D. K. Stukal, I. B. Philippov

PROMOTING A LEADER OR A CAUSE? AN AGENT-BASED MODEL OF SOCIAL MEDIA BOTS

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

22

D. K. Stukal, I. B. Philippov
Д. К. Стукал, И. Б. Филиппов

PROMOTING A LEADER OR A CAUSE? AN AGENT-BASED MODEL OF SOCIAL MEDIA BOTS

ПРОДВИЖЕНИЕ ЛИДЕРА ИЛИ ПОПУЛЯРИЗАЦИЯ ПОВЕСТКИ? АГЕНТНО-ОРИЕНТИРОВАННАЯ МОДЕЛЬ ПРИМЕНЕНИЯ БОТОВ В СОЦИАЛЬНЫХ СЕТЯХ

Denis K. STUKAL[1] *— Cand. Sci. (Polit.), Leading Research Fellow, Laboratory for Political Studies, Institute for Applied Political Studies*
E-MAIL: dstukal@hse.ru
https://orcid.org/0000-0001-6240-5714

Ilya B. PHILIPPOV[1] *— PhD Candidate*
E-MAIL: ibfilippov@gmail.com
https://orcid.org/0000-0002-1464-2923

[1] HSE University, Moscow, Russia

СТУКАЛ Денис Константинович — кандидат политических наук, ведущий научный сотрудник научно-учебной лаборатории политических исследований, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия
E-MAIL: dstukal@hse.ru
https://orcid.org/0000-0001-6240-5714

ФИЛИППОВ Илья Борисович — аспирант, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия
E-MAIL: ibfilippov@gmail.com
https://orcid.org/0000-0002-1464-2923

Abstract. Automated social media accounts, a.k.a. social media bots, have been gaining increasing interest among scholars studying human online behavior in recent years. Despite the abundant literature on bots, their substantive effects remain understudied. This paper bridges the existing gap by developing a realistic computational model of human interactions on Twitter, a popular social media platform, that includes leaders, ordinary users, and bots attached to leaders. First, we employ this model to study the effects of bots with different functions on promoting their leader by gaining them extra followers or retweets. Second, we explore the effects of bots on promoting their leader's cause through increasing the volume of tweets with the leader's ideology. We show that bots can be detrimental to the leaders' personal popularity, whereas the effect on cause promotion depends on the distribution of

Аннотация. В исследованиях политической активности в социальной сетях в последнее время все большее внимание уделяется автоматизированным аккаунтам, более известным как боты. Несмотря на обилие работ, посвященных этому явлению, воздействие ботов на социальную сеть и ее пользователей остается недостаточно исследованным. Данная статья направлена на заполнение этой лакуны и предлагает реалистичную вычислительную модель взаимодействий обычных пользователей, политических лидеров и их ботов в популярной социальной сети Twitter. В первую очередь мы используем модель для изучения последствий применения ботов для популяризации лидерского аккаунта с помощью приобретения новых подписчиков и ретвитов. Далее мы исследуем возможности ботов по продвижению лидерской политической позиции

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

23

D. K. Stukal, I. B. Philippov
Д. К. Стукал, И. Б. Филиппов

bots among leaders. These results can be used for developing suitable research designs for further empirical estimation of the effects of bots.

благодаря увеличению общего объема публикаций, распространяющих идеологию лидера. Мы показываем, что применение ботов может вредить личной популярности лидеров, в то время как успехи в продвижении идеологической позиции зависят от распределения ботов между лидерами. В дальнейшем данные результаты могут быть использованы для разработки дизайна эмпирической проверки результатов применения ботов.

Introduction

The effects of social media platforms that were initially and optimistically dubbed as a liberation technology [Diamond, 2010] proved much more nuanced and complex, as these platforms have paved the way not only for new forms of pseudo-activism including slacktivism[1] [Christensen, 2011], but also for novel ways of manipulating public opinion and suppressing civil and political participation through the spread of misinformation, armies of paid trolls, and networks of automated accounts known as bots [Gunitsky, 2015; Tucker et al., 2017; Feldstein, 2019]. The latter technology has spawned a particularly large and diverse body of academic research developing new methods for bot detection [Chavoshi, Hamooni, Mueen, 2016; Davis et al., 2016; Sayyadiharikandeh, 2020], identifying strategies behind the use of bots in diverse contexts [Shao et al., 2018; Uyheng, Carley, 2019]. However, the importance of bots for human online behavior remains understudied. Can bots affect what human users consume on social media platforms? Are bots an effective tool for boosting their creators' ability to reach out to larger online audiences?

Answering causal questions about the effects of bots empirically would require complex experimental designs that may be unfeasible or unethical. This paper takes a different approach by developing a computational agent-based model of human

---

[1] Morozov E. (2009) The Brave New World of Slacktivism. *Foreign Policy*. May 19. URL: https://foreignpolicy.com/2009/05/19/the-brave-new-world-of-slacktivism/ (accessed: 13.02.2022).

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

24

interactions on a social media platform like Twitter. The proposed model does not only allow us to capture important aspects of human online interactions on a popular social media platform or algorithm-induced patterns of human behavior but also to introduce different types of bots in the human network in a controlled way that enables us to measure the effects of bots under different scenarios.

We consider two main scenarios. First, bots are created and stay attached to a leader on only one side of the one-dimensional ideological space. Second, two leaders on different sides of the spectrum are equipped with bots. Both scenarios allow bots to operate under different regimes that involve doing nothing, posting tweets, following other users, or both. We exogenously vary the share of bots in the network and measure distinct metrics that gauge the ability of bots to gain new human followers or retweets to the leader on the one hand, or to contribute to the leader's cause by promoting tweets with her ideology through the network on the other.

We show that the presence of bots on only one side of the ideological spectrum generates qualitatively different results than the availability of bots to both leaders. In particular, we find that sophisticated bots equipped with multiple functions can damage the leader's ability to get human retweets, while assisting the leader in spreading the word through the network. The positive effect however goes away when bots are attached to both leaders, whereas the retweet-suppressing effect remains unchanged.

We make a two-fold contribution to the growing literature on the effects of bots on human behavior in social media environments. First, we develop a realistic and flexible model of human and non-human activity on a social media platform, thereby bridging the gap between empirical and theoretical research on online mobilization. Second, we use computational experiments to reveal and explore a previously understudied trade-off that social media users may face when choosing to launch a network of bots for promotion purposes. We show that different types of bots can be beneficial for the promotion of a cause, but detrimental to the promotion of a leader herself, thereby contributing to the growing body of literature on social media bots and online mobilization alike.

The paper proceeds as follows. Section two discusses the main lines of bot-related academic research and identifies some of the major gaps in the existing literature. Section three describes our computational model. Section four presents our findings. Section five concludes.

Literature review

The study of social media bots started with research on bot detection in the field of computer and data science. To date, scholars have proposed a variety of methods and tools for the automated detection of bots including diverse supervised [Davis et al., 2016; Varol et al., 2017; Stukal et al., 2017; Orabi et al., 2020] and unsupervised [Chavoshi et al., 2016; Wu et al., 2018; Khalil, Khan, Ali, 2020] machine learning techniques. Despite the voluminous literature on this topic, academics have expressed growing concerns about our technological capacity of identifying sophisticated inauthentic accounts that exhibit both automated and human behavior [Cresci et al., 2017; Grimme, 2017; Luceri et al., 2019].

The complexity of the bot-detection task is particularly worrisome, given a plethora of evidence that bots can be employed online for nefarious purposes, including ma-

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

25

nipulating public opinion about important political campaigns [Bastos, Mercea, 2019; Uyheng, Carley, 2019], spreading misinformation and propaganda [Shao et al., 2018], or threatening social and political activists [Treré, 2016]. On the other hand, previous research has also identified some cases of more positive use of bots for coordinating volunteer activities [Savage, Monroy-Hernandez, Höllerer, 2016] or assisting social media users in staying informed about recent news [Diakopoulos, 2019].

The case studies of bot deployment for the public good or public bad have been augmented with research on the activity strategies employed by bots. Empirical experimental research has shown that bots with higher levels of online activity and more developed algorithms for post generation tend to be more successful in gaining and retaining human followers [Freitas et al., 2015; Savvopoulos, Vikatos, Benevenuto, 2018]. It has also been shown that bots may infiltrate the network of social media users by randomly following, mentioning, or replying to other users [Shao et al., 2018]. This type of bot strategy was also highlighted in an agent-based computational model of the spread of information in a network of social media users[2] that revealed the potential superiority of the random targeting strategy over focusing on information hubs. This computational model is one of the few attempts to evaluate the effectiveness of bots in terms of bots' audience or the magnitude of the produced distortions in the network of users or posts. Systematic empirical research on this topic is hindered by both the lack of experimental data and ethical concerns on the one hand, and the complex mix of diverse algorithms that may control the behavior of bots on the other. As was previously claimed, bots would not necessarily be doing something they were told directly. Instead, they may be governed by abstract rules [Hegelich, Janetzko, 2016].

Given these complexities in solving the puzzle about the effects of bots empirically, a new line of theoretical research has emerged that has addressed the conundrum computationally through experiments with agent-based models. Bots have been studied in the spiral of silence context with the somewhat counter-intuitive finding that even small proportions of bots around 5 to 10 percent are able to change the opinion climate in the network [Ross et al., 2019; Cheng, Luo, Yu, 2020]. Alternatively, the activity of bots was also modeled in the context of disinformation spread where some more modest estimates of bot effects have been reported[3] [Beskow, Carley, 2019].

We continue this line of research by developing a realistic agent-based model that captures major aspects of user interactions on Twitter. We then introduce bots into the network and monitor the outcomes of their activity under different settings.

Computational Model

Broadly speaking, our model builds on the idea that a good model of social media communication requires taking into account the indirect nature of communication in social media environments. Indeed, a Twitter user cannot interact with others directly. Instead, all online encounters are mediated by the platform interface that involves multiple screens with diverse content and a limited set of available actions regarding this content. Some of the screens may be personalized for a particular individual,

---

[2] Lou X., Flammini A., Menczer F. (2020) Manipulating the Online Marketplace of Ideas. URL: https://arxiv.org/pdf/1907.06130v1.pdf (accessed: 13.02.2022).

[3] Ibidem.

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

26

whereas others can be identical across users. In the case of Twitter, an example of a personalized screen is the Twitter feed that shows a user the most relevant tweets that have been posted on the platform since the user's most recent login. The relevance of a tweet is measured by the platform's internal algorithms and depends on the user's previous online activity and her position in the network graph (i. e., who follows her and whom she follows). Put it differently, the feed content is unique for every user at all times. On the contrary, other screens, including a user's home page showing her original tweets and retweets, might look (almost) the same for everyone on Twitter.

From this screen-oriented perspective, any public communication in which platform users engage can change the screen content for everyone, because tweeting, retweeting, or commenting modify the user's home page and other users' feeds. In addition, retweeting and liking can also affect internal platform algorithms, thereby making changes in the personalized screens of other users.

Another aspect of mediated communication on Twitter is the central role of a tweet. Indeed, in many cases, Twitter users interact with each other through interacting with a tweet (the only exception being following or unfollowing other users). This mediation is a result of the platform architecture, design, and algorithms.

We build our model around these two aspects of mediated communication. The model does not seek to reproduce all the details of the platform functioning on Twitter but instead captures its fundamental characteristics. In particular, we model human interactions with the screens that are governed by internal algorithms that are in turn affected by users' activity. The key screen in our model is the feed. Users can interact with the tweets they can see in their feeds by retweeting (or not retweeting) those tweets and following or unfollowing their authors. All these actions make changes in the screens available to other users by changing the inputs for the algorithms that control individual feeds. In addition, users' feeds can get affected by bots, i. e., pseudo-users whose activity is controlled by algorithms. As the goals of bot creation and the strategies behind their deployment can be diverse, our model allows for bots with different types of functionalities.

Overall, the model includes three types of actors that are qualitatively and quantitatively different. First, we introduce ordinary users. Every time they are active (not necessarily at every iteration of the model), they can read their feed, look through and decide whether to follow back new followers, make a retweet or post an original tweet, follow, or unfollow another user based on her post.

Second, our model features leaders. These actors stand out among ordinary users due to their numeric characteristics. In particular, they get activated more often and can read or post larger volumes of tweets. Although our model allows for any number of leaders, we focus on the case of two leaders here for ease of presentation.

Finally, we introduce bots into the network. These actors can only get created by a leader, whom these bots follow. Besides, all bots of the same leader follow each other. Bots have the same activity characteristics as their leader, but the specific types of actions available to bots may be different and are controlled by a model hyperparameter that was introduced into the model in order to better understand the effects of bots. The available action types include tweeting or retweeting on the one hand, and random following of other users in order to get a reciprocal follow request

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

27

on the other. We implement this idea of random mutual followings within a probabilistic framework by introducing the 0.1 probability of a follow back. Hence, there are in total four combinations of bot functions ranging from no activity whatsoever to both functions activated. Interestingly, the no-activity bots can also play a role in the network, as they increase the number of followers their leader has, thereby potentially affecting the visibility of the leader.

A model run starts with a selection of a set of exogenous hyperparameters (shown in table 1). We then generate the network of users as a Barabasi-Albert random graph [Albert, Barabási, 2002] so that the nodes could be divided into a few elite nodes with a large number of followers and a large number of ordinary nodes (users). The resulting graph is only a starting point for our model, as the graph does not include bots at this stage. Besides, all ordinary users are very similar at initialization and have no features but the number and list of followers. However, this initial stage of network generation allows the model to identify leaders as top-2 users in terms of the numbers of followers. Once the two leaders are identified, they are also assigned ideological positions that are controlled with the model hyperparameter *leader_positions*. Then, once ideology was assigned to leaders, it also gets assigned to ordinary users. Every user gets a position between −1 and 1. The assignment process is sequential and makes sure that the users who follow those with negative positions would not receive high positive values. Put it differently, the assignment mechanism reproduces user homophily.

*Table 1. **Model hyperparameters***

Hyperparameter	Description
m	Number of outgoing edges network nodes have at initialization
numleaders	Number of leaders
totalP	Total number of users in the network
P	Number of non-bots in the network
b	Share of bots
leader_positions	Leaders' ideological positions
botshare_first	The share of bots attached to the first leader among all bots
alpha	Parameter controlling the distribution of activity levels over users
alpha2	Processing capacity parameter that controls the size of the processed feed
max_passiveness	Parameter that controls the minimum activity level
lifespan	Tweet lifespan
action_probability	Probability of tweeting; probability of following the author of the retweeted post; twice the probability of unfollowing a user
leader_boost	Leader's extra bonus to the probability of tweeting
tolerance	Maximum tolerated difference in ideological positions
steps	Number of iterations during a model run

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

28

Besides ideology, every user is characterized by the maximum time of inactivity and the processing capacity, i. e., the number of tweets they can read in one login session. These values are sampled from a power distribution so that only a few users receive high values of both features. In order to let the leaders stand out among ordinary users, they get particularly high values of two features through a special leaders' bonus value.

A user's activity status at any model iteration is probabilistic and depends on the number of the previous iterations when the user was not active (tracked with the clock parameter — see below) and the passiveness hyperparameter. For a leader, this probability gets a bonus boost. Overall, the probability of user activity is computed as follows:

$$P(activity) = \begin{bmatrix} \dfrac{1}{(passiveness - clock + 1)} | leader = 0 \\ \dfrac{1}{(passiveness - clock + 1)} + leader\_boost | leader = 1 \end{bmatrix}$$

where the clock parameter is set to zero if the user is active at the current model iteration, otherwise it increases by one.

Network initialization concludes with the inclusion of bots that follow one or both leaders and have the same ideology, activity, and processing capacity. The proportion of bots that are assigned to each leader is controlled with a model hyperparameter.

Once bots are added, the model is ready for computational experiments. Each model iteration involves the following steps (except steps 6 and 8 that can be skipped depending on the regime in which bots operate in):

(1) All previous unseen tweets are removed from user feeds.

(2) A user undergoes an activity check. If it does not pass the check, the following steps are not made for this node.

(3) A new user-specific feed is formed out of the set of all tweets available to the user. The user will read the content of the feed up to the user's processing capacity.

(4) If the user is not a leader or a bot, she can follow a new follower back with a given probability (0.1 in this paper).

(5) The user reads the feed up to her processing capacity, can follow or unfollow the authors of the read tweets, and can choose a tweet for retweeting.

(6) If the user is a bot that has not selected any tweet for retweeting, it randomly selects an old tweet of this leader for retweeting.

(7) The user posts an original tweet with a given probability fixed at 0.5 here.

(8) If the user is a bot, it can randomly follow a new user.

The sorting of the posted tweets in a feed depends on two different values. First, each tweet receives a value that reflects its objective characteristics (e. g., the popularity of the tweet and its author's metadata). This value — referred to as score — is particularly important when the number of tweets that are eligible into a feed exceed the user's processing capacity. In this case, tweets are sorted in the descending order based on this value, which thereby helps identify the tweets a user will actually read. The score is measured as follows:

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

29

$$\textbf{score} = n \left(\frac{\textbf{indegree}}{\textbf{max\_indegree} + 1} + \frac{\textbf{retweets}}{\textbf{max\_retweets} + 1} + \frac{\textbf{subscribed\_followings}}{\textbf{max\_subscribed\_followings} + 1} \right),$$

where n is the number of times the tweet could have entered into the feed through different channels; *indegree* is the number of followers the author of the tweet has; *retweets* is the number of times this tweet has been seen by other users; *subscribed_ followings* is the number of nodes that are followed by the user and follow the author of the tweet; prefix *max_* refers to the maximum values of the respective parameters in the whole feed.

The other value is more subjective and reflects a user's perception of a tweet. At step five of a model iteration, a user selected tweets for retweeting based on a value defined as follows:

$$\textbf{value} = \left(1 - \sqrt{(\textbf{position}_{\textbf{reader}} - \textbf{position}_{\textbf{tweet}})^2} \right) \times \left(\ln(\textbf{indegree} + 1) + 1 \right).$$

We refer to this value as utility. Importantly, in addition to scoring each tweet based on its utility, the user also checks if the absolute difference between $\textbf{position}_{\textbf{reader}}$ and $\textbf{position}_{\textbf{tweet}}$ falls below a threshold that is set as a hyperparameter. If it exceeds the threshold, the user unfollows the author of the tweet or the retweet maker with probability 0.25. Finally, with probability 0.5 the tweet with the largest utility gets retweeted and with the same probability the user starts following its author.

We implement this computational model in Python 3 and use it to perform a set of computational experiments in order to better understand the potential mechanisms behind the effects of bots on online political mobilization.

Model results

This section presents the results of a series of computational experiments performed using our model. In all the experiments, the first leader was assigned the ideological position of −0.5, whereas the other leader is ideologically located at 0.5.

While the leaders' ideological positions remain fixed throughout our experiments, we vary the share of bots in the network starting with the no-bots situation (share = 0) and up to the case where bots make up half the platform population (share = 0.5). This wide range of values considered for the share of bots is motivated by previous empirical research on the proportion of automated accounts on Twitter showing that the proportion of bots may vary dramatically depending on the national context and the segment of Twitter under study. In particular, existing empirical estimates range between under 10 percent and up to 50 percent [Chu et al., 2012; Subrahmanian et al., 2016; Stukal et al., 2017].

In addition to varying the share of bots across computational experiments, we also consider different action types available to bots. There are four main regimes of bot operation: bots posting original tweets or retweeting their leader, bots randomly following other users, bots doing all these things together, or not doing anything at all.

Model experiments with a given set of hyperparameters (including a preset share of bots and a bot regime) are referred to as model runs and are repeated 26 times with different starting values of the pseudo-random number generator (random seeds). Each model run includes 500 model iterations.

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

30

In order to evaluate the effects of bots under different regimes, we measure multiple performance metrics that represent distinct goals that could potentially be achieved with the use of bots. First, we measure the ability of bots to reach out to humans by measuring the number of people subscribed to bots. Although this metric is bot-centered and might not seem substantively interesting, we report it as it is often the focus in experimental research on bots [Freitas et al., 2015; Savvopoulos et al., 2018].

Second, we measure the ability of bots to promote the leader's ideology. For this purpose, we measure the distance between the leader's ideology and the average ideological position of the tweets that human users read at the last 100 iterations of each model run.

Finally, we measure the number of human followers the first leader has, and the number of times human users retweeted this leader. These two metrics aim to measure the effectiveness of bots in boosting the leader's personal popularity and providing her with extra resources for reaching out and mobilizing her audiences.

Below, we report these performance metrics averaged across random seeds and account for the random variance of these metrics via 95 percent Gaussian confidence intervals. Figure 1 reports the results for the four main bot regimes. The top left panel of Figure 1 shows the bot-centered performance metrics and reveals that pure random following is the best bot strategy for achieving and retaining followers. Interestingly, this bot regime outperforms the regime of the full-fledged bot activity that allows bots to use all their functions. The rationale behind this finding is as follows. Human users make probabilistic decisions to unfollow someone if that user's post (either an original tweet or retweet) is too dissimilar from their own ideological positions. However, if bots cannot tweet or retweet and can only randomly follow other users in order to get follow back requests, human users in our model will not have a chance to notice any ideological dissimilarities between themselves and the bots. Thereby, the share of human users who follow bots attains the maximum if bots can only send follow requests and there are enough bots in the network. At the same time, when all functions are available to bots, users are actually able to observe bots' ideological positions through tweets and retweets; hence, the metric can hardly exceed 50 percent of human users who are located on the left-hand side of the ideological spectrum close to the first leader. Nevertheless, the performance metric for the full-fledged regime is statistically better than tweeting/retweeting. As one can see from this result, random following can get extra followers to bots.

The right panels of Figure 1, both the top and bottom ones, present two different metrics related to the ability of bots to promote the leader's ideology. Unless bots can tweet or employ the full functionality, the average ideological position of the tweets retweeted by human users is located around zero, which is the center of the ideological spectrum (thus, the 0.5 distance from leader 1). However, if bots can tweet, they are able to amplify the leader's ideology. The more bots the network has, the smaller the distance between the average human-retweeted tweet and the first leader's position. The effect of the share of bots is particularly strong in the case of the full-fledged bot activity.

A substantively similar result can be inferred from the bottom right panel that shows the average number of retweets per capita that are posted by the users located on the left-hand side of the ideological spectrum, i.e., closer to leader 1. Here again, one can see that the number of tweets increases in the share of bots if these latter use the full functionality.

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

31

Importantly, these findings are substantially different from what one could infer from the top left panel. If the latter signals that silent bots are the best option for maximizing the network exposure to bots, the former goes beyond pure bot exposure and reveals other types of bots may be superior for the purposes of promoting a cause online.

The bottom left panel of figure 1 takes yet another perspective and looks at what bots can give the leader herself in terms of the number of retweets. This panel reveals a negative effect of the share of bots on the number of retweets that leader 1 gets in the case of tweeting bots; this negative effect persists regardless of whether bots can only tweet or combine this ability with other functionalities. This negative effect is driven by the fact that tweeting bots introduce extra tweets into the feeds of human users thereby distracting them and decreasing their ability to retweet the leader. From this perspective, even though having more tweeting bots does not damage the size of the leader's audience, it may have a negative impact on the leader's capability of getting her online audiences engaged.

Fig. 1. The effects of bots with different functionality (single-leader bots)
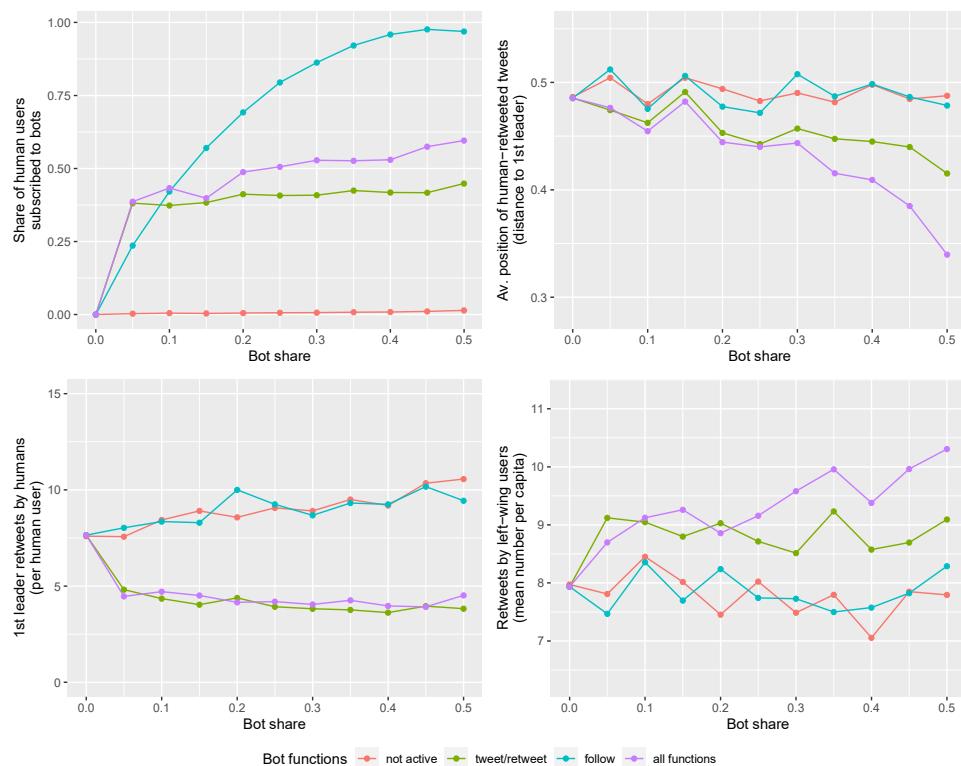
Summarizing the findings from figure 1, one can infer that bot deployment creates a trade-off for a leader. The most sophisticated bots can produce a strong boost to the leader's cause by exposing larger audiences to the cause-related message. However,

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

32

the same types of bots can substantially undermine the leader's ability to spread her voice through human retweets.

One significant limitation of these findings is due to the presence of bots on only one side of the ideological spectrum. In many situations, including the Russian political context [Stukal et al., 2019], bots are deployed on both sides of the spectrum. We now turn to this more realistic scenario and consider the case of bots attached to one of the two leaders. The results for this case are presented in figure 2.

Fig. 2. The effects of bots with different functionality (bots attached to both leaders)

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

33
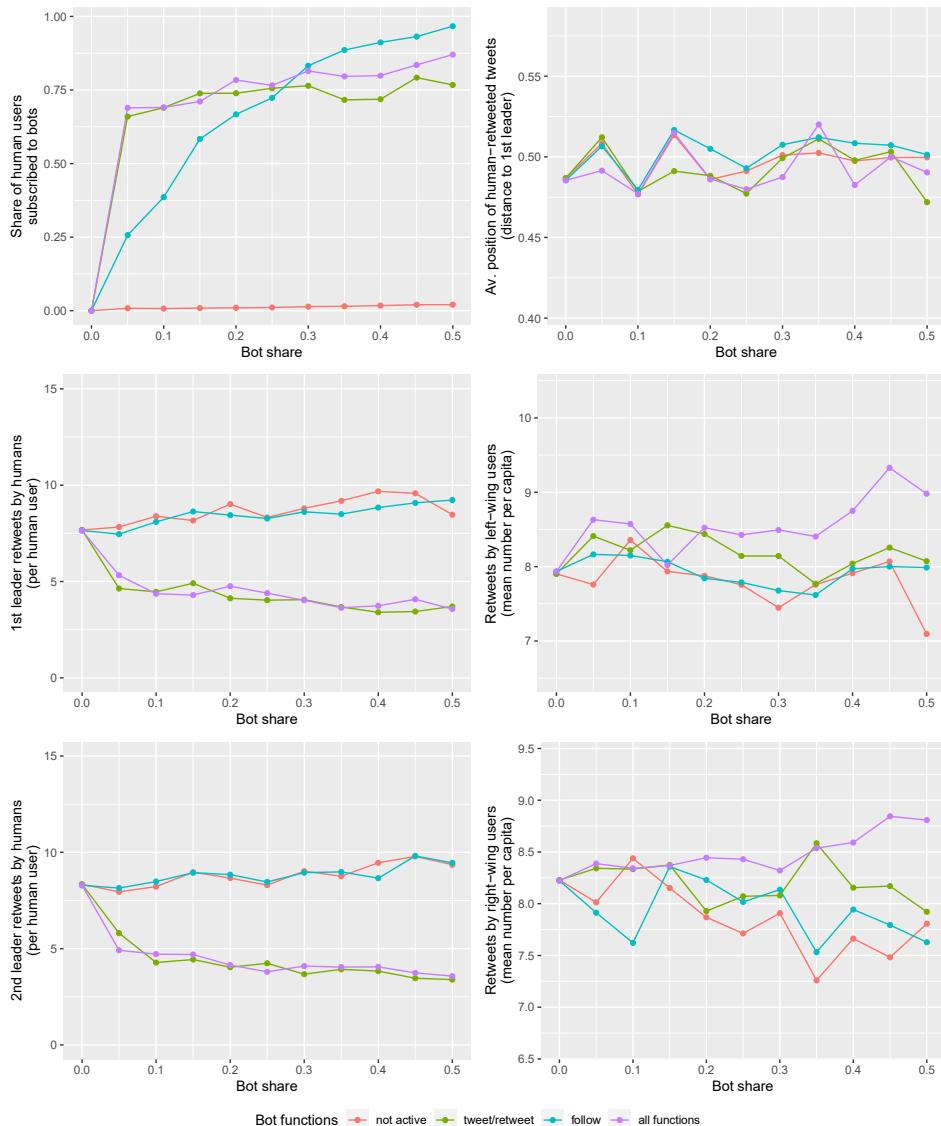
The top left-hand side panel corroborates our discussion of the results of the previous figure. As it was also the case before, the pure random following makes it possible for bots to achieve the maximum possible audience when enough bots are present. Tweeting bots and bots with the full-fledged activity, however, demonstrate even more impressing results than bots with the random following functionality. Indeed, all these bots could hardly exceed 50 percent on the left-hand side of figure 1, but they achieve almost complete coverage of platform users when bots are located at both sides of the spectrum. As it was also the case before, inactive bots remain unfollowed by human users.

The ability of bots to promote the first leader's cause (shown on the right-hand side of the top two rows in figure 2) gets trumped by the activity of the other leader's bots. In fact, the distance between the average ideological position of the tweets consumed by human users and the position of the first leader remains basically stable no matter what the share of bots is. The situation is identical also for the second leader (see the bottom right panel in figure 2).

The personal boosts that either leader can get from deploying bots (shown on the middle and bottom left panels in figure 2) reveal a very similar pattern to what was inferred from figure 1. In particular, one can see a detrimental effect of tweeting bots and the full-fledged bot activity on either leader.

Thus, this more realistic case reveals that even though the deployment of bots is not necessarily useful for promoting the leader herself, this might work as a defense strategy against the cause-promotion bot effects described in figure 1. Overall, the benefits of having bots depend on what the goal of bot deployment is. Bots can indeed help increase the number of followers (although only to a very limited extent) but may be unhelpful or even harmful for boosting leaders' retweets.

Conclusion

Social media bots have become a common element of the social media environment. Previous studies have shown that bots can exhibit large variation in their sophistication, activity levels, or types of produced content. Large bodies of literature exist on the technologies of bot detection; a plethora of studies have documented cases of the use of bots for commercial and political purposes in a number of countries. What remains however unclear is whether bots actually matter. Although the empirical puzzle is yet to be solved, this paper makes a two-fold contribution to the studies of the effects of bots on human behavior in social media environments.

First, we develop a realistic and flexible computational agent-based model of politically relevant interactions on Twitter. We consider three types of users, including leaders, ordinary users, and bots. All the users are assigned ideological positions and some tolerance towards ideological dissimilarity. Users can tweet, retweet, follow or unfollow other users based on their ideological positions. Bots are introduced in this network as attachments to leaders with different types of functionalities. We then vary the share of bots in the network and types of actions available to bots to see how the presence of different types of bots can change the metrics that may be relevant to the leaders.

Second, we use computational experiments to reveal and explore a previously understudied trade-off that social media users may face when choosing to launch

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

34

a network of bots for promotion purposes. In particular, we focus on two cases. In the first case, we consider the bots attached to one leader only. After running a series of computational experiments, we show that highly sophisticated bots and bots that are able to tweet or retweet can contribute to the leader's cause by making it more visible to the network audiences. However, the same types of bots can harm the leader's ability to engage with the audience, as the number of retweets the leader gets decreases in the share of bots. Hence, the deployment of bots creates a trade-off for a leader who would need to choose whether to promote the cause or herself.

In the second case, we consider a polarized situation with two leaders both having bots attached. In this case, the ability of bots to promote the leader's cause disappears. However, the same types of bots are still able to make it harder for a leader to get retweets. What is common for both cases is the small positive effect of bots on the leaders' ability to get extra followers, but the size of the effect is close to trivial.

Our results highlight some of the fundamental challenges for empirical research focused on measuring the effects of bots. One of these challenges is the dependence of the effects of bots on the network ecosystem. Introducing bots on the other side of the ideological spectrum resulted in important changes in our results. As the potential real-life cases of bot deployment are much more diverse than the ones we have considered in this paper, the empirical findings can be hard to generalize beyond the sample under study.

Another challenge is due to the variety of bots. Although this paper considered ideal-type situations with all bots having a predefined set of functions on, real-life situations would typically witness bots with diverse levels of sophistication and less clear patterns.

Further research is required to develop convincing empirical designs that would allow researchers to tease out the effects of the distinct types of bots on diverse groups of audiences and provide empirical measures of the effectiveness of bots as an amplification technology.

References

Albert R., Barabási A. (2002) Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*. Vol. 74. No. 1. P. 47—97. https://doi.org/10.1103/RevModPhys. 74.47.

Bastos M. T., Mercea D. (2019) The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*. Vol. 37. No. 1. P. 38—54. https://doi.org/ 10.1177/0894439317734157.

Beskow D. M., Carley K. M. (2019) Agent Based Simulation of Bot Disinformation Maneuvers in Twitter. In: *2019 Winter Simulation Conference (WSC)*. National Harbor, MD: IEEE. P. 750—761. https://doi.org/10.1109/WSC40007.2019.9004942.

Chavoshi N., Hamooni H., Mueen A. (2016) DeBot: Twitter Bot Detection via Warped Correlation. In: *2016 IEEE 16th International Conference on Data Mining (ICDM)*. Barcelona: IEEE. P. 817—822. https://doi.org/10.1109/ICDM.2016.0096.

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

35

Cheng C., Luo Y., Yu C. (2020) Dynamic Mechanism of Social Bots Interfering with Public Opinion in Network. *Physica A: Statistical Mechanics and its Applications*. Vol. 551. https://doi.org/10.1016/j.physa.2020.124163.

Christensen H. S. (2011) Political Activities on the Internet: Slacktivism or Political Participation by Other Means? *First Monday*. Vol. 16. No. 2. https://doi.org/10.5210/fm.v16i2.3336.

Chu Z., Gianvecchio S., Wang H., Jajodia S. (2012) Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? *IEEE Transactions on Dependable and Secure Computing*. Vol. 9. No. 6. P. 811—824. https://doi.org/10.1109/TDSC.2012.75.

Cresci S., Di Pietro R., Petrocchi M., Spognardi A., Tesconi M. (2017) The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race. In: *WWW'17 Companion: Proceedings of the 26th International Conference on World Wide Web Companion.* Geneva: International World Wide Web Conference Steering Committee. P. 963—972. https://doi.org/10.1145/3041021.3055135.

Davis C. A., Varol O., Ferrara E., Flammini A., Menczer F. (2016) BotOrNot: A System to Evaluate Social Bots. In: *WWW'16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web*. Geneva: International World Wide Web Conference Steering Committee. P. 273—274. https://doi.org/10.1145/2872518.2889302.

Diakopoulos N. (2019) Automating the News: How Algorithms Are Rewriting the Media. Cambridge, MA: Harvard University Press.

Diamond L. (2010) Liberation Technology. *Journal of Democracy*. Vol. 21. No. 3. P. 69—83.

Feldstein S. (2019) The Road to Digital Unfreedom: How Artificial Intelligence Is Reshaping Repression. *Journal of Democracy.* Vol. 30. No. 1. P. 40—52.

Freitas C., Benevenuto F., Ghosh S., Veloso A. (2015) Reverse Engineering Socialbot Infiltration Strategies in Twitter. In: *ASONAM'15: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. New York, NY: Association for Computing Machinery. P. 25—32. https://doi.org/10.1145/2808797.2809292.

Grimme C., Preuss M., Adam L., Trautmann H. (2017) Social Bots: Human-Like by Means of Human Control? *Big Data*. Vol. 5. No. 4. P. 279—293. https://doi.org/10.1089/big.2017.0044.

Gunitsky S. (2015) Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability. *Perspectives on Politics*. Vol. 13. No. 1. P. 42—54. https://doi.org/10.1017/S1537592714003120.

Hegelich S., Janetzko D. (2016) Are Social Bots on Twitter Political Actors? Empirical Evidence from a Ukrainian Social Botnet. In: *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM).* Palo Alto, CA: AAAI Press. P. 579—582.

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

36

Khalil H., Khan M. U., Ali M. (2020) Feature Selection for Unsupervised Bot Detection. In: *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET).* Piscataway, NJ: IEEE. P. 1—7. https://doi.org/10.1109/iCoMET48670.2020.9074131.

Luceri L., Deb A., Giordano S., Ferrara E. (2019) Evolution of Bot and Human Behavior During Elections. *First Monday*. Vol. 24. No. 9. https://doi.org/10.5210/fm.v24i9.10213.

Orabi M., Mouheb D., Al Aghbari Z., Kamel I. (2020) Detection of Bots in Social Media: A Systematic Review. *Information Processing & Management*. Vol. 57. No. 4. https://doi.org/10.1016/j.ipm.2020.102250.

Ross B., Pilz L., Cabrera B., Brachten F., Neubaum G., Stieglitz S. (2019) Are Social Bots a Real Threat? An Agent-Based Model of the Spiral of Silence to Analyse the Impact of Manipulative Actors in Social Networks. *European Journal of Information Systems*. Vol. 28. No. 4. P. 394—412. https://doi.org/10.1080/0960085X.2018.1560920.

Savage S., Monroy-Hernandez A., Höllerer T. (2016) Botivist: Calling Volunteers to Action Using Online Bots. In: *CSCW'16: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. New York, NY: Association for Computing Machinery. P. 813—822. https://doi.org/10.1145/2818048.2819985.

Savvopoulos A., Vikatos P., Benevenuto F. (2018) Socialbots' First Words: Can Automatic Chatting Improve Influence in Twitter? In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. Piscataway, NJ: IEEE. P. 190—193. https://doi.org/10.1109/ASONAM.2018.8508786.

Sayyadiharikandeh M., Varol O., Yang K.-C., Flammini A., Menczer F. (2020) Detection of Novel Social Bots by Ensembles of Specialized Classifiers. In: *CIKM'20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. New York, NY: Association for Computing Machinery. P. 2725—2732. https://doi.org/10.1145/3340531.3412698.

Shao C., Ciampaglia G. L., Varol O., Yang K.-C., Flammini A., Menczer F. (2018) The Spread of Low-Credibility Content by Social Bots. *Nature Communications*. Vol. 9. P. 1—9. https://doi.org/10.1038/s41467-018-06930-7.

Stukal D., Sanovich S., Bonneau R., Tucker J. A. (2017) Detecting Bots on Russian Political Twitter. *Big Data*. Vol. 5. No. 4. P. 310—324. https://doi.org/10.1089/big.2017.0038.

Stukal D., Sanovich S., Tucker J. A., Bonneau R. (2019) For Whom the Bot Tolls: A Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia. *Sage Open*. Vol. 9. No. 2. P. 1—16. https://doi.org/10.1177/2158244019827715.

Subrahmanian V. S., Azaria A., Durst S., Kagan V., Galstyan A., Lerman K., Zhu L., Ferrara E., Flamini A., Menczer F. (2016) The DARPA Twitter Bot Challenge. *Computer*. Vol. 49. No. 6. P. 38—46. https://doi.org/10.1109/MC.2016.183.

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167) январь — февраль 2022

37

Treré E. (2016) The Dark Side of Digital Politics: Understanding the Algorithmic Manufacturing of Consent and the Hindering of Online Dissidence. *IDS Bulletin.* Vol. 47. No. 1. P. 127—138. https://doi.org/10.19088/1968-2016.111.

Tucker J. A., Theocharis Y., Roberts M. E., Barberá P. (2017) From Liberation to Turmoil: Social Media and Democracy. *Journal of Democracy*. Vol. 28. No. 4. P. 46—59.

Uyheng J., Carley K. M. (2019) Characterizing Bot Networks on Twitter: An Empirical Analysis of Contentious Issues in the Asia-Pacific. In: Thomson R., Bisgin H., Dancy C., Hyder A. (eds.) *Social, Cultural, and Behavioral Modeling.* Cham: Springer. P. 153—162. https://doi.org/10.1007/978-3-030-21741-9_16.

Varol O., Ferrara E., Davis C. A., Menczer F., Flammini A. (2017) Online Human-Bot Interactions: Detection, Estimation, and Characterization. In: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)*. Vol. 11. No. 1. P. 280—289. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14871 (accessed: 13.02.2022).

Wu W., Alvarez J., Liu C., Sun H. M. (2018) Bot Detection Using Unsupervised Machine Learning. *Microsystem Technologies.* Vol. 24. P. 209—217. https://doi.org/10.1007/s00542-016-3237-0.

Monitoring of Public Opinion: Economic and Social Changes
Мониторинг общественного мнения: экономические и социальные перемены

No. 1   January — February 2022
№ 1 (167)   январь — февраль 2022

38