ORIGINAL ARTICLE

Collective Behavior of Social Bots Is Encoded in Their Temporal Twitter Activity

Andrej Duh,^{1,2} Marjan Slak Rupnik,^{2,3} and Dean Korošak^{1,2,4,*}

Abstract

Computational propaganda deploys social or political bots to try to shape, steer, and manipulate online public discussions and influence decisions. Collective behavior of populations of social bots has not been yet widely studied, although understanding of collective patterns arising from interactions between bots would aid social bot detection. In this study, we show that there are significant differences in collective behavior between population of bots and population of humans as detected from their Twitter activity. Using a large dataset of tweets we have collected during the UK-EU referendum campaign, we separated users into population of bots and population of humans based on the length of sequences of their high-frequency tweeting activity. We show that, while pairwise correlations between users are weak, they co-exist with collective correlated states; however the statistics of correlations and co-spiking probability differ in both populations. Our results demonstrate that populations of social bots and human users in social media exhibit collective properties similar to the ones found in social and biological systems placed near a critical point.

Keywords: Twitter bots; collective behavior; spin glass models; correlations; criticality

Introduction

Social bots are automated user accounts in online social networks owned and used by computers.^{1–3} Social media, such as Twitter or Facebook, which support high spreadability and convergence of content,^{4–6} particularly during influential political events,^{7–10} are particularly susceptible environments for such entities. Social bots have been successfully deployed in political context in social media despite their current low level of sophistication, but applying advances in artificial intelligence¹¹ and/or hybridization with human users into cyborgs^{12,13} can soon dramatically increase their human-like properties, coordination, and interaction potential, making them extremely difficult to detect and classify into strictly binary bot/human class.

Bots in the sense of software agents are not specific to online social media; they are present in many forms on the Internet, both with benevolent or malevolent intent.^{14,15} Although they are classified using different

taxonomies with different categories,^{13,15–17} bots are essentially autonomous agents as defined by Franklin and Graesser¹⁶:

“An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.”

To characterize a bot following this definition, one must describe its¹⁶

- environment,
- sensing,
- possible actions,
- drives, and
- decision-making.

Applying this description to, for example, a social bot on Twitter, we identify the environment in which such an entity exists as microblogging platform, where bots sense the content (tweets) and other users' data, and

¹Percipio, Ltd., Maribor, Slovenia.

²Faculty of Medicine, Institute of Physiology, University of Maribor, Maribor, Slovenia.

³Center for Physiology and Pharmacology, Institute for Physiology, Medical University of Vienna, Vienna, Austria.

⁴Faculty of Civil Engineering, Transportation Engineering and Architecture, University of Maribor, Maribor, Slovenia.

*Address correspondence to: Dean Korošak, Faculty of Medicine, Institute of Physiology, University of Maribor, Maribor SI-2000, Slovenia, E-mail: dean.korosak@um.si

can act with posts, retweets, and by following actions. The preferences of social bots are encoded intents that result in specific response and behavior of bots through a decision-making process. Starting from a general definition of bots as autonomous computational agents and using their characteristics as dimensions, various classification schemes and taxonomies for bots,^{14–16,18} and specifically for social bots, were derived and proposed.^{13,17,19,20} Stieglitz et al.¹⁷ introduced a two-dimensional categorization scheme with intent (drive) and imitation of human behavior (possible actions and decision-making) as variables. Gorwa and Giulbeault²⁰ instead of using categorical typology suggest a framework of three considerations for understanding bots: structure (combining environment and sensing), function (action), and use (drive and decision-making process).

Computational propaganda uses of politicized social bots in different roles, of malicious or of more mild nature,²¹ have been discovered in dissimilar political systems.²² Their effect on public opinion^{8,23–25} makes them important political actors and their analyses should be included in political communication study.²⁶ Social bots can manipulate, influence, and steer communication in social media or can also find themselves manipulated by human users.^{27,28} They were found to help spreading fake news,^{29,30} influence stock indicators,³¹ and serve as fake followers in promoting social media profiles.³² Grimme et al.¹³ proposed that the amplitude of influence of users in social media depends on two variables: orchestration and automation of users, suggesting that network of hybrid bots, agents that mimic temporal human behavioral patterns, while content is produced by humans, has the highest potential to influence social media.

An important computational task is the recognition, classification, and early detection of social bots using features extracted from user network, data, and metadata,^{12,19,33–36} a task that is getting more difficult as bots are getting better at mimicking human online activity and behavior, becoming more sophisticated utilizing artificial intelligence¹¹ and hybridization.¹³ Artificial intelligence methods such as deep neural networks are on the other hand employed to detect social bots.³⁷ One of the most sophisticated systems for detection and recognition of social bots on Twitter is BotOrNot³³ that uses more than 1000 features to classify a user as a human like or bot like. These features are grouped into six classes³³: network (statistical network features computed from retweet, mention, and hashtag

co-occurrences), user (account metadata), friends (statistics of social contacts), content (language features), sentiment, and temporal (tweet rate and intertweet time distributions).

Time-like features such as relationships between events,³⁸ combination of interarrival times and volume of events,³⁹ correlations,^{40,41} and time interval entropy⁴² were used to study temporal characteristics of individual users. In this study, we also focus solely on the temporal properties of Twitter users' activities, however, not to use them primarily as individual filter for bot/human classification, but to try to link the temporal features of activity to social aspects of bots and their collective behavior. Recently, the behavioral dynamics of bots and humans during activity in Twitter session was studied⁴³ to detect differences in bot and human behavior. We are not looking for individual differences between bots and humans, instead we are interested in population of heterogeneous social bots gathered around specific topic or event and their group dynamics, making a first step toward detecting and quantifying collective behavior of bots.

In this study, we show, using and analyzing Twitter data collected during the campaign around UK-EU referendum, that groups of social bots can be detected by specific temporal traces they leave in their tweeting activity. Although collective behavior in populations of bots and humans statistically differs, both groups have weak pairwise correlations co-existing with strongly coordinated states. We demonstrate that a simple Ising spin-glass model with random interactions and fields captures main features of the collective patterns such as scaling of average activity of users when they are represented as interacting spins with temporal activities binarized into spike trains. We discuss the possibility of adaptive behavior of population of social bots and their control of criticality.

Materials and Methods

Between March and September 2016, we used Twitter public API to track and collect tweets containing the word “brexit.” We collected timestamps, user IDs, tweets texts, hashtags, and URLs mentioned for each tweet that matched the search term and stored the tweets in the database for later analysis. We wrote custom software in Java for collecting and storing the tweets and set up Apache Cassandra database to store the tweets. For all the analyses and computations presented in this study, we developed custom Python scripts.

FIG. 1. Left panel—probability distribution of user activity (complete data). Scaling relationship is approximately described by $P(n) \sim n^{-a}$ with $a = 1.7$ (solid line). Right panel—rank distribution of tweetstorm lengths of top 10,000 most active users. In this study, the exponents of power plots (full lines) are 0.35 in the upper part and 0.95 in the tail.

Altogether we collected 33,145,488 tweets published by 4,658,780 unique users. We found that the distribution of user activity (number of tweets published by particular user, n) is heavy tailed and it can be approximately described by a power law probability distribution $P(n) \propto n^{-a}$, with the exponent $a \approx 1.7$ as shown in the left panel of Figure 1.

This shape of the distribution of user activity indicates that the majority of the collected tweets originate from a relative small subpopulation of users, while the large majority of users published only small number of tweets in the observed time period. For our analyses, we looked for a population of users in which each user has tweeted at least twice per day. This criterion led to population of the top 10^4 most active users (0.2% of all users) who published 25% of all tweets. An average user in this sample published approximately five tweets per day.

Within this sample, we looked for users that we could classify as bots or humans based on their tweeting activity. To this end, we defined a *tweetstorm*—a quantity describing a tweet sequence where time difference between consecutive tweets is less than 10 epoch timestamps (we converted timestamps given by Twitter API to epoch timestamps). We computed all tweetstorms for all users in the sample. When plotted as a rank distribution,

the tweetstorm lengths, w , follow a power-law shape with scaling law $P(w) \propto w^{-0.35}$ in the upper part and a Zipf-like distribution with $P(w) \propto w^{-0.95}$ in the lower part as shown in the right panel of Figure 1.

We sorted the users according to number of tweetstorms and length of tweetstorms in descending order. The users found in the top 5% of the intersection of both sorted lists were considered candidates for bot population, while the candidates for human population came from the bottom 5% of the intersection of both sorted lists. The final populations of 125 bots and 218 humans consisted of users which user ID we could identify through BotOrNot API (accessed in January and February 2017) and for which BotOrNot score we could obtain.*

To assay the collective behavior of bot and human populations, we chose to represent bot and human tweeting activity with spin variables $S_i(t)$. Each timeline of user's activity was transformed into a spike train with $S_i(t) = +1$ if i -th user has tweeted within the time interval $(t, t + \Delta t)$, and $S_i(t) = -1$; if not, we used bin width $\Delta t = 30$ minutes to binarize the tweeting activity.

*Available online at <http://truthy.indiana.edu/botornot/> and <https://github.com/truthy/botornot-python> (last accessed January 2017). This project has now moved to the webpage <https://botometer.iui.iu.edu/#/> and the Python API is now at <https://github.com/IUNetSci/botometer-python>

In this study, we were interested in three quantities that characterize collective state of spin populations: spin-spin correlations, co-spiking probability, and the average state of spins.

Spin-spin correlation coefficient is defined as follows:

$$c(i, j) = \langle S_i(t)S_j(t) \rangle - \langle S_i(t) \rangle \langle S_j(t) \rangle \quad (1)$$

where $\langle \rangle$ denote time averages. Average state of the system of N spins is the sum over all spin states:

$$m(t) = \frac{1}{N} \sum_j S_j \quad (2)$$

Existence of weak correlations between spins found in each population and co-spiking probability distributions greatly differing from independent model (in detail displayed and discussed in the Results section) led us to use spin glass model to try to describe and understand these results. A spin glass is a collection of interacting spins where the interaction between spins is a random quantity.⁴⁴ In this model, we have N spins with $S_i(t) = \pm 1$ at time t . At the next moment $(t+1)$, each spin updates its state according to the probability rule:

$$\begin{aligned} S_i(t+1) &= +1, \text{ with probability } p \\ S_i(t+1) &= -1, \text{ with probability } 1-p \end{aligned} \quad (3)$$

where the probability p depends on the effective field h_i that the i -th spin sees the following:

$$p = 1 / (1 + \exp(-2h_i)) \quad (4)$$

This effective field has two contributions: one from the spin interacting with all other spins with interaction strength J_{ij} , and one from external field $h_{i,ext}$:

$$h_i(t) = \frac{1}{N} \sum_{j=1}^N J_{ij} S_j(t) + h_{i,ext}(t) \quad (5)$$

The interaction strength J_{ij} and $h_{i,ext}$ are both fluctuating random quantities.

If we put simply $J\lambda(t)$ and $h\eta(t)$ for the interactions and fields, the average state of this spin system evolves with time in the mean-field approximation as follows:

$$m(t+1) = \tanh(J\lambda(t)m(t) + h\eta(t)) \quad (6)$$

where the fluctuations $\lambda(t)$ and $\eta(t)$ are the random variables uniformly distributed⁴⁵ in the interval $[-1, 1]$. As we show in the next section, even this simple model captures some of the collective behavior in both populations.

Results

For each user in bot ($N=125$) and human ($N=218$) populations, we obtained BotOrNot score through the BotOrNot API. The total score, a number between 0 and 1, is an estimate of a bot-like behavior of a user. The higher the score, the more likely it is that a user is a social bot. In the left panel of Figure 2, we show the statistics of bot scores for bot and human populations. The scores are indeed significantly different between the two groups with scores for the bots higher than the ones for the humans. In parallel, we also show the statistics of the user activity (the number of published tweets in observed time period) for both populations (right panel in Fig. 2). In this study, we also find significant differences between bots and humans, bots being typically more active, but we also find highly prolific users in the human population.

We have binarized the timelines of activity of the top 10,000 most active users with 30 minutes bin width, so the activity of each user was represented with a spin variable ($S_i(t) = \pm 1$). In Figure 3, we show samples of binarized spike trains for users from bot (left panel of Fig. 3) and human populations (right panel of Fig. 3) as raster plots over the period of two months. There are no obvious patterns visible in these spike trains. We can find highly active users as well as completely silent ones in these samples. However, user activity, represented as spike trains, is not random, but weakly correlated in both, bot and human population.

We have calculated pairwise correlation coefficients (Eq. 1) for the bots and humans, and for the whole 10,000 group of users. Figure 4 shows the distribution of the pairwise correlation coefficients for bots (left panel) and humans (right panel). We also show normal distributions with mean and standard deviation calculated from the data (blue line). Both populations show weakly correlated behavior, although the correlations between human users are more Gaussian-like distributed when compared with bots.

To quantify the difference between correlations in groups of bots and humans, we compared probability distribution of correlation coefficients of 150 randomly sampled users from group of all spins (10,000) with correlation coefficient distributions of bots and humans. We computed Jensen-Shannon divergences (JSD)⁴⁶ between distributions for many samplings of random users from all spins. In Figure 5, we plotted the distributions of JSD between bots and random users, and humans and random users. Both distributions are clearly separated with the mean of human-

FIG. 2. Left panel—comparison of bot and human BotOrNot score statistics; right panel—comparison of bot and human populations statistics of user activity.

FIG. 3. Left panel—raster plot of binarized activity of bots; right panel—raster plot of binarized activity of humans.

FIG. 4. Spin pairwise correlation distributions: population of bots (left panel), human population (right panel). Blue lines are Gaussian distribution with mean and standard deviation calculated from data.

random JSD equal to 0.029 and the mean of bots-random JSD equal to 0.129. The mean of the distribution of JSD between randomly picked groups is equal to 0.005 (not shown in Fig. 5). These results show that the population of humans (or their correlated temporal behavior) is significantly more similar to randomly picked group of users than the population of bots. Or, in other words, it would be hard to detect bots by just looking

at correlations between randomly picked users, but using cross-correlations or partial correlations between population of bots and randomly chosen users could lead to bot detection based on collective properties.

Besides the pairwise correlations, we looked at the collective states of bots and humans quantified with the probability of co-spiking behavior of K spins out of group of N . In each population (bots and humans),

FIG. 5. Distributions of JSD between human population and random sampling from all spins (10k) (left distribution in figure), and bot population and random sampling from all spins (10k). Full lines are Gaussian distributions with mean and standard deviations: 0.029 and 0.011 for human population, 0.129 and 0.018 for bot population. One hundred fifty random samples from 10,000 group were used in all cases, mean and standard deviation of these samplings were 0.005 and 0.004 (not shown in figure). JSD, Jensen-Shannon divergence.

we repeatedly randomly sampled $N=20$ users and computed the probability distribution $P(K)$ of K co-spiking users. Left panel in Figure 6 shows the obtained distributions along with the $P(K)$ for randomly shuffled spike trains. By randomly shuffling spike trains, we destroy all existing correlations in the population and $P(K)$ should be described with the independent spiking model. Indeed, as shown by the dashed line in Figure 6, the $P(K)$ of randomly shuffled spike trains (denoted by pluses) follow the Poisson distribution. However, $P(K)$ from the actual data, for both bots (circles) and humans (squares), is orders of magnitude larger than the independent model prediction, showing the existence of collective states in weakly correlated bot and human populations. $P(K)$ distributions for bots and humans are well described using beta-binomial distribution⁴⁷ (full lines).

The average spin $m(t) = \frac{1}{N} \sum_j S_j$ measures of the activity of a group of users at time t . In a spin-glass model of interacting group of users introduced in the previous section (Eqs. 3–6), the probability that a user will tweet in the next moment depends on the state of all other users in the group (tweeting or not) and on the

influence of external events. This is similar to the economic market models^{45,48} where the price of a commodity emerges as a result of decisions of interacting agents whether to buy the commodity or not. Following the analogy, we computed the logarithmic relative change of the mean activity or the return of the average activity $G(t) = \log(m(t)) - \log(m(t-1))$, which in real markets displays scaling properties.⁴⁹ The right panel of Figure 6 shows the distribution of returns for bots (circles) and humans (squares). The dashed lines are power-laws fits, $P(G) \propto G^{-b}$ to the tails of distributions (with slightly different exponents: $b=4.1$ for bots, and $b=3.8$ for humans) that hint to scaling properties of the activity returns in both populations.

To connect the observed data and the spin-glass model of the bot and human populations, we looked at the time evolution of the variance of average spin, connected to susceptibility in interacting spin systems: $\chi(t) = \text{var}(m(t') < t)$. We obtained the susceptibilities for both populations from the data and compared them to the results of the model computations for the average spin using Eq. 6. As shown in Figure 7, a good agreement with the measured data was obtained with the parameters $J=1.5$ for human population and

FIG. 6. Left panel—distribution of co-spiking activity in population $P(k)$. All random samplings were done with $N=20$, bots (circles), humans (squares), independent model—reshuffled spikes (pluses). Fits: dashed—Poisson distribution, full line—beta-binomial distribution. Right panel—plot of returns of average spin, humans (squares), bots (dots); fits with $J=1.5$ (humans), $J=1.75$ (bots), and $h=0.03$ (both populations); dashed power laws with exponents 4.1 and 3.8.

FIG. 7. Time evolution of susceptibility for human (lower trace) and bot (upper trace) populations. Red lines are results from the Ising spin-glass model with $J = 1.5$ (humans), $J = 1.75$ (bots), and $h = 0.03$ (both populations).

$J = 1.75$ for bot population. Amplitude of external field $h = 0.03$ was kept the same for both populations. The computations of the return of average spin using these same model parameters (shown with red lines in the right panel of Fig. 6) nicely fit to the return of average spin in both populations obtained from the actual data.

To check for predictive properties of tweetstorms as bot signature, we applied a logistic regression binary classifier⁵⁰ to a population of bots and humans, labeling bots with 1 and humans with 0. We divided the whole dataset into training (60%) and test (40%) and computed receiver operating characteristic (ROC) curve (Fig. 8). We repeated the experiment on shuffled data (each time series was randomly shuffled). The area under ROC curve was $AUC = 0.76$ for original data, and $AUC = 0.57$ for shuffled data. These results point to the conclusion that we can successfully classify users into bot-like or human-like class based on their tweeting activity, using supervised learning methods even though we use biased sample (high frequent activity) of accounts.

Discussion

Our results show that we can find users with social bot-like and human-like tweeting behavior within a large group of Twitter users by measuring the lengths and the number of high frequency tweeting sequences—tweetstorms—in their timelines. We used tweet frequency of one tweet every 10 seconds as an individual-level filter, similar to tweeting frequency that was found in other deployments of political bots.⁵¹ The key here is, however, not that our chosen temporal feature can discriminate between bot-like and human-like users (as classified by BotOrNot score) in a selected highly active group of users, but that we find differences in user behavior on group level between bot and human collectives. We found that the group of weakly correlated bots diverges more from the randomly sampled group of users than a group of humans. This suggests that the ac-

tivity of randomly picked users will likely resemble human-like correlated activity, or that bots stay well hidden in overall population. We should also note that our chosen tweetstorm metric does not discriminate between bots and humans based solely on their activity (total number of tweets) since we find highly active users in both groups. Our sampling procedure therefore yields heterogeneous groups of users with respect to overall user activity, so the discovered differences in collective behavior of bots and humans are free from

FIG. 8. ROC curve for predictions with logistic regression binary classifier. We performed supervised learning experiments on original and randomly shuffled datasets. Open dots and full line show the ROC curve for original and pluses and dashed line for shuffled data. Straight dashed line is the random guess line. FPR, false positive rate; TPR, true positive rate; ROC, receiver operating characteristic.

artifacts in selection process. Our approach is limited with some minimum threshold of user activity below which the probability of finding even weak pairwise correlations distinct from random processes is very low. Another challenge is that the user group dynamics is possibly time and context dependent, leading to time-dependent intergroup relationships. One approach to tackle this is to follow the approach developed by Pozzana and Ferrara⁴³ and divide the complete tweets timelines into appropriate activity sessions, and then study collective behavior dynamics across the ensemble of such parts of timelines.

After obtaining populations of bots and humans using an individual-level filter (tweetstorm), we used collective properties like bot-bot, human-human correlations or bot-human cross-correlations to identify bots that can otherwise not be separated well from both populations based on individual-level filter. As an example to demonstrate this approach, we looked at the published list of 2752 Twitter user accounts (IDs and screen names) released by Twitter[†] in response to U.S. Congress investigation of influence of Russian trolls on 2016 U.S. presidential elections. The list contains details on accounts for which Twitter believed were run by a troll farm identified as a Russian company tweeting on election. We cross-referenced the IDs from this list with the list of all users in our Brexit database and found 284 accounts (the order of magnitude consistent with what other researchers have found in their own databases),⁵² but only two (b#1, b#2) accounts between the top 10,000 most active users. To estimate whether b#1 and b#2 are more likely to be bots or humans, we computed their average pairwise correlation between population of bots and population of humans. We found that average correlations of b#1 (b#2) with the population of bots are 0.04 (0.027), while their correlations with population of humans are 0.019 (0.014). In case of these two accounts, we found them to be more likely bots than humans.

We observed that weak pairwise correlations between bots and between humans co-exist with collective, co-spiking, states in both populations. However, could the analysis of correlations between pairs of users lead to any insight into collective behavior in such social groups with complicated interactions between their members? Surprisingly, it does. Collective phenomena in biological and social systems as diverse

as population of neurons,⁵³ insulin-secreting beta cells,⁵⁴ flock of birds,⁵⁵ or U.S. Supreme Court⁵⁶ have been captured by simple, maximum entropy models with minimal structure using pairwise correlations. The key point is that even when correlations are weak, but spread extensively through the system, their effects cannot be treated perturbatively.⁵⁷

We showed that we can describe observations from the data of tweeting behavior of bots and humans by Ising model with random interactions and fields. This suggests that there might be other similarities between collective behaviors of users in social media and those found in other social or biological systems. One exciting possibility to explore is whether a population of interacting social bots can adapt its collective activity so that it is placed near a critical point in the parameter space; such critical behavior was found in many biological^{54,58} and small-scale social systems.⁵⁹ We see hints of criticality in Zipf-like rank distribution of tweetstorms (right panel in Fig. 1) and in scaling relationships of the distribution of return of average spin (right panel in Fig. 6). Zipf's law can emerge naturally without any fine tuning when a system is affected by fluctuating hidden variables.⁶⁰ In case of Twitter or other social media, such unobserved stimuli might be the social or political events around a topic that drive users' activities. Power laws that we found in distribution of return of average spin point to intermittent and bubbling underlying dynamics were similar to the one discovered in financial markets.

Why would being positioned near a critical point be beneficial to a population of social bots? A system at a critical point is highly susceptible to small changes and lacks robustness, information in the system spreads fast. A sophisticated population of social bots poised at critical point would therefore be able to quickly adapt to changes in the uncertain environment and thus become harder to detect and identify. A human population of users, on the other hand, would have to adopt an opposite strategy that would increase robustness to prevent social contagion and infiltration by bots. At the level of individual (human) user, the decisions on whether to follow, retweet, or mention another user of course depend on the contents of tweets and the current state of user's Twitter network (links to other users). What we observe on population level is the collective state of these network interactions, so increased robustness would be reflected in the change of the network topology. However, increased robustness would lower the ability to quickly spread information in the system, so the optimal strategy would be to

[†]Available online at https://democrats-intelligence.house.gov/uploadedfiles/exhibit_b.pdf

adaptively control the distance to criticality in accordance with the changes in the environment.

Conclusions

We have shown that bot-like and human-like behavior of highly active Twitter users can be detected using the peculiarities encoded in the timelines of their tweeting activity, but our key finding is that populations of bots and humans differ in their collective behavior expressed as collective states found in otherwise weakly correlated populations. We quantified these differences by computing distributions of pairwise correlations, co-spiking activity, and average states of each population. We found that some of the scaling properties of tweeting activities of bots and humans binarized into spike trains can be described with a simple Ising spin-glass model. We are intrigued by scaling relationships found in analyzed Twitter data that might hint to criticality and adaptive behavior in fluctuating environment⁶¹ of social bots similar to the one found in biological and small-scale social systems.

Even almost 2 years after the start of the UK-EU Brexit referendum campaigns on Twitter, there are reports on bot influence on referendum outcome,⁶² and ongoing analyses of datasets of Brexit tweets^{63,64} providing new insights into the life of social bots and their networks. In our further efforts, we will work toward the creation of an algorithm that would seek for social bots in large social networks using properties of their collective behavior rather than using individual-level filter. We propose combining our method of quantifying collective bot behavior presented in this study with the exploration of bot network. Instead of using bot followers³⁶ to build bot network, we suggest building a functional network from pairwise correlations between bots. New bot candidates can then be sought by link prediction methods⁶⁵ based on evolving network topology and the bot network expands with newly discovered bots.

We hope our work will stimulate further research of analogies and similarities between online social and biological collective phenomena, leading to new findings about the structure of communication and interaction in social media.

Acknowledgments

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P3-0396). We are grateful to Dr. Onur Varol, Center for Complex Network Research, Northeastern University, for helpful discussions.

Author Disclosure Statement

No competing financial interests exist.

References

1. Dubbin R. The rise of Twitter bots. *The New Yorker*, 14. 2013. Available online at www.newyorker.com/tech/elements/the-rise-of-twitter-bots (last accessed April 2, 2018).
2. Hwang T, Pearce I, Nanis M. Socialbots: Voices from the fronts. *Interactions*. 2012;19:38–45.
3. Ferrara E, Varol O, Davis C, et al. The rise of social bots. *Commun ACM*. 2016;59:96–104.
4. Jenkins H, Ford S, Green J. *Spreadable media: Creating value and meaning in a networked culture*. New York: NYU Press. 2013.
5. Duh A, Meznaric S, Korošak D. *Guerrilla media: Interactive social media*. In: Lugmayr A, Dal Zotto C (Eds.): *Media convergence handbook-vol. 1*. Berlin, Germany: Springer, 2016, pp. 307–324.
6. Mønsted B, Sapiezzyński P, Ferrara E, Lehmann S. Evidence of complex contagion of information in social media: An experiment using Twitter bots. *PLoS One*. 2017;12:e0184148.
7. Howard PN, Kollanyi B. Bots, # StrongerIn, and # Brexit: computational propaganda during the UK-EU Referendum. *arXiv preprint arXiv:1606.06356*, 2016.
8. Bessi A, Ferrara E. Social bots distort the 2016 US Presidential election online discussion. *First Monday*. 2016;21; DOI: 10.5210/fm.v21i1.7090.
9. Hegelich S, Janetzko D. Are social bots on twitter political actors? Empirical evidence from a Ukrainian social botnet. In: *Tenth International AAAI Conference on Web and Social Media*, Cologne, Germany, 2016, pp. 579–582.
10. Woolley SC. Automating power: Social bot interference in global politics. *First Monday*. 2016;21; DOI: 10.5210/fm.v21i4.6161.
11. Adams T. AI-powered social bots. *arXiv preprint arXiv:1706.05143*, 2017.
12. Chu Z, Gianvecchio S, Wang H, Jajodia S. Who is tweeting on Twitter: Human, bot, or cyborg? In: *Proceedings of the 26th Annual Computer Security Applications Conference*, ACM, 2010, pp. 21–30.
13. Grimme C, Preuss M, Adam L, Trautmann H. Social bots: Human-like by means of human control? *Big Data*. 2017;5:279–293.
14. Types of Bots: An Overview. Available online at <http://botnerds.com/types-of-bots/>
15. Tsvetkova M, García-Gavilanes R, Floridi L, Yasseri T. Even good bots fight: The case of Wikipedia. *PLoS One*. 2017;12:e0171774.
16. Franklin S, Graesser A. Is it an agent, or just a program?: A taxonomy for autonomous agents. In: *International Workshop on Agent Theories, Architectures, and Languages*, Berlin, Heidelberg, Germany: Springer, 1996, pp. 21–35.
17. Stieglitz S, Brachten F, Ross B, Jung AK. Do social bots dream of electric sheep? A categorisation of social media bot accounts. *arXiv preprint arXiv:1710.04044*, 2017.
18. Halfaker A, Riedl J. Bots and cyborgs: Wikipedia's immune system. *Computer*. 2012;45:79–82.
19. Oentaryo RJ, Murdopo A, Prasetyo PK, Lim EP. On profiling bots in social media. In: *International Conference on Social Informatics*, Cham, Germany: Springer, 2016, pp. 92–109.
20. Gorwa R, Guilbeault D. Understanding bots for policy and research: Challenges, methods, and solutions. *arXiv preprint arXiv:1801.06863*, 2018.
21. Velázquez E, Yazdani M, Suárez-Serrato P. Socialbots supporting human rights. *arXiv preprint arXiv:1710.11346*, 2017.
22. Howard PN. *Pax Technica: How the Internet of things may set us free or lock us up*. Yale University Press. 2015.
23. Gorodnichenko Y, Pham T, Talavera O. Social media, sentiment and public opinions: Evidence from # Brexit and # USElection. In: *Swansea University, School of Management Working Papers (2018-01)*. Swansea, UK, 2018.
24. Badawy A, Ferrara E, Lerman K. Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. *arXiv preprint arXiv:1802.04291*, 2018.
25. Shao C, Hui PM, Wang L, et al. Anatomy of an online misinformation network. *PLoS One*. 2018;13:e0196087.
26. Woolley S. The political economy of bots: Theory and method in the study of social automation. In: Kiggins R (Ed.): *The Political Economy of Robots*. London, UK: Palgrave Macmillan, 2018, pp. 127–155.

27. Lee K, Eoff BD, Caverlee J. Seven months with the devils: A long-term study of content polluters on Twitter. ICWSM. Barcelona, Spain, 2011, pp. 185–192.
28. Hunt E. Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. The Guardian. 2016. Available online at <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>
29. Shao C, Ciampaglia GL, Varol O, et al. The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 2017.
30. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. 2018;359:1146–1151.
31. Fan R, Talavera O, Tran V. Social media bots and stock markets. Swansea University, School of Management Working Papers (2018–30). 2018.
32. Confessore N, Dance GJX, Harris R, Hansen M. The Follower Factory, The New York Times. 2018. Available online at <https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html>
33. Davis CA, Varol O, Ferrara E, et al. BotOrNot: A system to evaluate social bots. In: Proceedings of the 25th International Conference Companion on World Wide Web, International World Wide Web Conferences Steering Committee, Montreal, Canada, 2016, pp. 273–274.
34. Varol O, Ferrara E, Davis CA, et al. Online human-bot interactions: Detection, estimation, and characterization. arXiv preprint arXiv:1703.03107, 2017.
35. Varol O, Ferrara E, Menczer F, Flammini A. Early detection of promoted campaigns on social media. arXiv preprint arXiv:1703.07518, 2017.
36. 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, New York, ACM, 2017, pp. 467–474.
37. Kudugunta S, Ferrara E. Deep neural networks for bot detection. arXiv preprint arXiv:1802.04289, 2018.
38. Radziwill NM, Benton MC. Bot or not? Deciphering time maps for tweet interarrivals. arXiv preprint arXiv:1605.06555, 2016.
39. Chino DY, Costa AF, Traina AJ, Faloutsos C. VolTime: Unsupervised anomaly detection on users' online activity volume. In: Proceedings of the 2017 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, Philadelphia, USA, 2017, pp. 108–116.
40. Chavoshi N, Hamooni H, Mueen A. Identifying correlated bots in twitter. In: International Conference on Social Informatics, Cham, Germany: Springer, 2016, pp. 14–21.
41. Chavoshi N, Hamooni H, Mueen A. Temporal patterns in bot activities. In: Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2017, pp. 1601–1606.
42. Pan J, Liu Y, Liu X, Hu H. Discriminating bot accounts based solely on temporal features of microblog behavior. Physica A 2016;450:193–204.
43. Pozzana I, Ferrara E. Measuring bot and human behavioral dynamics. arXiv preprint arXiv:1802.04286, 2018.
44. Mézard M, Parisi G, Virasoro M. Spin glass theory and beyond: An introduction to the replica method and its applications, Vol. 9. Teaneck: World Scientific Publishing Co, Inc. 1987.
45. Krawiecki A, Holyst JA, Helbing D. Volatility clustering and scaling for financial time series due to attractor bubbling. Phys Rev Lett. 2002;89:158701.
46. Lin J. Divergence measures based on the Shannon entropy. IEEE Trans Inf Theory. 1991;37:145–151.
47. Nonnenmacher M, Behrens C, Berens P, et al. Signatures of criticality arise from random subsampling in simple population models. PLoS Comp Biol. 2017;13:e1005718.
48. Bornholdt S. Expectation bubbles in a spin model of markets: Intermittency from frustration across scales. Int J Mod Phys C. 2001;12:667–674.
49. Gopikrishnan P, Plerou V, Amaral LAN, et al. Scaling of the distribution of fluctuations of financial market indices. Phys Rev E. 1999;60:5305.
50. We used Python implementation of Logistic Regression classifier. Available online at http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
51. Mustafaraj E, Metaxas PT. From obscurity to prominence in minutes: Political speech and real-time search. In: Proceedings of Web Science Conf., Raleigh, NC, April 26–27, 2010.
52. Llewellyn C, Cram L, Favero A, Hill RL. For whom the bell trolls: Troll behaviour in the Twitter Brexit Debate. arXiv preprint arXiv:1801.08754, 2018.
53. Schneidman E, Berry MJ, Segev R, Bialek W. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature. 2006;440:1007–1012.
54. Korošak D, Slak Rupnik M. Collective sensing of beta-cells generates the metabolic code. Front Physiol. 2018;9:31.
55. Bialek W, Cavagna A, Giardina I, et al. Statistical mechanics for natural flocks of birds. Proc Natl Acad Sci U S A. 2012;109:4786–4791.
56. Lee ED, Broedersz CP, Bialek W. Statistical mechanics of the US Supreme Court. J Stat Phys. 2015;160:275–301.
57. Azhar F, Bialek W. When are correlations strong?. arXiv preprint arXiv:1012.5987, 2010.
58. Mora T, Bialek W. Are biological systems poised at criticality? J Stat Phys. 2011;144:268–302.
59. Daniels BC, Krakauer DC, Flack JC. Control of finite critical behaviour in a small-scale social system. Nat Commun. 2017;8.
60. Schwab DJ, Nemenman I, Mehta P. Zipf's law and criticality in multivariate data without fine-tuning. Phys Rev Lett. 2014;113:068102.
61. Paarporn K, Eksin C, Weitz JS. Information sharing for a coordination game in fluctuating environments. bioRxiv. 2018:268268.
62. Available online at <https://www.theguardian.com/world/2017/nov/14/how-400-russia-run-fake-accounts-posted-bogus-brexit-tweets>
63. Bastos MT, Mercea D. The brexit botnet and user-generated hyperpartisan news. Soc Sci Comput Rev. 2017; DOI: 10.1177/0894439317734157.
64. Llewellyn C, Cram L. Distinguishing the wood from the trees: Contrasting collection methods to understand bias in a longitudinal brexit Twitter dataset. In: ICWSM. 2017, pp. 596–599.
65. Martinić-Ipsić S, Močibob E, Perc M. Link prediction on Twitter. PLoS One. 2017;12:e0181079.

Cite this article as: Duh A, Slak Rupnik M, Korošak D (2018) Collective behavior of social bots is encoded in their temporal Twitter activity. *Big Data* 6:2, 113–123, DOI: 10.1089/big.2017.0041.

Abbreviations Used

JSD = Jensen-Shannon divergences
ROC curve = receiver operating characteristic curve