturing consensus online,” *CompProp Research Project,* p. 22.

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

## Supplement

## 2016 data collection: list of keywords

#election2016, #elections2016, #tcot, #p2, #hillaryclinton, #donaldtrump, #presidentialdebate, #debates2016, #imwithher, #trump2016, #nevertrump, #neverhillary, #trumppence16, #hillary, #trumpwon, #debate, #trump, #garyjohnson, #jillstein, #jillnothill, #debatenight, #debates, #VPDebate