The devil is in the details: Studying the Influence and Content Diffusion Dynamics of Social Bots During 2014 Italian Mayoral Elections

Pujan Paudel[1] , Andrew H. Sung[2]

School of Computing Sciences and Computer Engineering,
The University of Southern Mississippi, Hattiesburg MS 39406, USA
pujan.paudel@usm.edu,andrew.sung@usm.edu

Abstract. Twitter bots have been exercised by political campaigns and state-level agencies in different settings, with different objectives. From spreading fake news and rumors to interacting with human users on political discussions, social bots have made their mark in election processes and social movements. But the question of how effective and successful the social engineering strategies of these social bots could be, in a similar context of human discussions during the same election is an important question. In this work, we study an identified network of Twitter bots in the Mayoral Election of Italy in 2014 and compare them with human users activity. Our analysis of the echo-chamber of political conversation, the emotional cost of information diffusion, and the mechanism of information spreading reveal to us the social bots being surprisingly able to reach audiences of wider political ideology by tweeting with lesser emotional cost in their tweets employing efficient information diffusion behavior. We then discuss the possible implications of these effective social engineering strategies of bots from multiple viewpoints and how twitter bots could be possibly effective in jeopardizing the political process.

Keywords: social bots · twitter · echo-chamber · information diffusion

1 Introduction

Social bots in Twitter which are primarily deployed with political objectives have evolved their strategies and modes of attack with time, and they can no more be marked off as merely content spammers or conversation polluters. The adversary bot-masters have adapted their automated agents to be highly capable of suppressing the flow of democratic voices through collective demobilization of social botnets and change perspectives of public on social actors, their discussions and the entire political regime through the volumetric mirage of automated influence. Most of the previous studies have been primarily focused on studying the channels of communication utilized by the automated accounts to influence the social movements. The question of the social engineering strategies they adopt

and the effectiveness in doing so still remains. We hope to answer this question in our work by comparing multiple behaviors of an identified political botnet in the 2014 Italian Mayoral election with the human accounts participating in the same election. Due to the space limitations, we refer our readers to [3] for a more detailed review of literature on political bots and their usage across different political contexts.

2 Data Collection

For our study of political social bots, we took the dataset of bots from the work of [10]. The dataset consists of a novel group of social bots that were deployed by one of the runner-up parties of the 2014 Mayoral election of Rome, employed through a social media marketing firm. In our recent work [9], we discovered how those bots have evolved over the traditional political bots : network wise, communication-wise and behavior-wise. We were further motivated by the complex and human-like resemblance shown by the political social bots to compare their activities with human users taking part in the same election. We identified the politically relevant keywords the bots were engaged in by initially creating a seed set of hashtags used in the re-tweets of the political accounts whom the bots were promoting. We then expanded our keywords by identifying the hashtags that co-occurred with the seed hashtags more than 10 times. The top keywords which we used to identify the politically relevant conversation, along with the parent topic of the keyword is listed in Table 1. It can be observed that we cover a wide variety of hashtags related to the primary political party leaders under concern, the opposition party, and external affairs to encompass the different dimensions of election conversations taking part on the social network.

To identify human users to compare the activity of bots with, we needed to identify Twitter users who took part in the political discussion during the surfacing event of 2014 Mayoral elections. Based on the work of [11], the required conversations on twitter can be collected using 3 different strategies, i) using hashtags ii) using random stream iii) using an official set of dictionary users known to be talking on the topic. We applied the methods i) and ii) to collect users and tweets for constructing the human user dataset as we were not able to use the random stream of Twitter to get archival data back from 2014. For the "hashtag" based collection approach, we collect all the tweets containing either of the keywords in Table 1 during the identical timeline of the tweets filtered from the bots during 2013-09-15 to 2014-11-30. We used the GetOldTweets-Python tool to collect the archival data. We then removed the tweets and users who occurred less than 10 times in our collected sets to remove possible hashtag hijackers and content spammers. We refer to the data collected through this approach as *users-samples-1*. For the "official set" data collection strategy, we initially identified the top 5 political candidates which were retweeted the most by the twitter bots. We then expanded through all the retweeters of the 5 political accounts, during the identical period and eliminated verified accounts and also the accounts with less than 10 retweets originating from the political accounts.

This approach gives us a relevant list of human users who were tweeting on the same topic of the political discussion as the political bots during the election period. We also removed the accounts that followed either of the botnet accounts. We refer to the data collected through this approach as *users-samples-2*. We denote the combined datasets (*users-samples-1* and *users-samples-2*) with *users-samples-combined*.

Completely reconstructing the conversation landscape of an event that happened in the past would be a very difficult task, but we believe the two approaches we used to provide us a close approximation of the discussions. The tweets collected through the official strategy somewhat mitigate the lack of tweets that could have been collected from a random stream, as it has been studied that the "streaming strategy" is most similar to the "official set strategy" in terms of relevance and frequency. Our approach also minimizes the drawback of "official set strategy" to oversample broadcast accounts reported on [11] as we remove the non-seed verified users from our data collection pipeline. As the final process of our pipeline of finding the suitable users to represent human communication in our studies, we used the tool Botometer [17] and computed the metric CAP(Complete Automation Probability) of the users present in the *user-samples-combined* dataset. We removed all the users above the threshold CAP score of 0.3. We used a relatively conservative threshold score for the CAP to eliminate any possible presence of automated accounts in our *user-samples-combined* dataset.

Table 1: Seed hashtags used for data collection strategy

Politicians	**Party/ Slogans**	**Opposition Party**	**External Affairs**
#Renzi	#PD	#Brunetta	#Palestina
#cuperlo	#BelloeDemocratico	#Grillo	#Hamas
#gotti	#congressopd	#Lupi	#Gaza
#bersani	#M5s		#IsraelUnderAttack
#letta			#TelAviv

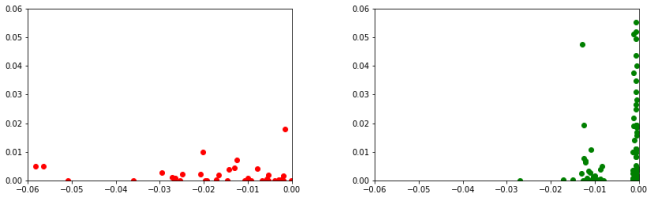
3 Methods

3.1 The extent of Echo Chambers

The presence of echo-chambers in online discussion is an inevitable scenario produced as a result of selective exposures of the platforms and ideological segregation of the users on it [14]. Prior works [14] have highlighted the strong presence of echo-chambers in retweet based interaction in tweets discussing political contexts. We compared the quantitative presence of echo-chambers induced by the

bots and human users by studying how politically polarized the users are who spread or retweet the content of them. In the case of bots, we collected the human user accounts who tweeted more than 10 tweets originally authored by the bots of the botnets, to make sure the human users we included in the study are not random content distributors, but occasionally share the posts of the Twitter bots. In the case of human users, we used the *users-samples-2* dataset we collected, to make sure the human users we studied are actively engaged in a discussion about the political activity occurring at the same time the botnet was deployed. We selected the 500 most frequently used URLs from each of the datasets, and filtered 200 of them being news outlets or media portals. We initially used the lists of partisan media outlets compiled by third-party organization Media Bias/ Fact Check. Since most of the sources of media outlets were in Italian language and the outlets not being available on the partisan list, we manually annotated the news outlets by verifying their political alignment cross-checking with Wikipedia. We labeled the polarity score of news sources falling under the left and the left-center with -1, right and right-center with 1 and the partisan media outlets with a score of 0. An overview of the media outlets and their polarity scores are listed in Table 2. We then assign two different polarity scores to the users in our study, i) Production Polarity: the mean polarity score of the tweets produced by a user, which contains links to the media outlet of known political polarization. ii) Consumption Polarity: the mean polarity score of the tweets produced by the users who share the contents of the user.

In Table 3, we list multiple statistics regarding the production and consumption polarities for the bots and human users. The Pearson correlation between Production polarity and Consumption polarity for the bots understudy is -0.0444 compared to the human user's 0.258. The negative correlation observed in bot's sharing activity and its spreaders are very interesting compared to a relatively stronger Pearson correlation in the case of humans, highlighting a strong case of echo-chambers present in the community of human users. On average, the bots produce content with a lower polarity score than the humans, and the mean polarity score of its content consumers is also lower than the human users. The Standard Deviation (S.D) of the consumption polarities of the bots is higher for the bots than humans, suggesting the bots were able to reach audiences with relatively broader political spectrum. Secondly, we shift the analysis asking the question of how the increase in production polarity affects the variance/spread of polarity scores of the consumers reached by the producers. In Figure 1, we plot the spreader influence variance against the production polarity. It can be seen that in the case of bots, with the polarity scores being very close to zero itself, the spread of polarity scores is very high as compared to the humans. The production polarity of bots are in very lower ranges compared to humans, and they still attract higher ranges of variance in polarity scores of spreaders than humans. In the case of the human users, as the produced content becomes more polar, the variance of the spreaders reached increases significantly, again reflecting a very strong presence of echo-chambers.



(a) (b)

Fig. 1: Production Polarity against Spread of Influence in a) Humans b)Bots

Table 2: Top Media Outlets and polarity Score

Left (-1)	**Right(+1)**
ilmanifesto	iltempo
messianicradio	lefigaro
carmillaonline	breitbart
partitodemocratico	ecodibergamo
rifarelitalia	sicilia5stelle
daysofpalestine	conservativescores
lapresse	algemeiner
primariepd2013	ohalright
volkskrant	ilfoglio

Table 3: Measures of Polarity

Measure	**Bots**	**Humans**
Mean Production Polarity	-0.0021	-0.0387
Mean Consumption Polarity	-0.00036	-0.0311
Std. Production Polarity	0.0055	0.07432
Std. Consumption Polarity	0.18765	0.12334

3.2 Emotional Cost of Information Diffusion

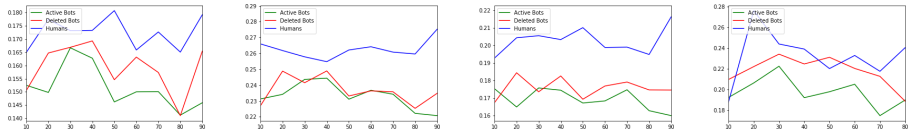
We studied how the degree of influence gained by a tweet varies with the emotional content present on it. We used the tweets by the bots and the users from our *users-samples-combined* dataset with the number of retweets obtained between 5 and 90. For getting the numerical scores of the multiple dimensions of emotional attributes conveyed by the tweets, we used the Perspective API. We translated the tweets in our datasets using Google Translate API as the tweets were in the Italian Language. We used 5 different dimensions of emotional analysis offered by Perspective API, out of the 16 total. The dimensions of emotion we study, alongside the description of the model annotating the score, are :

i) Severe Toxicity: rude, disrespectful or unreasonable comments. ii) Threat: intention to inflict pain, injury, or violence against an individual or group. iii) Profanity: swear words, curse words, or other obscene or profane language. iv) Insult: Insulting, Inflammatory, or negative comments towards a person or a group of people. v) Identity Attack: negative or hateful comments targeting someone because of their identity. We analyzed the tweets of the users in bots dataset with further detail by dividing the tweets of the bots into the active tweets, authored by the users who are still present on the Twitter platform remaining active, whereas deleted tweets, who have been taken down by the platform, alongside the account tweeting it.

We plot each of the emotional dimensions of the tweets ranged from (0-1) in Y-axis against the influence (retweet) gained by the tweet in X-axis. We study how the emotional attribute score changes with the increase in the amount of influence gained by the tweets. In Figure 2, for all the emotional attributes of the tweets we are studying, we can observe two different types of behaviors: firstly, the negative emotion scores of the diffused tweets by humans are always higher than those by the bots (both deleted as well as active), with increasing amount of gathered influence. The bots were able to gain the same amount of influence as humans by tweeting content with a significantly lesser amount of toxicity, identity attack, profanity. We also observed that with the increasing amount of influence achieved, the tweets of the deleted bots were slightly in the upper ranges of the emotional score, which can be interpreted as one of the possible reasons those bots were deleted from the platform, while the one that is active still prevalent. Most interestingly, we can observe that in the case of the human users, the increasing influence comes with the cost of an increase of negative emotional scores, for all of the negative emotions under study. Whereas, the bots had only a little or no rise of negative emotion with respect to the increase of influence obtained. This symbolizes that the emotional cost of diffusion was higher for human users than the bots. The phenomenon of positivity bias in information spread has been validated in prior works thus explaining tweets with positive sentiments spreading more in volume than the negative ones. Our findings suggest that the phenomenon was stronger for social bots than human users. Our findings disagree with the one reported by [16], who reported that emotionally neutral tweets failed to garner much attention in terms of retweet. Building upon our findings of the bots tweeting with lower emotionally harmful sentiments than the human users, we also studied how the speed of diffusion correlates with the negative emotional attributes of a tweet. We re-use the definition of the speed of diffusion used in prior works [13], as the duration between the time the tweet was posted and it's first retweet observed. We observed that the bots were able to gain much lesser response time to the first re-tweets on average (15839.53 seconds), implying greater speed than the humans (81371.83 seconds), by tweeting with a lesser magnitude of all emotional dimensions on their text at the same time. Our findings are contrasting with the work of [13], where the author reported tweets with higher negative valence having a higher speed in spreading than the positive and neutral ones.

3.3 Mechanism of Information Spreading

First, we studied whether the broadcast model or the viral model dominated the message spreading of the bots and humans in the social space. The viral diffusion model represents interpersonal communication through a chain of the individual-to-individual spreading process. Whereas, the broadcast diffusion model closely relates to diffusion patterns of mass media or marketing efforts where there are limited information sources disseminating information to a large number of individuals.



(a) (b) (c) (d)

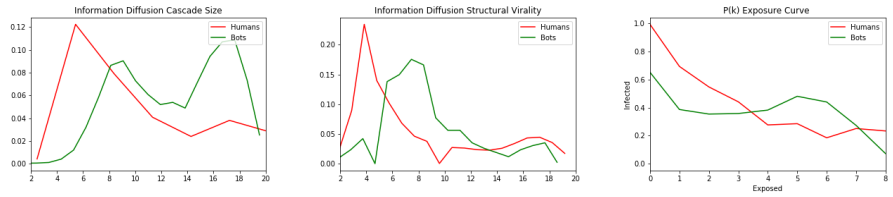
Fig. 2: Emotional Cost of Diffusion a)Severe Toxicity, b)Threat c)Identity Attack & d) Obscene

We created retweet cascades for all the tweets in our analysis, by using the metadata attached with the tweets, the followers of the users in our study, and the retweet information. The retweet cascade we constructed represents the path of diffusion of the tweets from the seed node to the message spreaders. To solve the inherent problem of re-tracing the true diffusion path of the retweet tree of a diffusion process, we applied an approach previously used before in [15]. After re-constructing the diffusion path of the retweet cascades we used the measure of structural virality to quantify the mode of message diffusion, which is the average distance between all pairs of retweeters in the tree. There are multiple measures to calculate the structural virality of a diffusion tree as discussed briefly in the work of [15], and we use their proposed approach of using Weiner's index to compute the structural virality of the diffusion cascade.

We also investigate the spread of the tweets from the perspective of information adoption, asking the question, How do successive exposures to a tweet through mutual retweets to a user affect the probability that the user will retweet the information as well ? Different from the work of [12], we build a network on the users from the structure of interaction via retweet signals, instead of mention signals. We use the retweet metadata and friendship graph available to get the exact timestamp and diffusion paths of the adoption of information through retweets. Borrowing the terminology from [12], we call a user is k-exposed to a seed user h, if he has not retweeted a message from h, but k of his friends have already broadcasted a message of h. We then estimate the probability that user u, which has been k-exposed to a user h, will retweet a message from h before becoming (k+1) exposed.

After reconstructing the diffusion tree using the methodology discussed, we measure the structural virality of information cascades. To get an initial idea about the popularity of the seed messages for our analysis, we plot the distribution of cascade sizes for both bots and humans in Figure 3. We can observe that the messages diffused by bots had larger cascade sizes, denoting the greater popularity of messages authored by bots than humans. We then plotted the structural virality of the diffusion cascades in Figure 3. A higher value of structural virality would represent a viral mode of diffusion being more dominant in the communication patterns. We can observe that the information cascades of the bots have higher values of structural virality than by humans. This signifies that the spread of information of tweets authored by humans mainly followed

a broadcasting pattern, indicating a star network of retweets from the original tweets but without many further retweets and selective sharing with declining diversity over time. Whereas, the higher value of structural virality for the diffusion cascades of the bots in our study signifies a complex process of viral spreading diffusion occurring in the case of bots. This signifies the more effective mode of communication pattern occurring in the diffusion process of twitter bots, where messages went "viral" through multiple chains of an individual-to-individual diffusion process, also indicating a higher probability of cross-ideological sharing across heterogeneous communities, which might also help explain the results of section I.



(a) (b) (c)

Fig. 3: Information Diffusion a) Cascade Size and b) Structural Virality c) P(k) Exposure Curve

We calculated the exposure curves P(k) for each of the information diffusion chains of the bots and humans and averaged those curves in Figure 3. The stickiness of the information spreading, signified by the maximum value of P(k) is considerably higher for the human users than the bots as we can see a ramp-up to peak value reached relatively early followed by a decline for larger values of k. Now, when we compare the behavior of persistence, visible through the rate of decay after the peak of the curve, we can observe significant differences. In the case of humans, the adoption probability drastically falls with increasing magnitude of exposure, indicating less effect of repeated exposure in increasing the adoption by their followers. Interestingly, for the bots, the adoption rate remains constant with increased exposure level from k: 1 - 4, and suddenly increases for k : 4-6 before falling off eventually. The shape of the curve represents a higher persistence of conversation for tweets by the bots compared to humans. The more sticky and less persistent behavior of the human exposure curve suggests that if the audience of the human users does not adopt or retweet the idea put forward by them after the initially small number of exposures, the chances of the idea getting adopted later on are very marginal. The repeated exposures having significant marginal effects on the probability of retweeting the seed user after k^{th} exposure hints towards the behavior of complex contagion present in the communication demonstrated by the tweets of the bots.

4 Discussion and Conclusion

Our work tries to shed light on the influence of Twitter bots on human users beyond the statistical metrics of influence volume garnered by them. The research questions we frame are inspired by the social engineering strategies the automated accounts could possibly adopt to meet their campaign level objectives. The bots we studied, in the context of the Italian Mayoral election, garnered influence from wider users in the political spectrums, reflecting that they were not merely strengthening the echo-chambers, but also influencing Twitter users with different political ideologies. It can be argued that much of the success of automated accounts also depends upon influencing users outside their echo chambers, rather than merely gaining traction from users on the same side of the spectrum. Our content level analysis of the tweets using the multiple attributes of negative emotions reveal that the bots were able to gain larger influence using a relatively lesser amount of toxicity level on their text. The social bots showing an upper hand on the information diffusion mechanisms and the novelty on the diffusion audience they reached are concerning findings as it is getting even more difficult to detect them.

Our work contributes to the broader research call on identifying and quantifying the social strategies used by the automated agents in social networks. Our future work constitutes the important need for devising methods that can help us identify the strategies used by the social bots using graph-based and machine learning metrics on the conversation graph of social networks. We also need to replicate the research settings on the datasets of other political scenarios of different countries where social bots were discovered to be used and influential to the public. The similarities and differences of the social engineering methods by bots across different contexts, complemented by the understanding of the underlying political atmosphere can help us understand more about the threats these agents can possess on social movements of the future, and possibly predict it before time. We believe uncovering the strategies of the social bots and the tools they use for information manipulation will aid us in designing better systems to detect them and gain an upper hand on the arms race of social digital spam polluting the conversation landscape of our social networks

References

1. Bessi, A. and Ferrara, E., 2016, November. Social bots distort the 2016 US Presidential election online discussion. First Monday, 21(11-7).
2. Howard, P. N. and Kollanyi, B., 2016, June. Bots,# StrongerIn, and# Brexit: computational propaganda during the UK-EU referendum. Available at SSRN 2798311.
3. Woolley, S. C., 2016, March. Automating power: Social bot interference in global politics. First Monday, 21(4).
4. Mustafaraj, E., and Metaxas, P. T., 2010. From obscurity to prominence in minutes: Political speech and real-time search.
5. Ratkiewicz, J., Conover, M. D., Meiss, M., Gonçalves, B., Flammini, A., and Menczer, F. M., 2011, July. Detecting and tracking political abuse in social media. In Fifth international AAAI conference on weblogs and social media.

6. Rathnayake, C., and Buente, W., 2017, Feb. Incidental effects of automated retweeting: An exploratory network perspective on bot activity during Sri Lanka's presidential election in 2015. Bulletin of Science, Technology & Society, 37(1), 57-65.
7. Everett, R. M., Nurse, J. R., and Erola, A., 2016, April. The anatomy of online deception: What makes automated text convincing?. In Proceedings of the 31st Annual ACM symposium on applied computing (pp. 1115-1120). ACM.
8. Zhang, J., Zhang, R., Zhang, Y., and Yan, G., 2013, Oct. On the impact of social botnets for spam distribution and digital-influence manipulation. In 2013 IEEE Conference on Communications and Network Security (CNS) (pp. 46-54). IEEE.
9. Paudel, P., Nguyen, T.T., Hatua, A. and Sung, A.H., 2019. How the Tables Have Turned: Studying the New Wave of Social Bots on Twitter Using Complex Network Analysis Techniques. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 501-508). ACM.
10. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., and Tesconi, M. 2017, April. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In Proceedings of the 26th international conference on World Wide Web companion (pp. 963-972). International World Wide Web Conferences Steering Committee.
11. Llewellyn, C., Cram, L., and Favero, A., 2016, June. Avoiding the Drunkard's search: Investigating collection strategies for building a Twitter dataset. In 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL) (pp. 205-206). IEEE.
12. Romero, D. M., Meeder, B., and Kleinberg, J., 2011, March. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In Proceedings of the 20th international conference on World wide web (pp. 695-704). ACM.
13. Ferrara, E., and Yang, Z., 2015, Sep. Quantifying the effect of sentiment on information diffusion in social media. PeerJ Computer Science, 1, e26.
14. Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R., 2015, Oct. Tweeting from left to right: Is online political communication more than an echo chamber?. Psychological science, 26(10), 1531-1542.
15. Liang, H., 2018, Apr. Broadcast versus viral spreading: the structure of diffusion cascades and selective sharing on social media. Journal of Communication, 68(3), 525-546.
16. Kušen, E. and Strembeck, M., 2018. Why so emotional? An analysis of emotional bot-generated content on Twitter. In Proc. of the 3rd International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS) (pp. 13-22). SciTePress.
17. Davis, C.A., Varol, O., Ferrara, E., Flammini, A. and Menczer, F., 2016, Apr. Botornot: A system to evaluate social bots. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 273-274). International World Wide Web Conferences Steering Committee.