Dormant Bots in Social Media: Twitter and the 2018 U.S. Senate Election

Richard Takacs
Applied Physics Laboratory
Johns Hopkins University
Laurel, MD, United States
Richard.takacs@jhuapl.edu

Ian McCulloh
Accenture
Washington, DC, United States
Ian.mcculloh@accenturefederal.co

***Abstract*— Bots are often identified on social media due to their behavior. How easily are they identified, however, when they are dormant and exhibit no measurable behavior at all, except for their silence? We identified "dormant bot networks" positioned to influence social media discourse surrounding the 2018 U.S. senate election. A dormant bot is a social media persona that does not post content yet has large follower and friend relationships with other users. These relationships may be used to manipulate online narratives and elevate or suppress certain discussions in the social media feed of users. Using a simple structure-based approach, we identify a large number of dormant bots created in 2017 that begin following the social media accounts of numerous US government politicians running for re-election in 2018. Findings from this research were used by the U.S. Government to suspend dormant bots prior to the elections to prevent any malign influence campaign. Application of this approach by social media providers may provide a novel method to reduce the risk of content manipulation for online platforms.**

Keywords—Twitter, bots, social media, elections, political

I. Introduction

Social media ecosystems offer incentives, both economic and political, for systems that create and manage accounts exhibiting human-like behavior. The introduction of social bots has accelerated the development of these systems. Broadly speaking, bot software is designed to run automated scripts over networks to perform specific, usually repetitive functions [1], [2], [3]. Their introduction to social media ecosystems, through platform-specific application programming interfaces (APIs) allows much of the same functionality as a human user. For example, on Twitter a bot has the ability to post content, respond to others, retweet other user's content, create relationships with other users, and direct message other accounts. When organized into simple networks, all of these activities can be accomplished in large scale. As a versatile yet inexpensive method to alter existing networks and content propagation patterns on social media, bot capabilities have been utilized by governments [4], [5], political campaigns [2], [6], [7], private corporations [8] and individuals [9] for a variety of economic and social applications. "Computational Propaganda", a term coined by Oxford University's Internet Institute, summarizes automated systems that spread misinformation and utilize data-driven methods to shape public opinion [10]. These accounts are used to artificially inflate the perceived importance of content; overwhelm narratives counter to the ideological, political or commercial interests of network owners [2], [11]; and conduct search engine optimization of social media content – a process of ensuring content managed by the bot network appears higher in search results than opposition material [12]. While examples of this online behavior are numerous, recent studies find that within a sample of 1.3 million Russian Twitter accounts discussing political and military topics between 2014 and 2015, as much as 50% of activity was managed by highly automated accounts [13].

Despite attention given to the issue, detection of bots in social media remains problematic. Social media platforms themselves limit transparency through their APIs [14]. Techniques generally used by researchers to sniff out bots on Twitter rely on behavior anomalies [15], [16], [17] such as post frequency, topic-sentiment analysis and network analytics. These methods lag behind the increasingly sophisticated bots they are supposed to detect. While these methods have been useful in illuminating some networks, they remain limited by the requirement that bot accounts take an active role in posting on social media platforms and are ineffective in highlighting bot networks that have been created but have not, at the time of study, posted content; a phenomenon we term "dormant bots".

In this paper, we identify dormant bots positioned in the Twitter networks of United States Senators eligible for re-election in 2018 and highlight the potential scale of the dormant bot issue within the U.S. political social media environment.

ASONAM '19, August 27–30, 2019, Vancouver, BC, Canada

ACM ISBN 978-1-4503-6868-1/19/08$15.00
https://doi.org/10.1145/3341161.3343852

II. Background

Twitter is a social networking and online news service that allows users to post and interact with 280-character messages referred to as "tweets". As of July 2017, Twitter averaged 328 million monthly active users, making it the eighth most popular social network in the world [17]. As more and more users join the platform, its perceived influence in politics and social debate rise with Twitter's popularity.

Politicians increasingly use Twitter to communicate with constituents and increase their popularity among voters during election campaigns [2], [6], [7]. As of the writing of this paper, the current US president, Donald Trump was the twentieth most popular Twitter persona with almost 47 million followers and the former US president, Barack Obama was third with 99 million followers. While other US politicians hold significantly fewer followers on Twitter, the social media platform is generally considered an important tool for reaching the voting population and main-stream media and politicians increasingly use the platform [7].

The United States experienced an increase in support for populist candidates during its 2016 election [18]. Much of the populist movement was credited to the increased popularity of social media platforms such as Twitter and Facebook. No doubt, Twitter played an important role in the 2018 US election. Politicians may use social media platforms to increase voting participation, communicate their campaign platform, and influence voters to support them. As social media campaign tactics become more sophisticated, political campaigns are likely to evolve with them.

This paper focuses on the Twitter accounts of incumbent U.S. senators running for reelection in 2018. Following the 2016 election, 46 senators were in the democratic political party, 52 were in the republican party, and 2 were independent. There were 33 of 100 U.S. Senate seats open for election in November, 2018. Of these seats, 25 were held by democrats and 8 were held by republicans. Four of the serving democrats and four republicans stated they will not run for re-election, leaving 25 incumbent senators that campaigned in 2018.

III. Method

Research consisted of three phases: data collection, persona analysis, and follower analysis. The first phase, data collection, involved extracting Twitter data on all incumbent U.S. Senators running for re-election in November 2018. The second phase focused on analysis of follower accounts' creation date, posting activity, and follower and friends networks. The third phase explored differences in attributes between bot and human accounts. All analysis was conducted in the open-source R coding language [21] using publicly available packages.

Twitter's REST API was utilized to collect the follower networks of the 21 Democratic and 4 Republican incumbent senators seeking re-election. Combined, all Senatorial accounts had 5,976,176 followers at the time of collection in January 2018. The data set provides values including screen names, numbers of followers, friends and posts, account creation dates and an optional user-defined location. Again, data collection was not funded by research sponsors.

In phase two, accounts were segmented by the month of follower account creation to analyze the growth in followership for the 25 studied personas. A general pattern in follower account creation was observed, with spikes in new accounts occurring in January 2017 and July 2017. Account creation distributions for followers occurring before July 2017 and those after July 2017 were therefore analyzed for statistically significant variations using the Kruskal-Wallis tests [22] to examine internal variance and determine the independence of the samples. Next, either the Wilcoxon rank sum [23] or Mann-Whitney test [24] were used, depending on the independence of samples, to determine if followers created before and after July 2017 represented statistically different populations.

For phase three, two methods were applied to determine behavioral differences between follower accounts created before and after July 2017. First, Kendall Rank Correlation was used to examine the relationships between friends, followers, and account posts. Next, samples of follower accounts were examined for bot behaviors using Indiana University Network Science Institute's Botometer API [5].

All analysis was conducted with median calculations and nonparametric tests to account for heavily skewed distributions within variables and the research intent to maintain primary data by not utilizing normalization techniques.

IV. Findings

Three large increases in follower account creation are observed in the data. Figure 1 displays a histogram of the Twitter account creation dates by month. This figure indicates three spikes in new account creation for followers of incumbent senators running for reelection. The first spike occurs in 2008 and can be attributed to the growing popularity of Twitter at that time. The second spike occurs in January 2017 around the time of the US presidential inauguration. The third spike occurs around the end of July and beginning of August, 2017. According to our data, 2,803,999, or roughly 47% of all Twitter accounts following U.S Senators standing for election in 2018 were created between July and December 2017. Comparing account creation between political parties highlights growth at different scales. 2,757,262 of these new accounts are found following Democratic candidate personas with the remaining 46,737 following Republican accounts. On average, democrat incumbents had 11 times as many newly created Twitter accounts following them between August and December of 2017.

The difference in behavior of the Twitter accounts was analyzed for those accounts who began following the U.S. politicians before and after July 2017, when the last spike in account creation began. Behavior was analyzed by measuring the correlation between the number of friends, followers, and posts for the follower accounts. Kendall's correlation was utilized to provide a distribution free test of dependence between all variables. The Wilcox signed-rank test was used for each of the three comparisons (follower- friend, follower-post, friend-post) to test the null hypothesis that the difference

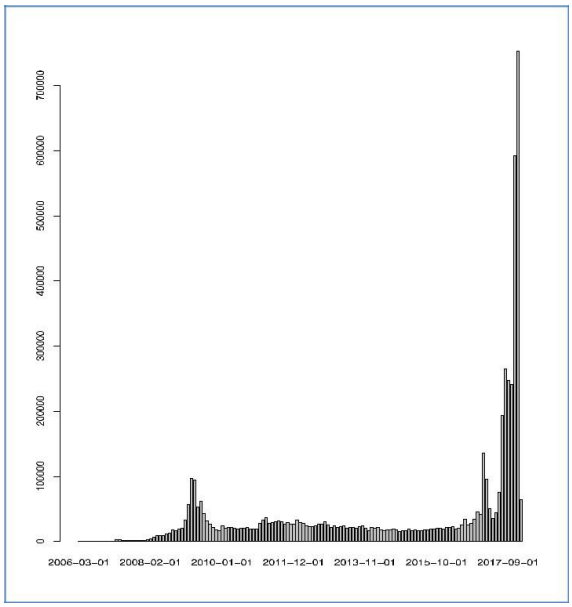

Fig. 1. Histogram of account creation dates for Twitter followers of incumbent US senators campaigning in 2018.

in correlations was zero against the alternative hypothesis that the difference is not zero. The p-value for all three comparisons was much less than 0.0001, thus we reject the null hypothesis at the greater than 99.9% confidence level. We conclude that the relationship between the follower-friend-post patterns are statistically different for accounts created before and after July 2017.

For accounts created prior to July 2017, there is a high correlation between the number of followers and the number of friends, 0.88; the number of followers and the number of posts, 0.93; and the number of friends and the number of posts, 0.88. These correlations suggest a strong relationship between the posting activities of a Twitter user and their networks of friends and followers.

For accounts created after July 2017, however, correlation coefficients fall significantly. The correlations between the number of followers and number of friends is 0.60; between the number of followers and number of posts is 0.64; and between the number of friends and number of posts is 0.30.

Further analysis of post-July 2017 follower behavior indicates a cause for these differences. We find that 1,154,894 accounts created since July 2017 (41%) that follow incumbent Senators have never posted to Twitter, making them a potential "dormant bot". Here too, we observe differences between political parties with 1,137,910 accounts following democratic senators and 16,984 following republicans. There is therefore an approximate average of 54 thousand potential dormant bots following each democrat incumbent compared with an average four thousand following their republican peer. These potential dormant bots that have never posted a tweet have an average of 1.7 followers and 99.0 friends. Without further explanation, it appears odd that a social media persona would hold such high degree in a social network when they don't communicate.

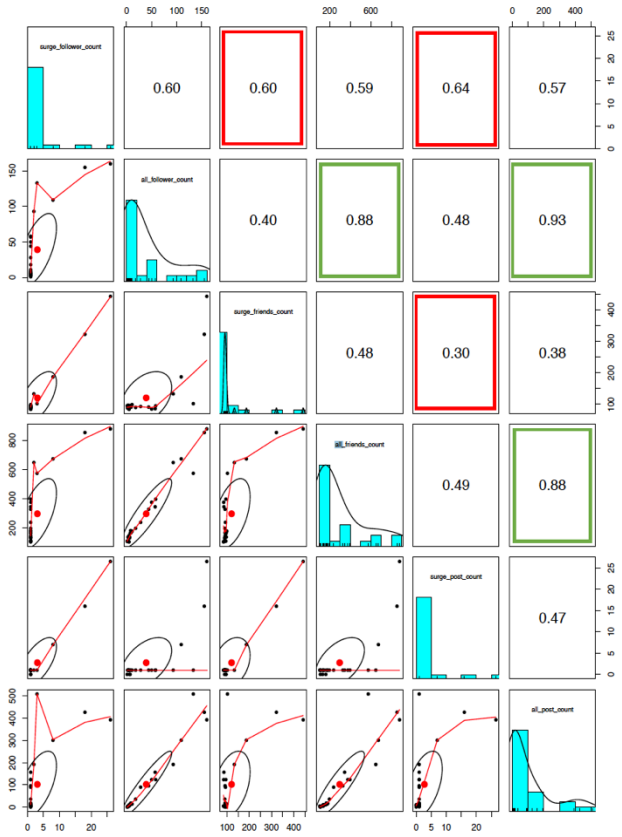

Fig. 2. Kendall Rank Correlation Table: posts, friends, and followers for all accounts and accounts created in July 2017.

One limitation of machine learning approaches to detect automated bots is that they require posting content to evaluate behavior. If a bot is dormant, there is no behavior to consider. The remaining 59% of newly created accounts that did post content were randomly sampled and tested for "bot behavior" using the University of Indiana Botometer. Results indicate that 71.9% of the accounts that tweet are likely bots at 65% confidence. While there are several limitations to the Botometer methodology, this indicates an unusually high presence of bots in the online political discourse. If we assume accounts that never post may be bots and extrapolate the number of Botometer detected bots, then as many as 83.5% of newly created Twitter accounts following senate incumbents are bots. Furthermore, democrat incumbents are far more likely to have bot followers than their republican counterparts.

V. Conclusion

Early findings of this research were shared with the U.S. Federal Bureau of Investigations in April 2018, who used this research to obtain a court order requiring Twitter to suspend "dormant bots". This immediately preceded Twitter suspension of 70 million accounts [25] and a subsequent 23% drop in their stock price. Twitter responded by making changes to their API access the following month [26]. This paper not only contributes to the growing body of knowledge regarding social media, online influence, and

automated content manipulation through social bots; it also highlights the importance of research for protecting the legitimacy of democratic elections in an increasingly connected world.

This paper presents two important contributions. First, we find an unusual pattern in account creation following U.S. politicians running for re-election in 2018. Accounts that do not exhibit any social media behavior and remain dormant may indicate a new type of bot tactic or behavior. While these accounts do not directly shape the online discussion, their presence as followers may affect search engine rankings making followed candidates easier to find in searches or amplify their content. These potential dormant bots are also pre-positioned to increase attention to followed accounts by liking, retweeting, or mentioning their content in the future. These dormant bots are also well-positioned within online friend networks, which may reduce detection from structural bot identification methods and position these accounts for greater influence in affecting the online discourse.

The second contribution brings attention to the presence of a high number of potential bots in the 2018 US congressional election race. While the existence of bots in Twitter is not surprising, the volume of potential bots is truly alarming. With approximately half of Twitter accounts following incumbent political candidates being created in the last six months and as many as 83.5% of them being bots, there are many challenges that must be addressed. The least of these challenges is the veracity of Twitter as a lens into American voting dynamics. Some voters who rely on social media may be unduly influenced by the large bot behavior. Most concerning may be the potential attacks against the legitimacy of U.S. democratic elections following public realization of bot activity.

This research is not without limitations. Our work focused on U.S. Senate incumbents running for election in 2018. This work could be expanded to investigate the U.S. House of Representatives and politicians who are not running for re-election. It would also be interesting to investigate similar patterns among Twitter accounts representing government agencies, corporations, and celebrities. This would help determine whether this behavior is localized to political campaigns or more broadly to social media. Future work could test the veracity of the dormant bot assumption by sending tweets to a random sample of suspected accounts and look for a response. The mere presence of an account that exhibits no activity is suspect however.

Given the growing concern over bots and their potential impact in the online discourse, we should be concerned. Social media providers should consider implementing simple algorithms to detect potential dormant bots and investigate their validity. They should consider policies to either deactivate accounts after some period of time, require users to verify account legitimacy, or some policy to reduce the risk of automated manipulation of the online discourse.

REFERENCES

[1] Akimoto, A. 2011. Japan the Twitter nation. *The Japan Times*, 18.

[2] Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. 2011. The socialbot network: when bots socialize for fame and money. In *Proceedings of the 27th annual computer security applications conference.* 93-102. ACM.

[3] Dunham, K. 2008. *Malicious Bots*. London: Taylor and Francis

[4] BBC News. 2012. Russian Twitter political protests swamped by spam. Retrieved from: http://bbc.com/news/technology-16108876 January 17, 2018

[5] Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A. 2016. The rise of social bots. *Communications of the ACM*, 59(7): 96-104.

[6] Metaxas, P.T. and Mustafaraj, E. 2012. Social media and the elections. *Science*, 338(6106): 472-473.

[7] Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A. and Menczer, F. 2011. Truthy: mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on World wide web*, 249-252. ACM.

[8] [Kelly, J., Barash, V., Alexanyan, K., Etling, B., Faris, R., Gasser, U., & Palfrey, J. 2012. *Mapping Russian Twitter*. Cambridge MA: Harvard Berkman Center

[9] Lotan, G. 2014. Mining Twitter gold, at five bucks a pop. *Los Angeles Times*. Retreived from: http://www.latimes.com/opinion/op-ed/la-oe-0601-lotan- buying-followers-20140601-story.html

[10] Woolley, S.C. and Howard, P.N 2017. *Computational Propaganda Worldwide.* Oxford: Oxford Internet Institute.

[11] Subrahmanian, V.S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A. and Menczer, F. 2016. The DARPA Twitter bot challenge. *Computer*, 49(6): 38-46. IEEE.

[12] Killoran, J.B. 2013. How to use search engine optimization techniques to increase website visibility. *IEEE transactions on professional communication*, 56(1): 50-66.

[13] Stukal, D., Sanovich, S., Bonneau, R. and Tucker, J.A. 2017. Detecting Bots on Russian Political Twitter. *Big data*, 5(4): 310- 324.

[14] Driscoll, K. and Walker, S. 2014. Big data, big questions| working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication,* 8: 1745–1764.

[15] Grier, C., Thomas, K., Paxson, V. and Zhang, M. 2010. @ spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security,* 27-37. ACM.

[16] Pitsillidis, A., Levchenko, K., Kreibich, C., Kanich, C., Voelker, G.M., Paxson, V., Weaver, N. and Savage, S. 2010. Botnet Judo: Fighting Spam with Itself. In *Proceedings Network and Distributed System Security Symposium*

[17] Stringhini, G., Kruegel, C. and Vigna, G. 2010. Detecting spammers on social networks. In *Proceedings of the 26th annual computer security applications conference*, 1-9. ACM.

[18] Ratkiewicz, J., Conover, M., Meiss, M.R., Gonçalves, B., Flammini, A. and Menczer, F. 2011. Detecting and Tracking Political Abuse in Social Media. *ICWSM*, 11: 297-304.

[19] Dunn, J. 2017. Facebook totally dominates the list of most popular social media apps. *Business Insider*.

[20] Groshek, J. and KocMichalska, K. 2017. Helping populism win? Social media use, filter bubbles, and support for populist presidential

candidates in the 2016 US election campaign. *Information, Communication & Society*, 1-19.

[21] R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

[22] Kruskal, W.H. and Wallis, W.A. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260): 583-621.

[23] Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6): 80-83.

[24] Mann, H.B. and Whitney, D.R. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of mathematical statistics*, 50-60.

[25] Timberg, C. and Dwoskin, E., 2018. Twitter is sweeping out fake accounts like never before, putting user growth at risk. *The Washington Post*.

[26] Anon. New developer requirements to protect our platform. Retrieved April 7, 2019 from https://blog.twitter.com/developer/en_us/topics/tools/2018/new-developer-requirements-to-protect-our-platform.htm