

# Bots, Elections, and Social Media: A Brief Overview



**Emilio Ferrara**

**Abstract** Bots, software-controlled accounts that operate on social media, have been used to manipulate and deceive. We studied the characteristics and activity of bots around major political events, including elections in various countries. In this chapter, we summarize our findings of bot operations in the context of the 2016 and 2018 US Presidential and Midterm elections and the 2017 French Presidential election.

**Keywords** Social media · Bots · Influence · Disinformation

## 1 Introduction

Social media have been widely portrayed as enablers of democracy [12, 15, 47, 48, 50]. In countries where freedom to communicate and organize are lacked, social media provided a platform to openly discuss political [2, 9, 13, 23, 25, 55, 87] and social issues [8, 18, 19, 37, 38, 77, 82], without fears for safety or retaliation. Such platforms have also been used to respond to crises and emergencies [34, 49, 75, 88, 89]. It is hard to overstate the importance of these platforms for the billions of people who use them every day, all over the world.

However, as it happens with most powerful emerging technologies, the rise of popularity led to abuse. Concerns about the possibility of manipulating public opinion using social media have been brought a decade before they materialized [39]. Ample evidence was provided by the scientific community that social media can influence people's behaviors [5, 14, 31, 32, 45, 60]. These concerns have been corroborated by numerous recent studies [26–28, 40, 58, 66, 68, 81].

Social media can be used to reach millions of people using targeted strategies aimed to maximize the spread of a message. If the goal is to manipulate public

---

E. Ferrara (✉)

USC Information Sciences Institute, Marina del Rey, CA, USA

e-mail: [emiliofe@usc.edu](mailto:emiliofe@usc.edu)

opinion, one way to achieve it is by means of bots, software-controlled social media accounts whose goal is to mimic the characteristics of human users, while operating at much higher pace at substantially no downside for their operators. Bots can emulate all basic human activity on social media platforms, and they become increasingly more sophisticated as new advancements in Artificial Intelligence emerge [30, 41, 57, 70, 80].

In this chapter, we focus on the use of bots to manipulate the political discourse. The first anecdotal accounts of attempts to steer public opinion on Twitter date back to the 2010 US Midterm election [65] and similarly during the 2010 US Senate special election in Massachusetts [58, 62], where bots were used to generate artificial support for some candidates and to smear their opponents.

Attribution, i.e., the determination of the actors behind such operations, has proven challenging in most such cases [30]. One notorious exception is represented by the attribution of an interference campaign occurred during the 2016 US Presidential election to a Russian-sponsored operation. This was as a result of a thorough investigation on Russian interference led by the US Senate Select Committee on Intelligence (SSCI). They found that “The Russian government interfered in the 2016 U.S. presidential election with the goal of harming the campaign of Hillary Clinton, boosting the candidacy of Donald Trump, and increasing political and social discord in the United States.”<sup>1</sup> Numerous studies have investigated the events associated with this operation [7, 10, 44].

It is worth noting that bots have been used for other purposes, for example social spam and phishing [29, 42, 43, 61, 69, 78, 79, 85]. Albeit much work has been devoted to the challenges of detecting social spam [35, 56, 90] and spam bots [11, 51, 52, 61, 72], only recently the research community started to investigate the effects that bots have on society, political discourse, and democracy. The goal of this chapter is to summarize some of the most important results in this space.

## 1.1 Contributions of This Chapter

The aim of this chapter is to connect results of our investigations into three major political events: (i) the 2016 US Presidential election; (ii) the 2017 French Presidential election; and (iii) the 2018 US Midterm elections. We will discuss the role of bots in these events, and highlight the influence they had on the online political discourse. The contributions of this chapter are as follows:

- We first provide a brief overview of how bots operate and what are the challenges in detecting them. Several recent surveys have been published on the problem of characterizing and detecting bots [71, 86], including our own on *Communications of the ACM* [30].

---

<sup>1</sup>See Wikipedia: [https://en.wikipedia.org/wiki/Russian\\_interference\\_in\\_the\\_2016\\_United\\_States\\_elections](https://en.wikipedia.org/wiki/Russian_interference_in_the_2016_United_States_elections)

- We then illustrate our first, and maybe the most prominent, use case of bots-driven interference in political discourse, discussing how bots have been used during the 2016 US Presidential election to manipulate the discussion of the presidential candidates. This overview is based on our results that appeared prior to the November 8, 2016 election events [10].
- We then illustrate how bots have been used to spread disinformation prior to the 2017 French Presidential election to smear Macron’s public image.
- Finally, we overview recent results that suggest how bots have been evolving over the course of the last few years, focusing on the 2018 US Midterm elections, and we discuss the challenges associated to their detection.

## 2 Anatomy of a Bot

### 2.1 What Is a Bot

In this chapter, we define *bot* (short for *robot*, a.k.a., social bot, social media bot, social spam bot, or sybil account) as a social media account that is predominantly controlled by software rather than a human user. Although the definition above inherently states nothing about the intents behind creating and operating a bot, according to published literature, malicious applications of bots are reported significantly more frequently than legitimate usage [30, 71].

While in this chapter we will focus exclusively on bots that aim to manipulate the public discourse, it is worth noting that some researchers have used bots for social good [4, 60], as illustrated by a recent taxonomy that explores the interplay between intent and characteristics of bots [71]. Next, we describe some techniques to create and detect bots.

### 2.2 How to Create a Bot

In the early days of online social media, in the late 2000s, creating a bot was not a simple task: a skilled programmer would need to sift through various platforms’ documentation to create a software capable of automatically interfacing with the front-end or the back-end, and operate functions in a human-like manner.

These days, the landscape has completely changed: indeed, it has become increasingly simpler to deploy bots, so that, in some cases, no coding skills are required to setup accounts that perform simple automated activities: tech blogs often post tutorials and ready-to-go tools for this purposes. Various source codes for sophisticated social media bots can be found online as well, ready to be customized and optimized by the more technically-savvy users [44].

We recently inspected some of the readily-available Twitter bot-making tools and compiled a non-comprehensive list of capabilities they provide [10, 28].

Most of these bots can run within cloud services or infrastructures like *Amazon Web Services* (AWS) or Heroku, making it more difficult to block them when they violate the Terms of Service of the platform where they are deployed.

A very recent trend is that of providing Bot-As-A-Service (BaaS): Advanced conversational bots powered by sophisticated Artificial Intelligence are provided by companies like *ChatBots.io* that can be used to carry digital spam campaigns [29] and scale such operations by automatically engaging with online users.

Finally, the increasing sophistication of Artificial Intelligence (AI) models, in particular in the area of *neural-based natural language generation*, and the availability of large pre-trained models such as OpenAI's GPT-2 [64], makes it easy to programmatically generate text content. This can be used to program bots that produce genuine-looking short texts on platforms like Twitter, making it harder to distinguish between human and automated accounts [3].

### 2.3 How to Detect Bots

The detection of bots in online social media platform has proven a challenging task. For this reason, it has attracted a lot of attention from the computing research community. Even DARPA, the U.S. *Defense Advanced Research Projects Agency*, became interested and organized the 2016 DARPA Twitter Bot Detection [74], with University of Maryland, University of Southern California, and Indiana University topping the challenge, focused on detecting bots pushing anti vaccination campaigns. Large botnets have been identified on Twitter, from dormant [24, 24], to very active [1].

The literature on bot detection has become very extensive. We tried to summarize the most relevant approaches in a survey paper recently appeared on the *Communications of the ACM* [30]: In that review, we proposed a simple taxonomy to divide the bot detection approaches into three classes: (i) bot detection systems based on social network information; (ii) systems based on crowd-sourcing and leveraging human intelligence; (iii) machine learning methods based on the identification of highly-predictive features that discriminate between bots and humans. We refer the interested reader to that review for a deeper analysis of this problem [30]. Other recent surveys propose complementary or alternative taxonomies that are worth considering as well [20, 20, 71, 86].

As of today, there are a few publicly-available tools that allow to do bot detection and study social media manipulation, including (i) Botometer,<sup>2</sup> a popular bot detection tool developed at Indiana University [21], (ii) BotSlayer,<sup>3</sup> an application

---

<sup>2</sup>Botometer: <https://botometer.iuni.iu.edu/>

<sup>3</sup>BotSlayer: <https://osome.iuni.iu.edu/tools/botslayer/>

that helps track and detect potential manipulation of information spreading on Twitter, and (iii) the Bot Repository,<sup>4</sup> a centralized database to share annotated datasets of Twitter social bots.

In conclusion, several algorithms have been published to detect bots using sophisticated machine learning techniques including deep learning [46], anomaly detection [22, 36, 59], and time series analysis [16, 73].

### 3 Social Media Manipulation

Bots have been reportedly used to interfere in political discussions online, for example by creating the impression of an organic support behind certain political actors [58, 62, 65, 66]. However, the apparent support can be artificially generated by means of orchestrated campaigns with the help of bots. This strategy is commonly referred to as social media *astroturf* [66].

#### 3.1 2016 US Presidential Election

Our analysis of social media campaigns during the 2016 US Presidential Election revealed the presence of social bots. We here summarize our findings first published in [10], discussing data collection, bot detection, and sentiment analysis.

**Data Collection** We manually crafted a list of hashtags and keywords related to the 2016 US Presidential Election with 23 terms in total, including 5 terms specifically for the Republican Party nominee Donald Trump, 4 terms for the Democratic Party nominee Hillary Clinton, and the remainder terms relative to the four presidential debates. The complete list of search terms is reported in our paper [10]. By querying the Twitter Search API between September 16 and October 21, 2016, we collected a large dataset. After post-processing and cleaning procedures, we studied a corpus constituted by 20.7 million tweets posted by nearly 2.8 million distinct users.

**Bot Detection** We used Botometer *v1* (the version available in 2016) to determine the likelihood that the most active accounts in this dataset were controlled by humans or were otherwise bots. To label accounts as bots, we use the 50% threshold—which has proven effective in prior studies [21, 30]—an account was considered to be a bot if the bot score was above 0.5. Due to the Twitter API limitations, it would have been impossible to test all the 2.78 million accounts in short time. Therefore, we tested the top 50 thousand accounts ranked by activity volume, which account for roughly 2% of the entire population and yet are responsible for producing over 12.6 million tweets, which is about 60% of the total

---

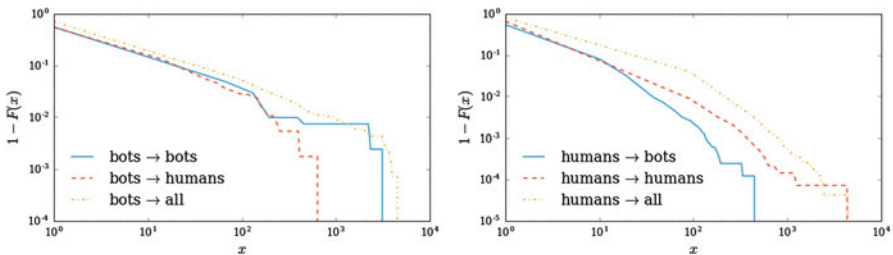
<sup>4</sup>Bot Repository: <https://botometer.iuni.iu.edu/bot-repository/>

conversation. Of the top 50 thousand accounts, Botometer classified as likely bots a total of 7,183 users (nearly 15%), responsible for 2,330,252 tweets; 2,654 users were classified as undecided, because their scores did not significantly diverge from the classification threshold of 0.5; the rest—about 40 thousand users (responsible for just 10.3 million tweets, less than 50% of the total)—were labeled as humans. Additional statistics are summarized in our paper [10].

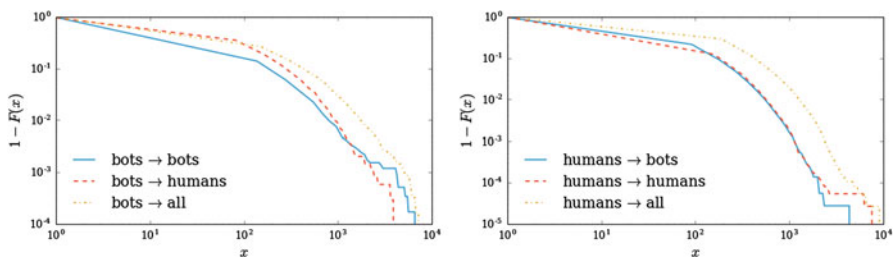
**Sentiment Analysis** We leveraged sentiment analysis to quantify how bots (resp., humans) discussed the candidates. We used SentiStrength [76] to derive the sentiment scores of each tweet in our dataset. This toolkit is especially optimized to infer sentiment in short informal texts, thus ideally suited for social media. We tested it extensively in prior studies on the effect of sentiment on tweets’ diffusion [32, 33]. The algorithm assigns to each tweet  $t$  a positive  $P^+(t)$  and negative  $P^-(t)$  polarity score, both ranging between 1 (neutral) and 5 (strongly positive/negative). Starting from the polarity scores, we captured the emotional dimension of each tweet  $t$  with one single measure, the sentiment score  $S(t)$ , defined as the difference between positive and negative polarity scores:  $S(t) = P^+(t) - P^-(t)$ . The above-defined score ranges between  $-4$  and  $+4$ . The negative extreme indicates a strongly negative tweet, and occurs when  $P^+(t) = 1$  and  $P^-(t) = 5$ . Vice-versa, the positive extreme identifies a strongly positive tweet labeled with  $P^+(t) = 5$  and  $P^-(t) = 1$ . In the case  $P^+(t) = P^-(t)$ —positive and negative sentiment scores for a tweet  $t$  are the same—the sentiment  $S(t) = 0$  of tweet  $t$  is considered neutral as the polarities cancel each other out.

**Partisanship and Supporting Activity** We used a simple heuristic based on the 5 Trump-supporting hashtags and the 4 Clinton-supporting to attribute user partisanship. For each user, we calculated their top 10 most used hashtags: If the majority supported one particular candidate, we assigned the given user to that political group (Clinton or Trump supporter). Compared to network-based techniques [6, 17], this simple partisanship assignment yielded a smaller yet higher-confidence annotated dataset, constituted by 7,112 Clinton supporters (590 bots and 6,522 humans) and 17,202 Trump supporters (1,867 bots and 15,335 humans).

**Summary of Results: Engagement** Figures 1 and 2 illustrate the Complementary Cumulative Distribution Functions (CCDFs) of replies and retweets initiated by bots



**Fig. 1** Complementary cumulative distribution function (CCDF) of replies interactions generated by bots (left) and humans (right) (from [10])



**Fig. 2** Complementary cumulative distribution function (CCDF) of retweets interactions generated by bots (left) and humans (right) (from [10])

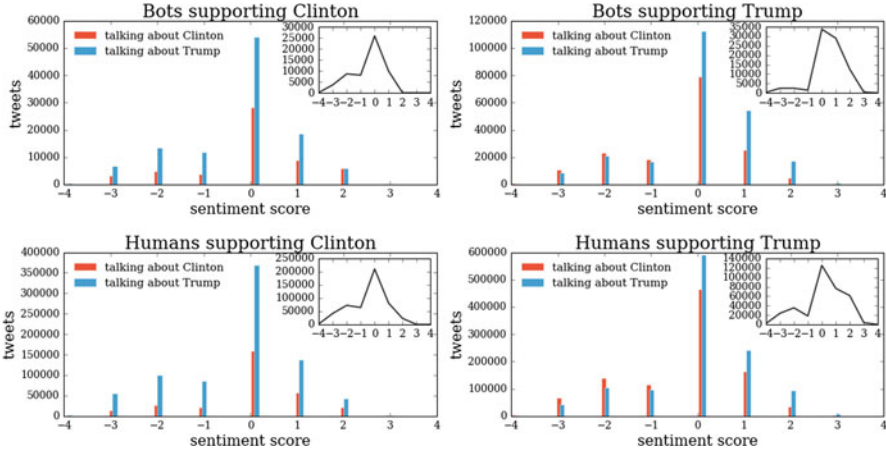
and humans in three categories: (i) within group (for example bot-bot, or human-human); (ii) across groups (e.g., bot-human, or human-bot); and, (iii) total (i.e., bot-all and human-all). The heavy-tailed distributions, typically observed in social systems, appear in both. Hence, further inspection of Fig. 1 suggests that (i) humans replied significantly more to other humans than to bots and, (ii) conversely, bots receive replies from other bots significantly more than from humans. One hypothesis is that unsophisticated bots could not produce engaging-enough questions to foster meaningful exchanges with humans.

Figure 2, however, demonstrates that retweets were a much more vulnerable mode of information diffusion: there is no statistically significant difference in the amount of retweets that humans generated by resharing content produced by other humans or by bots. In fact, humans and bots retweeted each other substantially at the same rate. This suggests that bots were very effective at getting their messages reshared in the human communication channels.

Our study highlighted a vulnerability in the information ecosystem at that time, namely that content was reshared often without a thorough scrutiny on the information source. Several subsequent studies hypothesized that bots may have played a role in the spread of false news and unverified rumors [67, 83].

**Summary of Results: Sentiment** We further explored how bots and humans talked about the two presidential candidates. Next, we show the sentiment analysis results based on *SentiStrength*. Figure 3 illustrates four settings: the top (resp., bottom) two panels show the sentiment of the tweets produced by the bots (resp., humans). Furthermore, the two left (resp., right) panels show the support for Clinton (resp., Trump). The main histograms in each panel show the volume of tweets about Clinton or Trump, whereas the insets show the difference between the two. By contrasting the left and right panels we note that the tweets mentioning Trump are significantly more positive than those mentioning Clinton, regardless of whether the source is human or bot. However, bots tweeting about Trump generated almost no negative tweets and indeed produced the most positive set of tweets in the entire dataset (about 200,000 or nearly two-third of the total).

The fact that bots produce systematically more positive content in support of a candidate can bias the perception of the individuals exposed to it, suggesting that



**Fig. 3** Distributions of the sentiment of bots (top) and humans (bottom) supporting the two presidential candidates. The main histograms show the disaggregated volumes of tweets talking about the two candidates separately, while the insets show the absolute value of the difference between them (from [10])

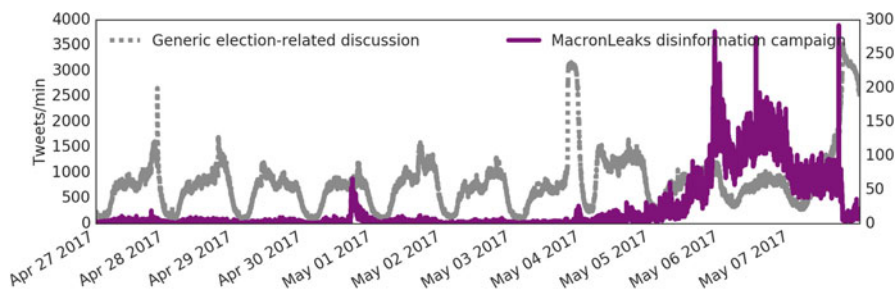
there exists an organic, grassroots support for a given candidate, while in reality it is in part artificially inflated. Our paper reports various examples of tweets generated by bots, and the candidate they support [10].

### 3.2 2017 French Presidential Election

A subsequent analysis of the Twitter ecosystem highlighted the presence and effects of bots prior to the 2017 French Presidential Election. We next report our findings summarizing the results published in 2017 [28]. We provide a characterization of both the bots and the users who engaged with them.

**Data Collection** By following the same strategy as in the 2016 US Presidential election [10], we manually selected a set of hashtags and keywords related to the 2017 French Presidential Election. By construction, the list contained a roughly equal number of terms associated with each of the two candidates, namely Marine Le Pen and Emmanuel Macron, and various general election-related terms: we ultimately identified 23 terms, listed in our paper [28]. We collected data by using the Twitter Search API, from April 27 to the end of election day, on May 7, 2017: This procedure yielded a dataset containing approximately 17 million unique tweets, posted by 2,068,728 million unique users. Part of this corpus is a subset of tweets associated with the *MacronLeaks* disinformation campaign, whose details are described in our paper [28]. The timeline of the volume of posted tweets is illustrated in Fig. 4.

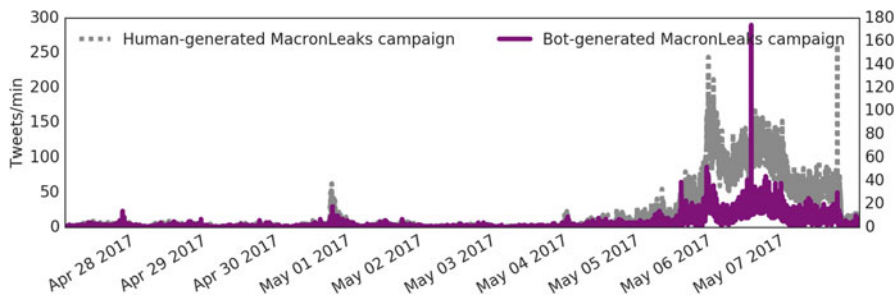




**Fig. 4** Timeline of the volume of tweets generated every minute during our observation period (April 27 through May 7, 2017). The purple solid line (right axis) shows the volume associated with MacronLeaks, while the dashed grey line (left axis) shows the volume of generic election-related discussion. The presidential election occurred on May 7, 2017 (from [28])

**Bot Detection** Due to the limitations of the Twitter API, and the time restrictions for this short period of unfolding events, we were unable to run in real time the bot detection relying upon Botometer. For this reason, we carried out a post-hoc bot detection on the dataset using an offline version of the bot-detection algorithm inspired by Botometer’s rationale. Specifically, we exclusively leveraged user metadata and activity features to create a simple yet effective bot detection classifier, trained on same data as Botometer, which is detailed in our paper [28]. We validated its classification accuracy and assessed that it was similar to Botometer’s performance, with above 80% in both accuracy and AUC-ROC scores. Manual validation corroborated the performance analysis. Hence, we used this simplified bot detection strategy to unveil bots in the dataset at hand.

**Summary: Temporal Dynamics** We started by exploring the timeline of the general election-related discussion on Twitter. The broader discussion that we collected concerns the two candidates, Marine Le Pen and Emmanuel Macron, and spans the period from April 27 to May 7, 2017, the Presidential Election Day, see Fig. 4. Let us discuss first the dashed grey line (left axis): this shows the volume of generic election-related discussion. The discussion exhibits common circadian activity patterns and a slightly upwards trend in proximity to Election Day, and spikes in response to an off-line event, namely the televised political debate that saw Le Pen facing Macron. Otherwise, the number of tweets per minute averages between 300 and 1,500 during the day, and quickly approaches de facto zero overnight, consistently throughout the entire observation window. Figure 4 also illustrates with the purple solid line (right axis) the volume associated with MacronLeaks, the disinformation campaign that was orchestrated to smear Macron’s reputation. The temporal pattern of this campaign is substantially different from the general conversation. First, the campaign is substantially silent for the entire period till early May. We can easily pinpoint the inception of the campaign on Twitter, which occurs in the afternoon of April 30, 2017. After that, a surge in the volume of tweets, peaking at nearly 300 per minute, happens in the run up to

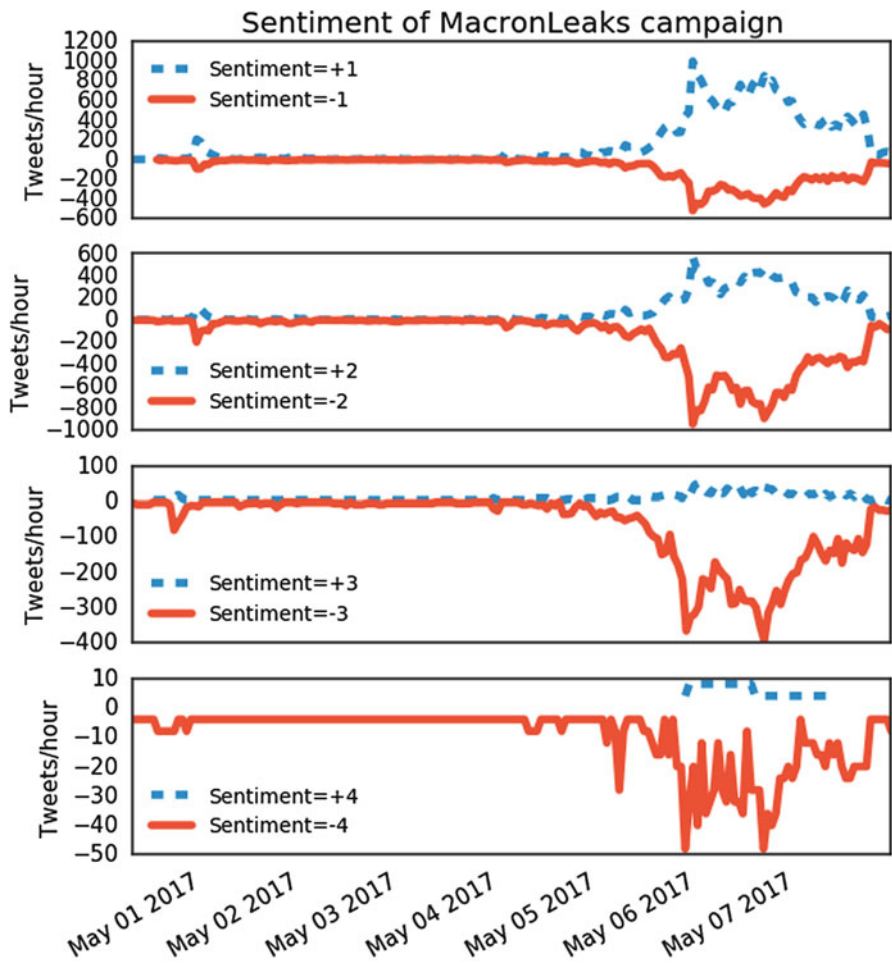


**Fig. 5** Timeline of the volume of tweets generated every minute, respectively by human users (dashed grey line) and social bots (solid purple line), between April 27 and May 7, 2017, and related to MacronLeaks. Spikes in bot-generated content often slightly precedes spikes in human posts, suggesting that bots can trigger cascades of disinformation (from [28])

Election Day, between May 5 and May 6, 2017. It is worth noting that such a peak is nearly comparable in scale to the volume of the regular discussion, suggesting that for a brief interval of time (roughly 48 h) the MacronLeaks disinformation campaign acquired significant attention [28].

**Summary: Bot Dynamics** Like in the previous study, we here provide a characterization of the Twitter activity, this time specifically related to MacronLeaks, for both bot and human accounts. In Fig. 5, we show the timeline of the volume of tweets generated respectively by human users (dashed grey line) and bots (solid purple line), between April 27 and May 7, 2017, and related to MacronLeaks. The amount of activity is substantially close to zero until May 5, 2017, in line with the first coordination efforts as well as the information leaks spurred from other social platforms, as discussed in the paper [28]. Spikes in bot-generated content often appear to slightly precede spikes in human posts, suggesting that bots can trigger cascades of disinformation [67]. At peak, the volume of bot-generated tweets is comparable with the that of human-generated ones. Further investigation revealed that the users who engaged with bots pushed the *MacronLeaks* disinformation campaign were mostly foreigners with pre-existing interest in alt-right topics and alternative news media, rather than French users. Furthermore, we highlighted an anomalous account usage pattern where hundreds of bot accounts used in the 2017 French Presidential elections were also present in the 2016 US Presidential Election discussion, which suggested the possible existence of a black market for reusable political disinformation bots [28].

**Summary: Sentiment Dynamics** Identically to the 2016 US Presidential Election study, we annotated all tweets in this corpus using *SentiStrength*, and subsequently studied the evolution of the sentiment of tweets in the 2017 French Presidential Election discussion. Figure 6 shows the temporal distribution of tweets' sentiment disaggregated by intensity: the four panels illustrate the overall timeline of the volume of tweets that exhibit positive and negative sentiment at the hourly resolution, for sentiment polarities ranging from 1 (lowest) to 4 (highest) in both positive



**Fig. 6** Temporal distribution of sentiment disaggregated by sentiment intensity (hourly resolution). The sign on the y-axis captures the amount of tweets in the positive (resp., negative) sentiment dimension

and negative spectra. What appears evident is that, as Election Day approaches, moderately and highly negative tweets (sentiment scores of  $-2$ ,  $-3$ , and  $-4$ ) significantly outnumber the moderately and highly positive tweets, at times by almost an order of magnitude. For example, between May 6 and 7, 2017, on average between 300 and 400 tweets with significant negative sentiment (sentiment scores of  $-3$ ) were posted every hour, compared with an average of between 10 and 50 tweets with an equivalently positive sentiment (score scores of  $+3$ ). Since the discussion during that period was significantly driven by bots, and bots focused against Macron, our analysis suggested that bots were pushing negative campaigns

against that candidate aimed at smearing his credibility and weakening his position in the eve of the May 7's election.

### 3.3 2018 US Midterms

The notorious investigation on Russian interference led by the US Senate Select Committee on Intelligence (SSCI) put social media service providers (SMSPs) at the center-stage of the public debate. According to reports, SMSPs started to devote more efforts to “sanitize” their platforms, including ramping up the technological solutions to detect and fight abuse. Much attention has been devoted to identifying and suspending *inauthentic activity*, a term that captures a variety of tools used to carry out manipulation, including bot and troll accounts.

Hence, it is natural to ask whether these countermeasures proved effective, or if otherwise the strategies and technologies bots typically used until 2017 evolved, and to what extent they successfully adapted to the changing social media defenses and thus escaped detection. We recently set to answer these questions: to this purpose, we monitored and investigated the online activity surrounding the 2018 US Midterm elections what were held on November 6, 2018.

**Data Collection** We collected data for six weeks, from October 6, 2018 to November 19, 2018, i.e., one month prior and until two weeks after election day. Tweets were collected using the Twitter Streaming API and following these keywords: *2018midtermelections*, *2018midterms*, *elections*, *midterm*, and *midtermelections*. Post-processing and cleaning procedures are described in detail in our paper [53]: we retained only tweets in English, and manually removed tweets that were out of context, e.g., tweets related to other countries' elections (Cameroon, Congo, Biafra, Kenya, India, etc.) that were present in our initial corpus because they contained the same keywords we tracked. The final dataset contains 2.6M tweets, posted by nearly 1M users.

**Bot Detection** Similarly to the 2016 US Presidential election study, since this study was a post-mortem (i.e., not in real time but after the events), we adopted Botometer to infer the bot scores of the users in our dataset. The only distinction worth mentioning is that we used the Botometer API version v3 that brings new features and a non-linear re-calibration of the model: in line with the associated study's recommendations [86], we used a threshold of 0.3 (which corresponds to a 0.5 threshold from previous versions of Botometer) to separate bots from humans (note that the results remain substantially unchanged if a higher threshold was used). As a result, we obtained that 21.1% of the accounts were categorized as bots, which were responsible for 30.6% of the total tweets in our dataset. Manual validation procedures assessed the reasonable quality of these annotations. The resulting evidence suggests that bots were still present, and accounted for a significant amount of the tweets posted in the context of the political discourse revolving around the 2018 US Midterms.

Interestingly, about 40 thousand accounts were already inactive at the time of our analysis, and thus we were not able to infer their bot scores using the Twitter API. We manually verified that 99.4% of them were suspended by Twitter, corroborating the hypothesis that these were bots as well, and were suspended by Twitter in the time between the events and our post-mortem analysis, which was carried out in early 2019.

**Political Leaning Inference** Next, we set to determine if bots exhibited a clear political leaning, and if they acted according to that preference. To label accounts as conservative or liberal, we used a label propagation approach that leveraged the political alignment of news sources whose URLs were posted by the accounts in the dataset. Lists of partisan media outlets were taken from third-party organizations, namely AllSides.Org and MediaBiasFactCheck.Com. The details of our label propagation algorithm are explained in our paper [53]. Ultimately, the procedure allowed us to reliably infer, with accuracy above 89%, the political alignment of the majority of human and bot accounts in our corpus. These were factored into the subsequent analyses aimed at determining partisan strategies and narratives (see [53]).

**Summary: Bot Activity and Strategies** Provided the evidence that bots were still present despite the efforts of the SMSPPs to sanitize their platforms, we aimed at determining the degree to which they were embedded in the human ecosystem, specifically in the retweet network. This network is of central importance in our analysis, because it conveys information diffusion dynamics; many recent studies suggested a connection between bots and the spread of unverified and false information [67, 83]. It is therefore of paramount importance to determine if bots still played a role in the retweet network of election-related social media discourse as of 2018.

To this aim, we resorted to perform the  $k$ -core decomposition analysis. In social network theory, a  $k$ -core is a subgraph of a graph where all nodes have degree at least equal to  $k$ . The intuition is that, as  $k$  grows, one is looking at increasingly more highly-connected nodes' subgraphs. Evidence suggests that high  $k$ -cores are associated with nodes that are more embedded, thus influential, for the network under investigation [84].

If bots were still influential in the 2018 US Midterm election discussion, our hypothesis is that we would find them in high concentration predominantly into high  $k$  cores. This would be consistent with our findings related to the 2016 US Presidential Election discussion [10].

Figure 7 corroborates our intuition. Specifically, we show the percentage of both conservative and liberal human and bot accounts as a function of varying  $k$ . Two patterns are worth discussing: first, as  $k$  increases, the fraction of conservative bots grows, while the prevalence of liberal bots remains more or less constant; conversely, the prevalence of human accounts decreases, with growing  $k$ , more markedly for liberal users than conservative ones. We summarize these findings suggesting that conservative bots were situated in a premium position in the retweet network, and therefore may have affected information spread [53].

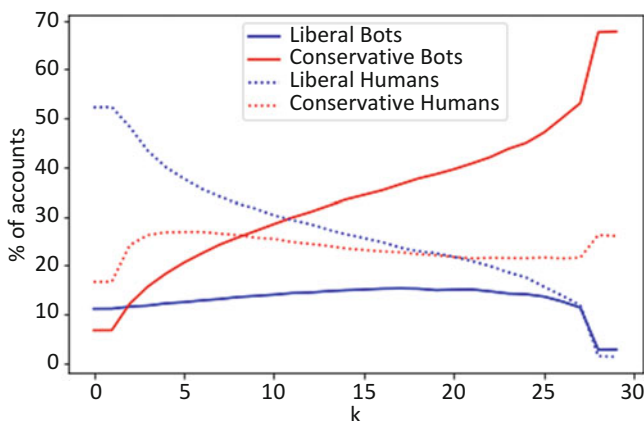


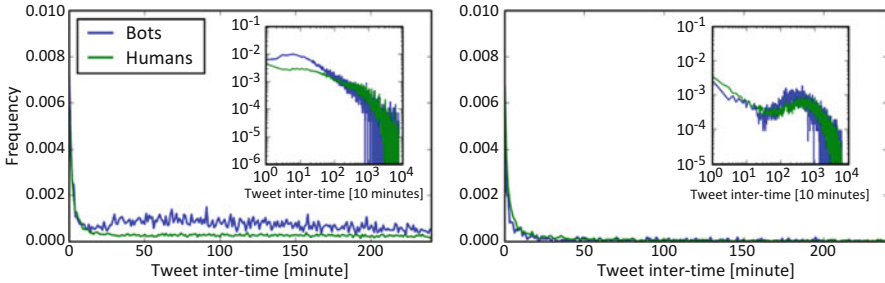
Fig. 7 K-core decomposition: liberal vs. conservative bots and humans (from [53])

### 3.4 2016 vs 2018: A Comparative Bot Analysis

Having identified and analyzed the activity of human and bot accounts in the context of the political discourse associated to US election events in both 2016 and 2018, it is natural to ask whether these studies involved a similar set of accounts. In other words, it is worth determining whether there exists a continuum of users that are active in both time periods under investigation. If this is the case, it would be interesting to study the users present in both periods, determine whether any of them are the bots under scrutiny in the previous studies, and ultimately understand if the strategies they may have exhibited evolved, possibly to escape detection or avoid further scrutiny of SMSPs.

**Data Collection** To answer the questions above, we isolated the users present in both the 2016 and 2018 datasets described above. This process yielded over 278 thousand accounts, active in both periods. Further processing and cleaning procedures, as detailed in our paper [54], brought the dataset down to 245 K users, accounting for over 8.3 M tweets in 2016 and 660 K in 2018. Botometer was used to determine the bot scores of these accounts. As a result, 12.6% of these accounts scored high in bot scores and were therefore classified as bots. We used this dataset to study the evolution of behavior of bots over the time period of study.

**Summary: Bot Evolution Dynamics** One advantage of bots over humans is their scalability. Since bots are controlled by software rather than human users, as such they can work over the clock, they don't need to take rests and don't have the finite cognitive capacity and bandwidth that dictates how humans operate on social media [63]. In principle, a bot could post continuously without any break, or at regular yet tight intervals of time. As a matter of fact, primitive bots used these simple strategies [58, 65]. However, such obvious patterns are easy to spot automatically, hence not



**Fig. 8** Tweet inter-event time by bots and humans in 2016 (left) and 2018 (right). A clear distinction in temporal signature between bots and humans was evident in 2016, but vanished in 2018 (from [54])

very effective. There is therefore a trade-off between realistic-looking activity and effectiveness. In other words, one can investigate the patterns of *inter-event time* between a tweet post and its subsequent, and lay out the frequency distribution in an attempt to distill the difference between human and bot accounts' temporal dynamics.

Figure 8 illustrates the tweet inter-time distribution by bots and humans in 2016 (left) and 2018 (right). It is apparent that, while in 2016 bots exhibited a significantly different frequency distribution with respect to their human counterparts, in 2018 this distinction has vanished. In fact, statistical testing of distribution differences suggests that human and bot temporal signatures are indistinguishable in 2018. The discrepancy is particularly relevant in the time range between 10 min and 3 h, consistent with other findings [63]: in 2016, bots shared content at a higher rate with respect to human users.

Our work [54] corroborates the hypothesis that bots are continuously changing and evolving to escape detection. Further examples that we reported also illustrate other patterns of behavior that have changed between 2016 and 2018: for instance, the sentiment that was expressed in favor or against political candidates in 2018 reflects significantly better what the human crowd is expressing. However, in 2016, bots' sentiment drastically diverged, in a manner easy to detect, from that of the human's counterpart, as we discussed earlier.

## 4 Conclusions

In this chapter, we set to discuss our latest results regarding the role of bots within online political discourse in association with three major political events.

First, we described the results of our analysis that unveiled a significant amount of bots distorting the online discussion in relation to the 2016 US Presidential election. We characterized the activities of such bots, and illustrated how they successfully fostered interactions by means of retweets at the same rate human users

did. Other researchers suggested that this played a role in the spread of false news during that time frame [67, 83].

Second, we highlighted the role of bots in pushing a disinformation campaign, known as *MacronLeaks*, in the run up to the 2017 French Presidential election. We demonstrated how it is possible to easily pinpoint the inception of this disinformation campaign on Twitter, and we illustrated how its popularity peak was comparable with that of regular political discussion. We also hypothesized that this disinformation campaign did not have a major success in part because it was tailored around the information needs and usage patterns of the American alt-right community rather than French-speaking audience. Moreover, we found that several hundreds of bot accounts were re-purposed from the 2016 US Election. Ultimately, we suggested the possibility that a black market for reusable political bots may exist [28].

Third, we studied the 2018 US Midterms, to investigate if bots were still present and active. Our analysis illustrated that not only bots were almost as prevalent as in the two other events, but also that conservative bots played a central role in the highly-connected core of the retweet network. These findings further motivated a comparative analysis contrasting the activity of bots and humans in 2016 and 2018. Our study highlighted that a core of over 245 K users, of which 12.1% were bots, was active in both events. Our results suggest that bots may have evolved to better mimic human temporal patterns of activity.

With the increasing sophistication of Artificial Intelligence, the ability of bots to mimic human behavior to escape detection is greatly enhanced. This poses challenges for the research community, specifically in the space of bot detection. Whether it is possible to win this arms race is yet to be determined: any party with significant resources can deploy state of the art technologies to enact influence operations and other forms of manipulation of public opinion.

The availability of powerful neural language models lowers the bar to adopt techniques that allow to build credible bots. For example, it may be already in principle possible to automatize almost completely the generation of genuine-looking text. This may be used to push particular narratives, to artificially build traction for political arguments that may otherwise have little or no human organic support.

Ultimately, the evidence that our studies, and the work of many other researchers in this field, have brought strongly suggest that more policy and regulations may be warranted, and that technological solutions alone may not be sufficient to tackle the issues of bot interference in political discourse.

**Acknowledgements** The author is grateful to his collaborators and coauthors on the topics covered in this paper, in particular Adam Badawy, Alessandro Bessi, Ashok Deb, and Luca Luceri, who contributed significantly to three papers widely discussed in this chapter [10, 53, 54].



## References

1. Abokhodair, N., Yoo, D., McDonald, D.W.: Dissecting a social botnet: growth, content and influence in twitter. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 839–851. ACM (2015)
2. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 us election: divided they blog. In: 3rd International Workshop on Link Discovery, pp. 36–43. ACM (2005)
3. Alarifi, A., Alsaleh, M., Al-Salman, A.: Twitter turing test: identifying social machines. *Inf. Sci.* **372**, 332–346 (2016)
4. Allem, J.-P., Ferrara, E., Uppu, S.P., Cruz, T.B., Unger, J.B.: E-cigarette surveillance with social media data: social bots, emerging topics, and trends. *JMIR Public Health Surveill.* **3**(4), e98 (2017)
5. Aral, S., Walker, D.: Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Manag. Sci.* **57**(9), 1623–1639 (2011)
6. Badawy, A., Ferrara, E., Lerman, K.: Analyzing the digital traces of political manipulation: the 2016 russian interference twitter campaign. In: Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2018 (2018)
7. Badawy, A., Lerman, K., Ferrara, E.: Who falls for online political manipulation? In: Companion Proceedings of the 2019 World Wide Web Conference, pp. 162–168 (2019)
8. Barberá, P., Wang, N., Bonneau, R., Jost, J.T., Nagler, J., Tucker, J., González-Bailón, S.: The critical periphery in the growth of social protests. *PLoS One* **10**(11), e0143611 (2015)
9. Bekafigo, M.A., McBride, A.: Who tweets about politics? political participation of twitter users during the 2011 gubernatorial elections. *Soc. Sci. Comput. Rev.* **31**(5), 625–643 (2013)
10. Bessi, A., Ferrara, E.: Social bots distort the 2016 us presidential election online discussion. *First Monday* **21**(11) (2016)
11. Boshmaf, Y., Muslukhov, I., Beznosov, K., Ripeanu, M.: The socialbot network: when bots socialize for fame and money. In: Proceedings of the 27th Annual Computer Security Applications Conference, pp. 93–102. ACM (2011)
12. Boyd, D., Crawford, K.: Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Inf. Commun. Soc.* **15**(5), 662–679 (2012)
13. Carlisle, J.E., Patton, R.C.: Is social media changing how we understand political engagement? an analysis of facebook and the 2008 presidential election. *Polit. Res. Q.* **66**(4), 883–895 (2013)
14. Centola, D.: An experimental study of homophily in the adoption of health behavior. *Science* **334**(6060), 1269–1272 (2011)
15. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P.: Measuring user influence in twitter: the million follower fallacy. In: Fourth International AAAI Conference on Weblogs and Social Media (ICWSM 2010), pp. 10–17. AAAI Press (2010)
16. Chavoshi, N., Hamooni, H., Mueen, A.: Debot: twitter bot detection via warped correlation. In: ICDM, pp. 817–822 (2016)
17. Conover, M., Ratkiewicz, J., Francisco, M.R., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on twitter. *ICWSM* **133**, 89–96 (2011)
18. Conover, M.D., Davis, C., Ferrara, E., McKelvey, K., Menczer, F., Flammini, A.: The geospatial characteristics of a social movement communication network. *PLoS One* **8**(3), e55957 (2013)
19. Conover, M.D., Ferrara, E., Menczer, F., Flammini, A.: The digital evolution of occupy wall street. *PLoS One* **8**(5), e64679 (2013)
20. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: The paradigm-shift of social spambots: evidence, theories, and tools for the arms race. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 963–972. International World Wide Web Conferences Steering Committee (2017)
21. Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F.: Botnot: a system to evaluate social bots. In: WWW '16 Companion Proceedings of the 25th International Conference Companion on World Wide Web, pp. 273–274. ACM (2016)

22. De Cristofaro, E., Kourtellis, N., Leontiadis, I., Stringhini, G., Zhou, S., et al.: Lobo: evaluation of generalization deficiencies in twitter bot classifiers. In: Proceedings of the 34th Annual Computer Security Applications Conference, pp. 137–146. ACM (2018)
23. DiGrazia, J., McKelvey, K., Bollen, J., Rojas, F.: More tweets, more votes: social media as a quantitative indicator of political behavior. *PLoS One* **8**(11), e79449 (2013)
24. Echeverria, J., Besel, C., Zhou, S.: Discovery of the twitter Bursty botnet. *Data Science for Cyber-Security* (2017)
25. Effing, R., Van Hillegersberg, J., Huibers, T.: Social media and political participation: are facebook, twitter and youtube democratizing our political systems? In *International Conference on Electronic Participation*, pp. 25–35. Springer (2011)
26. El-Khalili, S.: Social media as a government propaganda tool in post-revolutionary Egypt. *First Monday* **18**(3) (2013)
27. Ferrara, E.: Manipulation and abuse on social media. *ACM SIGWEB Newsletter* (4) (2015)
28. Ferrara, E.: Disinformation and social bot operations in the run up to the 2017 french presidential election. *First Monday* **22**(8) (2017)
29. Ferrara, E.: The history of digital spam. *Commun. ACM* **62**(8), 2–91 (2019)
30. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. *Commun. ACM* **59**(7), 96–104 (2016)
31. Ferrara, E., Varol, O., Menczer, F., Flammini, A.: Detection of promoted social media campaigns. In: *Tenth International AAAI Conference on Web and Social Media*, pp. 563–566 (2016)
32. Ferrara, E., Yang, Z.: Measuring emotional contagion in social media. *PLoS One* **10**(11), e0142390 (2015)
33. Ferrara, E., Yang, Z.: Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Comput. Sci.* **1**, e26 (2015)
34. Gao, H., Barbier, G., Goolsby, R.: Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intell. Syst.* **26**(3), 10–14 (2011)
35. Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., Zhao, B.Y.: Detecting and characterizing social spam campaigns. In: *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, pp. 35–47. ACM (2010)
36. Gilani, Z., Kochmar, E., Crowcroft, J.: Classification of twitter accounts into automated agents and human users. In: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pp. 489–496. ACM (2017)
37. González-Bailón, S., Borge-Holthoefer, J., Moreno, Y.: Broadcasters and hidden influentials in online protest diffusion. *Am. Behav. Sci.* **57**(7), 943–965 (2013)
38. González-Bailón, S., Borge-Holthoefer, J., Rivero, A., Moreno, Y.: The dynamics of protest recruitment through an online network. *Sci. Rep.* **1**, 197 (2011)
39. Howard, P.N.: *New Media Campaigns and the Managed Citizen*. Cambridge University Press, Cambridge (2006)
40. Howard, P.N., Kollanyi, B.: Bots, #strongerin, and #brexit: computational propaganda during the uk-eu referendum. Available at SSRN 2798311 (2016)
41. Hwang, T., Pearce, I., Nanis, M.: Socialbots: voices from the fronts. *Interactions* **19**(2), 38–45 (2012)
42. Jagatic, T.N., Johnson, N.A., Jakobsson, M., Menczer, F.: Social phishing. *Commun. ACM* **50**(10), 94–100 (2007)
43. Jin, X., Lin, C., Luo, J., Han, J.: A data mining-based spam detection system for social media networks. *Proceedings VLDB Endowment* **4**(12), 1458–1461 (2011)
44. Kollanyi, B., Howard, P.N., Woolley, S.C.: Bots and automation over twitter during the first us presidential debate. Technical report, *COMPROM Data Memo* (2016)
45. Kramer, A.D., Guillory, J.E., Hancock, J.T.: Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci.* **111**(24), 8788–8790 (2014)
46. Kudugunta, S., Ferrara, E.: Deep neural networks for bot detection. *Inf. Sci.* **467**, 312–322 (2018)

47. Kümpel, A.S., Karnowski, V., Keyling, T.: News sharing in social media: a review of current research on news sharing users, content, and networks. *Soc. Media Soc.* **1**(2), 2056305115610141 (2015)
48. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web, pp. 591–600 (2010)
49. Latonero, M., Shklovski, I.: Emergency management, twitter, and social media evangelism. In: Using Social and Information Technologies for Disaster and Crisis Management, pp. 196–212. IGI Global, Hershey (2013)
50. Lazer, D., Pentland, A.S., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al.: Life in the network: the coming age of computational social science. *Science* (New York, NY) **323**(5915), 721 (2009)
51. Lee, K., Caverlee, J., Webb, S.: The social honeypot project: protecting online communities from spammers. In: Proceedings of the 19th International Conference on World Wide Web, pp. 1139–1140. ACM (2010)
52. Lee, K., Caverlee, J., Webb, S.: Uncovering social spammers: social honeypots+ machine learning. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 435–442. ACM (2010)
53. Luceri, L., Deb, A., Badawy, A., Ferrara, E.: Red bots do it better: comparative analysis of social bot partisan behavior. In: Companion Proceedings of the 2019 World Wide Web Conference, pp. 1007–1012 (2019)
54. Luceri, L., Deb, A., Giordano, S., Ferrara, E.: Evolution of bot and human behavior during elections. *First Monday* **24**(9) (2019)
55. Lutz, C., Hoffmann, C.P., Meckel, M.: Beyond just politics: a systematic literature review of online participation. *First Monday* **19**(7) (2014)
56. Markines, B., Cattuto, C., Menczer, F.: Social spam detection. In: Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web, pp. 41–48 (2009)
57. Messias, J., Schmidt, L., Oliveira, R., Benevenuto, F.: You followed my bot! transforming robots into influential users in twitter. *First Monday* **18**(7) (2013)
58. Metaxas, P.T., Mustafaraj, E.: Social media and the elections. *Science* **338**, 472–473 (2012)
59. Minnich, A., Chavoshi, N., Koutra, D., Mueen, A.: Botwalk: efficient adaptive exploration of twitter bot networks. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp. 467–474. ACM (2017)
60. Mønsted, B., Sapieżyński, P., Ferrara, E., Lehmann, S.: Evidence of complex contagion of information in social media: an experiment using twitter bots. *PLoS One* **12**(9), e0184148 (2017)
61. Mukherjee, A., Liu, B., Galance, N.: Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st International Conference on World Wide Web, pp. 191–200 (2012)
62. Mustafaraj, E., Metaxas, P.T.: From obscurity to prominence in minutes: political speech and real-time search (2010)
63. Pozzana, I., Ferrara, E.: Measuring bot and human behavioral dynamics. arXiv preprint arXiv:1802.04286 (2018)
64. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. *OpenAI Blog* **1**(8) (2019)
65. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A., Menczer, F.: Detecting and tracking political abuse in social media. *ICWSM* **11**, 297–304 (2011)
66. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., Menczer, F.: Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th International Conference Companion on World Wide Web, pp. 249–252. ACM (2011)
67. Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.-C., Flammini, A., Menczer, F.: The spread of low-credibility content by social bots. *Nat. Commun.* **9**(1), 4787 (2018)
68. Shorey, S., Howard, P.N.: Automation, algorithms, and politics! automation, big data and politics: a research review. *Int. J. Commun.* **10**, 24 (2016)
69. Song, J., Lee, S., Kim, J.: Spam filtering in twitter using sender-receiver relationship. In: International Workshop on Recent Advances in Intrusion Detection, pp. 301–317 (2011)

70. Stella, M., Ferrara, E., De Domenico, M.: Bots increase exposure to negative and inflammatory content in online social systems. *Proc. Natl. Acad. Sci.* **115**(49), 12435–12440 (2018)
71. Stieglitz, S., Brachtel, F., Ross, B., Jung, A.-K.: Do social bots dream of electric sheep? a categorisation of social media bot accounts. arXiv preprint arXiv:1710.04044 (2017)
72. Stringhini, G., Kruegel, C., Vigna, G.: Detecting spammers on social networks. In: *Proceedings of the 26th Annual Computer Security Applications Conference*, pp. 1–9. ACM (2010)
73. Stukal, D., Sanovich, S., Bonneau, R., Tucker, J.A.: Detecting bots on russian political twitter. *Big Data* **5**(4), 310–324 (2017)
74. Subrahmanian, V., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A., Menczer, F.: The darpa twitter bot challenge. *Computer* **49**(6), 38–46 (2016)
75. Sutton, J.N., Palen, L., Shklovski, I.: Backchannels on the Front Lines: Emergency Uses of Social Media in the 2007 Southern California Wildfires. University of Colorado, Boulder (2008)
76. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. *J. Am. Soc. Inf. Sci. Tech.* **61**(12), 2544–2558 (2010)
77. Theocharis, Y., Lowe, W., van Deth, J.W., García-Albacete, G.: Using twitter to mobilize protest action: online mobilization patterns and action repertoires in the occupy wall street, indignados, and aganaktismenoi movements. *Inf. Commun. Soc.* **18**(2), 202–220 (2015)
78. Thomas, K., Grier, C., Song, D., Paxson, V.: Suspended accounts in retrospect: an analysis of twitter spam. In: *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, pp. 243–258. ACM (2011)
79. Thomas, K., McCoy, D., Grier, C., Kolcz, A., Paxson, V.: Trafficking fraudulent accounts: the role of the underground market in twitter spam and abuse. In: *Usenix Security*, vol. 13, pp. 195–210 (2013)
80. Varol, O., Ferrara, E., Davis, C., Menczer, F., Flammini, A.: Online human-bot interactions: detection, estimation, and characterization. In: *International AAAI Conference on Web and Social Media* (2017)
81. Varol, O., Ferrara, E., Menczer, F., Flammini, A.: Early detection of promoted campaigns on social media. *EPJ Data Sci.* **6**(1), 13 (2017)
82. Varol, O., Ferrara, E., Ogan, C.L., Menczer, F., Flammini, A.: Evolution of online user behavior during a social upheaval. In: *Proceedings 2014 ACM Conference on Web Science*, pp. 81–90 (2014)
83. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. *Science* **359**(6380), 1146–1151 (2018)
84. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge university press, New York (1994)
85. Yang, C., Harkreader, R., Zhang, J., Shin, S., Gu, G.: Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In: *Proceedings of the 21st International Conference on World Wide Web*, pp. 71–80. ACM (2012)
86. Yang, K.-C., Varol, O., Davis, C.A., Ferrara, E., Flammini, A., Menczer, F.: Arming the public with artificial intelligence to counter social bots. *Hum. Behav. Emerg. Technol.* **1**, e115 (2019)
87. Yang, X., Chen, B.-C., Maity, M., Ferrara, E.: Social politics: agenda setting and political communication on social media. In: *International Conference on Social Informatics*, pp. 330–344. Springer (2016)
88. Yates, D., Paquette, S.: Emergency knowledge management and social media technologies: a case study of the 2010 haitian earthquake. *Int. J. Inf. Manag.* **31**(1), 6–13 (2011)
89. Yin, J., Lampert, A., Cameron, M., Robinson, B., Power, R.: Using social media to enhance emergency situation awareness. *IEEE Intell. Syst.* **27**(6), 52–59 (2012)
90. Zhang, X., Zhu, S., Liang, W.: Detecting spam and promoting campaigns in the twitter social network. In: *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pp. 1194–1199. IEEE (2012)