2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Socialbots' First Words: Can Automatic Chatting Improve Influence in Twitter?

Alkiviadis Savvopoulos, Pantelis Vikatos†, Fabrício Benevenuto*
Computer Engineering and Informatics, University of Patras, Greece
*Universidade Federal de Minas Gerais (UFMG), Brazil
†IIT-CNR, Pisa, Italy
asavv@ceid.upatras.gr, p.vikatos@iit.cnr.it, fabricio@dcc.ufmg.br

***Abstract*—Online social networks (OSNs) such as Twitter and Facebook constitute an open space for developers to create sophisticated machines that imitate human users by automating their social network activities. The existence of socialbots is so powerful that bots can be transformed to influential users. Previous studies use fundamental functions such as posting a tweet or creating links to a specific target group to investigate the ability to infiltrate of these accounts. Our study analyzes the role of automated chatting in bots' infiltration. Our analysis is compared with the state of the art of this kind and reveals that the chat functionally was able to improve Klout and follow ratio on about 24% and 123% respectively. Also, the advanced communications skills contribute to more message interactions between socialbots and other accounts. Based on our empirical experimental study the conversational socialbots infiltrate more successfully in the Twitter-sphere.**

***Index Terms*—Twitter social network, Social bots, Infiltration strategies**

I. Introduction

Create bots, artificial intelligence entities, able to pass by humans has long been a challenge for the AI community. Recently, social networks provide a suitable environment for bots to pass by humans due to the weak form identify allowed on these systems, associated with lack of context, informality, and overload of information experienced by their users. Many efforts showed that social networks, particularly Twitter, are full of socialbots [1, 2, 3]. Their goals vary from commercial chatbots, i.e Pandora Bot [1], to attempts of mass manipulation during political events [4, 5].

Broadly speaking, there are two existing kinds of research efforts related to socialbots: 1) approaches that aim at identifying and detecting socialbots; and 2) approaches that aim at deploying socialbots to measure the social network response to its design options. The main solutions for the first kind of approach are currently based on machine learning techniques and explore behavioural and structural features [3, 6, 7]. Even DARPA launched a Twitter bot challenge in which different teams attempted to identify the bots in a labeled dataset [8]. Overall, these efforts are of extreme importance as bots can be used to automatize any malicious activity in social networks, including search spam [6], manipulation of trending topics [9], and link farming [10]. The limitation of this approach is usually associated with the construction of the labeled dataset based on humans evaluation, making the trained model of the classifiers limited by the capacity of humans to recognize a socialbot. Thus, aiming at identifying other behaviors that favor socialbots, the second type of effort in this field aims at creating and deploying social bots with different behaviors and design options in order to evaluate their infiltration capacity and the response of the social network to each different socialbot new feature or implemented behavior [11, 12].

In this paper, we contribute to the second form of existing efforts by investigating if chatting, i.e., establishing automatic conversations through direct messages, can help socialbots to infiltrate in Twitter. In similar previous studies, the functionality of chat was not investigated, assuming that this could count negatively to the socialbot and even be used by users to decide to follow back or not. We also explore other new functions from Twitter, such as liking a tweet from other users, to check if these new interactive actions support are able to create engagement among socialbot and users. To do that we deployed a set of socialbots in twitter that use state of the art techniques to chat and we compared its infiltration with existing bots. We found that the chat functionally was able to improve Klout and follow ratio on about 24% and 123% respectively. Regarding message interactions, the socialbot that has the ability to chat receives four more messages than the one with the basic activities. Next, we describe the methodology used to develop and evaluate our socialbots. Then, we provide empirical results and summarize our findings and conclusions.

II. Methodology

Our methodology is inspired by the study [12], which describes in detail infiltration strategies of social bots and presents a comparison study between an alternative set of attributes in order to examine the factors of the most influential bots.

Overall, authors show that a female bot tends to become more influential, a high activity account that implements actions such as posting, re-posting, following, and mentioning others increases the opportunity for other users to interact with the bot, reflecting on its final influence score. In addition, the bots that post tweets automatically tend to be more reliable than the bots with only the re-posting function. Another factor

[1] http://www.pandorabots.com

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

which affects how socialbots are able to engage socially is the set of target users with whom the socialbot attempts to interact. The study [12] shows that bots that only post about specific topics trends and those target users who are interested in this topic (hashtag), tend to be more successful. This study did not evaluate a one-to-one chat as a feature for the socialbots. So, if a user contacts the bot privately, she does not receive an answer.

A. Social bots

We implemented as the baseline the socialbot configurations described in [12]. Precisely, we configure the sociabots with the following setup.

Profile: We used an English female first name and English last name. The email account was created based on a combination of the first and last name using Gmail. As profile picture, we have used a female image from friends who allowed us to use their pictures for experimental purposes. The biography contains a small description of job, interests and location (e.g, Journalist - Covering World News around the world, London, England). Finally, we used as the background a cover image related to the interests that biography describes.

Activities: Our socialbots performs one of the possible activities and sleeps a random time between 1-60 minutes. The baseline activities consist of posting automatically generated tweets using the Markov strategy described in [12] or re-posting recent tweets from a set of previously selected hashtags. Another activity consists of following users with common interests, which are the users who post one of our previously selected target hashtags.

B. New Socialbot Activities

Twitter provides functions for users to like, retweet or comment on a specific post also allow users to establish one-to-one conversations trough direct private messages. We consider that by using these extra functions, socialbots can improve their interactions and communication skills and then infiltrate better the Twitter network. We differentiate from the socialbots using additional functions as the following lists presents. We call the proposed socialbot as conversational socialbot. More specifically, we implemented the following additional activities.

- **Like**: The socialbot selects a person from the corresponding group and performs a 'Like' in a recent post.
- **Retweet**: The socialbot posts a retweet from a user that already posted a target hashtag.
- **Comments**: The socialbot sends a comment to a recent post from a random user of a target group.
- **Direct Messages**: The socialbot replies to direct messages from other users. Additionally, the socialbot can initiate a conversation through a direct message with her followers as part of one of her actions.

Next, we evaluate each of this new feature separately in order to evaluate the outcome for a socialbot that implements each of them.

C. Ethical considerations of the study

In this study, the tweets and statuses of the socialbots are either re-tweets of existing public messages or a result of a word-generating model which operates on a tweet based corpus. User mentions and URLs are excluded from the tweet posting process, to establish spam-free and user-friendly posts. More important, users are free to decide if they want to follow a socialbot or not, and they can unfollow if they disliked the content they receive in their timelines. The experimentation concluded after thirty days when all the socialbot accounts were deactivated. The usernames of the socialbots and of the users who connected and interacted with them will remain undisclosed.

III. Experimental Results

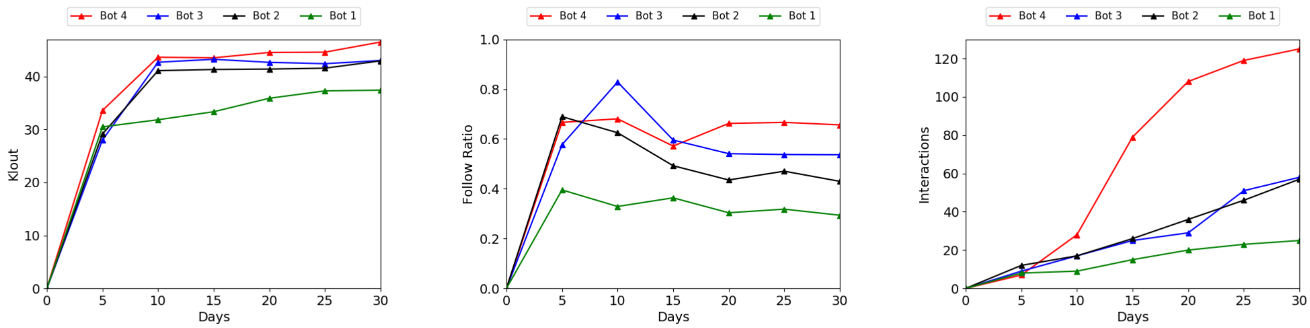
A. Implementation

We created a set of accounts on Twitter. The socialbots were differentiated in the functions as Table I presents. We contracted our experiments using Twitter API and tweepy [2] which is python module for activation of the automated functions.

Also, we defined the socialbots to follow a number of initial accounts and to have a number of posts and likes. More specifically the socialbots followed each other. Our experiment started with 10 re-posts from trend tweets and 10 likes for each socialbot. We implemented the automated generation of posts using a simple Markov model [3], as described in [12], created based on twenty articles from CNN news and Telegraph.

The chat function was implemented using PyAIML [4], which is an interpreter package for AIML, the Artificial Intelligence Markup Language with a standard set of conversations. The conversations set was expanded with crawled dialogues from trend hashtags of Twitter in the same period that the experiment was conducted. Every time the socialbot received a message the listener of Twitter streaming API activated the PyAIML module. The response of PyAIML was sent back to the initial sender, maintaining a proper conversation. In terms of comments, the socialbot selected a recent post from a random user of the corresponding group and isolated the post's text field. Then the module of PyAIML produced the comment based on the extracted text. Furthermore, socialbots initiated a conversation by selecting randomly followers and igniting a controversy regarding their own tweets. The socialbots were running over a period of 30 days. In addition, all socialbots go to sleep in a random time between 00:00 - 02:00 and are reactivated at some time 10:00-12:00 GMT+2 time zone, mimicking the expected sleeping time of human users. We checked to what extent socialbots can gain popularity and influence in the Twitter social network and the best performance of each bot type set is presented using the following metrics:

[2] $http://www.tweepy.org$
[3] $https://pypi.python.org/pypi/markovify$
[4] $https://pypi.python.org/pypi/PyAIML$



(a) Klout Comparison

(b) Follow Ratio Comparison

(c) Message Interactions Comparison

Fig. 1: Metrics Performance

TABLE I: Socialbots Configuration

Bot Type	Activities
Bot 1	copy post, markov post, follow
Bot 2	Bot 1 activities + like
Bot 3	Bot 2 activities + comment and retweet
Bot 4	Bot 3 activities + direct message

- Follow ratio: This is a standard metric for estimating the popularity of followers an account related to his/her Twitter friends (following).
- Klout score: Klout score is a popular measure for online influence which ranges from 1 to 100, with higher scores implying a higher online social influence of a user.
- Number of message-based interactions: We measure the number of times other users interact with a socialbot through messages (tweets), such as when some user @mentions the socialbot, replies to the socialbot, retweets or favorites a tweet posted by the socialbot.

B. Results

We analyze the improvement in the performance by using the new set of activities and comparing with the basic activities of Bot 1 as they described in the previous section. Figure 1a provides the Klout of each bot during the 1 month period. We infer that there is no significant difference between socialbots until the end of the fifth day. The "like" activity of Bot 2 improves the Klout performance about 15%. The additional activities of the comments and the retweets of Bot 3 lead to almost the same Klout scores as Bot 2. However, during the 10 - 30 day the Bot 4 reaches significantly higher Klout performance. The 10 day constitutes the peak of Klout improvement with 37% for Bot 4 and at the end of 30 day the Klout metric reaches the highest value 46.4, which corresponds to 24% improvement.

Follow ratio express the level of visibility for each bot. We infer from Figure 1b that the follow ratio of Bot 4 is higher than the others after 2 weeks activity. The Bot 4 has the highest follow ratio score that equals 0.65 that in terms of improvement means 123% of better performance. Also, we infer a significant improvement in Bot 2 and Bot 3 that equals 35% and 68% respectively.

The metric of message interactions is related to the communication of bots with other users. Figure 1c provides the comparison between the different type of socialbots. The new features contribute to the enhancement of the communication skills of the socialbots. Hence, the Bot 2 and Bot 3 reach 57 and 58 total message interactions which correspond to 1.284 and 1.38 more messages to each socialbot respectively. We infer that the existence of direct message functionality improves much higher than other socialbots. The Bot 4 reaches the 125 total interactions which it corresponds to four times more message interactions.

C. The Impact of Automatic Chatting

We consider that user's attention is gained through reacts to personal messages or comments. The behaviour of users shows that our conversational social bots are able to maintain a discussion alive and react appropriately to users' direct messages. The use of AIML contributes to this dimension in developing the rules and the knowledge base. Regarding AIML rule is consisted of a pair of an input sentence and the reply to it. All pair elements are mapped to the AIML tags $< pattern >$ and $< template >$. All patterns are transformed into the rules of an atomic category and the rules are formed in the following example:

```
<category>
<pattern>Hi</pattern>
<template>Hello world!</template>
</category>
```

Based on the users' reactions in our conversational socialbots, we infer that the socialbot of type Bot 4 is able to maintain alive a discussion. In the following text boxes, there are examples of conversations between Twitter users and our socialbots. The first example constitutes a part of the discussion which is initiated by a Twitter user.

User: Have a good afternoon
Bot 4 : Afternoon has been good so far.
User : What did you do then ?
Bot 4 : I talked to people on the web.
User : And who do you support ?
Bot 4 : Charlie Parker.

TABLE II: The rate of unfollow during a month period for each bot type

Bots	Unfollow Rate
Bot 1	0.297
Bot 2	0.304
Bot 3	0.291
Bot 4	0.218

In the next, Bot 4 is able to handle conversations that use slang words and phrases.

User : fu
Bot 4 : That kind of language will get you nowhere.
User : haha
Bot 4 : I'm glad you thought it was funny.

The following example represents an initiation of a conversation from socialbot side.

Bot 4 : I think I dont entirely agree with your last tweet...
User : Ok legend may be a strong word, but it was only because I called him a legend...
Bot 4 : Are you sure this is a fact ?
User : Never said it was a fact, just said in gest

Furthermore, we checked how chatting functionally influences the rate of Twitter users that unfollow the socialbots. Our results in Table II present that socialbots with high conversational skills have less possibility to lose their linkages than the others.

IV. Conclusion

We investigate a set of new features that were not explored before for socialbots in Twitter. The use of likes, as well as comments and retweets, showed an increase of about 15% in Klout metric. Also, the follow ratio was improved about 35% using the ability to like posts of the users in the target group. Retweets and comments improved more the follow ratio reaching the 68% improvement. Our main research question was to check if the ability to chat would not affect negatively the performance of a socialbot, but it turns out that it affected positively. Our results showed that in terms of Klout, follow ratio and messages interactions social bot with advanced communications skills have better performance. We inferred that at the end of our empirical experimental procedure the Klout achieve 24% better performance. Also, the visibility and message interactions of conversational socialbots increased 1.23 times and 4 times more respectively. Furthermore, we inferred that socialbots that are able to chat can also keep their follow linkages more than other types of bots. Existing behavior-based detection strategies consider that socialbots might not interact too much. We show that "off-the-shelf" libraries to allow a socialbot to automatically chat can be deployed and be used to favor socialbots and make it even better to infiltrate twitter. As future work, we are interested in examining the infiltration strategies of socialbots in different social networks such as Facebook and Instagram.

V. Acknowledgments

This work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme.

References

[1] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Communications of the ACM*, vol. 59, no. 7, pp. 96–104, 2016.

[2] J. Edwards, "There are 20 million fake users on twitter, and twitter can't do much about them–business insider," 2013.

[3] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?" *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6, pp. 811–824, 2012.

[4] A. Bessi and E. Ferrara, "Social bots distort the 2016 u.s. presidential election online discussion," *First Monday*, November 2016. [Online]. Available: http://firstmonday.org/ojs/index.php/fm/article/view/7090

[5] B. Kollanyi, P. N. Howard, and S. C. Woolley, "Bots and automation over twitter during the first us presidential debate," *Comprop Data Memo*, 2016.

[6] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on twitter," in *Proceedings of the 7th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*, 2010.

[7] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on twitter." in *Proc. The 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.

[8] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, and F. Menczer, "The darpa twitter bot challenge," *Computer*, vol. 49, no. 6, pp. 38–46, 2016.

[9] A. Chakraborty, J. Messias, F. Benevenuto, S. Ghosh, N. Ganguly, and K. P. Gummadi, "Who makes trends? understanding demographic biases in crowdsourced recommendations," in *Proceedings of the International AAAI Conference on Web and Social Media*, ser. ICWSM'17.

[10] S. Ghosh, B. Viswanath, F. Kooti, N. K. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. P. Gummadi, "Understanding and combating link farming in the twitter social network," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 61–70.

[11] J. Messias, L. Schmidt, R. Oliveira, and F. Benevenuto, "You followed my bot! transforming robots into influential users in twitter," *First Monday*, vol. 18, no. 7, 2013.

[12] C. Freitas, F. Benevenuto, S. Ghosh, and A. Veloso, "Reverse engineering socialbot infiltration strategies in twitter," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 2015, pp. 25–32.