

Why a Socialbot is Effective in Twitter? A Statistical Insight

Mohd Fazil

Department of Computer Science
Jamia Millia Islamia, New Delhi, India
Email: mohdfazil.jmi@gmail.com

Muhammad Abulaish, SMIEEE

Department of Computer Science
South Asian University, New Delhi, India
Email: abulaish@sau.ac.in

Abstract—Twitter, a popular microblogging platform, facilitates users to express views and thoughts on any topic of discussion using short messaging texts limited to 140 characters. Due to its open and real-time information sharing and dissemination nature, it is abused by socialbots for political astroturfing, advertising, spamming, and other illicit activities. To this end, we injected an army of 98 socialbots associated to top six Twitter using countries to study socialbots’ infiltration behaviour. In this paper, we present a statistical insight derived through the analysis of the captured data by our socialbots. Socialbots’ profile features, such as age, gender, etc. and their behavioural impact on infiltration performance are studied and presented, wherein a user’s *following* activity to a socialbot is considered as an infiltration. Experimental results and subsequent statistical analyses show that socialbots’ profiles belonging to India were the successful in duping highest number of users, whereas Indonesian socialbots were least infiltrative. Moreover, among various Twitter activities, *following* is found to be the most effective activity for infiltrating a user. Among the intruded users, trace of the presence of botnets, spammers, and other malicious users have also been observed and presented in this paper.

Index Terms—Social network analysis, Twitter data analysis, Socialbots identification, Socialbots characterization, Twitter spam.

I. INTRODUCTION

Now a days majority of people, particularly younger generation, are registered on one or more Online Social Networks (OSNs) that facilitate them to connect and keep in touch with their family members, friends, acquaintances, and colleagues irrespective of their geographical location and boundary. OSNs are now enormously used for news propagation, entertainment, gaming, thought expression, and so on. In the contrary, the affordable accessibility and easy to use functionality of OSNs have also proved them as heaven for criminals and defaulters provoking to heavily use them for all possible ill activities causing breeding of severe problems, such as spamming, cyberbullying, cyberstalking, identity theft, botnets, and many other ill activities. Most of the cyber crimes are generally executed by creating fake profiles which is prevalent since the inception of the internet technology, particularly OSNs. Fake profiles are evolving and getting sophisticated every day. Socialbot is one such sophisticated version of the malicious bots. Socialbots are automated programs that mimic human behaviour to gain users’ trust and then exploit it to carry out complicated activities [1]. Their injection and deployment is

mostly state supported and generally used for political astroturfing, propaganda diffusion, etc. against rivals [2]. Twitter, as a microblogging platform, seems ideal for socialbots to carry out such sophisticated attacks. In order to understand the socialbots’ working behaviour and their characterization, and to identify the category of users and regions that are vulnerable to socialbots’ infiltration, we have captured data through injecting socialbots in Twitter network and analyzed it at different levels of granularity.

Many researchers have done the experiment of socialbots injection in OSNs [1], [3], [4] but, to the best of our knowledge, no one has ever analyzed to reveal the socialbots’ profile features that are more infiltrative in nature. Moreover, the regional behaviour of the socialbots has not been analyzed in the existing literature. Although, there is no provision of regional networks in Twitter, we can assign a user in the proximity of a particular country by setting profile’s time-zone as of the time-zone of the respective country. In this study, an army of 98 socialbots were injected in the Twitter network, where the number of socialbots assigned to a country was proportional to the user-base of the country. We have adjusted the bots’ characteristics such as age, gender, etc. as per the Twitter statistics. Our socialbots network was active for about 1 month before being detected and suspended by the Twitter defense mechanism. After analyses of the logged data, we have identified some interesting facts that are summarized below:

- Among the top six Twitter using countries, socialbots associated to India were most successful in intrusion, whereas Indonesia associated socialbots were least effective. This finding can be justified by the Symantec corporation’s annual security report highlighting that social media related scams are at surge in India with second highest in Asia, and facts from other report¹ also justify it.
- Among the various users’ activities on Twitter, “following” is proved to be most affluent activity for socialbots in alluring and duping users to follow them.
- Trace of other socialbot networks operating in Twitter has been observed, but ambitions does not seem clear as

¹<http://timesofindia.indiatimes.com/tech/tech-news/cyberattack-india-among-the-most-vulnerable-nations/articleshow/51346401.cms>

they stay away themselves from spamming or information polluting.

II. RELATED WORK

In recent past, researchers have tried to conceive socialbots' behavior and their impact on OSNs and users. There are number of instances where socialbots and their misuse have been observed and reported in the form of propaganda diffusion, political astroturfing [5], [6], identity theft [7], etc. In addition, some competitions have also been organized to observe their impact. In [1], [8], socialbots network creation is defined along with exposure of inherent vulnerabilities existing in OSNs. In [1], authors have thoroughly analyzed economic feasibility of the attack and reported the inherent vulnerabilities of social network. Aiello et al. [9] analyzed the individual bot's capability by creating a profile on aNobii network to study how a passive user without reputation and trust at creation time, moves in the list of top influencing users of the network, simply by surfing and investigating other users profile. In [10], authors have characterized and predicted users that can be easily persuaded to engage with socialbots using certain set of features along with ranking the features. Unlike [9], authors in [4], [11] have used active approach for infiltration and intentionally target specific group of people. In [4], socialbots have breached a technical organization using the information revealed by its employees on Facebook, demolishing the belief that security aware users can not be infiltrated. In [12], authors proved through experiment that socialbots can easily manipulate OSNs' reputation metrics such as Klout score and Twitalyzer and some of the socialbot gained score near to celebrities.

Our study is first of its kind in which vulnerability of the users of top six Twitter using countries is analyzed albeit Twitter has no provision of partitioning network as regional network. We have adjusted socialbots' profile attributes as per their country and accordingly adjusted time-zone. With all these settings, friend and other recommendations by the Twitter were from the same region as of the profile.

III. PRELIMINARIES AND EXPERIMENTAL SETUP

In this section, we present the socialbots injection process in detail – starting from profile creation to running the whole socialbots network for approx. one month time duration.

A. Profile Creation and Distribution

Presence of a person on an OSN is determined by an account having his/her personal information, such as name, address, age, gender, and so on. Therefore, we created socialbots' profile manually to set attributes as per the requirements rather than purchasing it from available vendors or using automated profile creation tool. Number of socialbots for a country, C , is set as a proportion of the user-base share of C to the total user-base of the top six Twitter using countries. It is calculated using equation 1, where NS_i is the number of socialbots assigned to the i^{th} country, C_i is the user base share of the i^{th} Twitter using country. Distribution of Twitter

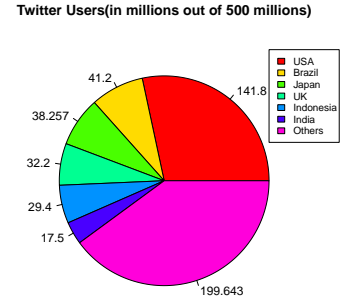


Fig. 1: Twitter user distribution

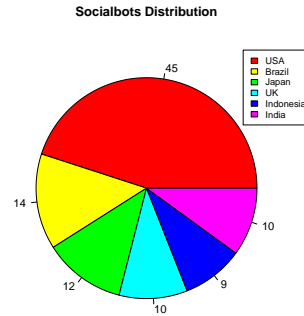


Fig. 2: Country-wise socialbots distribution

user-base around the world with USA leading the list and the socialbots are shown in figures 1 and 2, respectively. In figure 2, the number of socialbots for a country is slightly different from the one calculated using equation 1 to make the socialbot count for a country to be at least 10% of the total number of socialbots. Number of male and female socialbots is kept as 47 and 51, respectively meeting the Twitter gender distribution. Genders of the profiles are exhibited by name, profile picture, and description. Within a country, there were equal number of male and female socialbots. Since date of birth is available only for few profiles, profile pictures were used to determine the age group of profiles. Number of socialbots for different age segment was as per Twitter age distribution and within age segments ages are determined randomly. Thus, in our experiment, we have tried our best to mimic Twitter distributions to set different users profile features. All socialbot profiles were created between 1st November 2015 to 3rd January 2016.

$$NS_i = \frac{C_i}{\sum_{i=1}^6 C_i} \times 100 \quad (1)$$

B. Socialbots Injection and Operation

Following the profile setting process, we operated them on Twitter. An activity by a user requires the user to be

logged in, but this can be avoided by accessing the Twitter through API using an application. Therefore, we developed an application called *TrueBots* using the python library codebird-php. Application authenticated every socialbot using OAuth and saved the generated secret keys in a profile database. After authentication, profiles got activated at any time between 30 minutes to (8 times the number of authenticated socialbots) minutes for the first time activity. Subsequent activities by a socialbot require only the credentials. Application accessed the Twitter using the REST API. First activity was performed as per the following rule, where rn is a binary random variable.

$$Activity = \begin{cases} \text{Follow 5 to 10 celebrities,} & \text{if } rn = 0 \\ \text{Follow 10 to 20\% of intra-} & \\ \text{country socialbots,} & \text{Otherwise} \end{cases}$$

For the first case, socialbots followed celebrities irrespective of their home country, whereas for the second case socialbots followed 10 to 20% of the socialbots from their own country within a time period of 30 minutes to 2 days. This is just for first activity, and subsequent activities were performed as described above. Next activation time for a bot was determined at current activation, and target user set for the socialbots were crawled either from the follower list of a celebrity hailing from the same country or from the follower list of the followers of the socialbot. It is due to the fact that though celebrities have fans from all over the globe, generally majority of their followers used to be from their home country. In order to evade network-based detection approach, we maintained follower to following ratio for a socialbot to a threshold level of 0.25. Whenever this ratio dropped below the threshold, socialbots randomly unfollowed friends, and whenever users followed socialbots, they started following back users to comply with social etiquette that ultimately increases the number of followers for the socialbots. In Twitter, tweets are generally users personal thoughts, whereas retweets show agreement with other users' thought. During whole experiment, tweets were posted in either of the three ways – (i) posted quotes stored in a database, (ii) posted tweets crawled from the trending topics as of their own with some tuning or retweeted the tweets from trending topics, (iii) posted tweets that were crawled and logged from followers timeline. We have not used any automated sentence generation technique like Markov chains for tweet generation because algorithmic sentences can be easily identified as proved in [3]. The whole network was active for one month between 4th January 2016 to 3rd February 2016 and all activities were monitored and logged in local data repository.

IV. INFILTRATION PERFORMANCE ANALYSIS

In this section, we present an analysis of profile features efficacy for manipulating network shape and infecting users trust. Based on socialbots grouping, our analysis is grouped into two parts – (i) General analysis, and (ii) Country-based analysis.

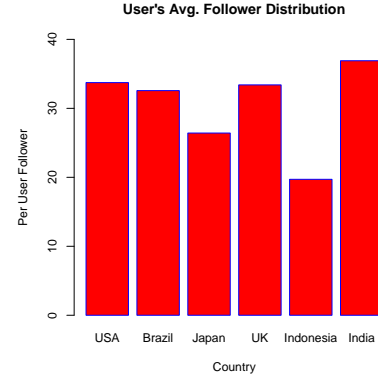


Fig. 6: Socialbots' average followers per country

A. General Analysis

In this, we present statistical analysis of all socialbots as a single entity without any grouping. We have analyzed the effect of the socialbots' age and gender on Twitter users and on overall intrusion mechanism. Further details about these analyses are presented in the following sub-sections.

1) *Age-Based Analysis*: This section presents an analysis of the efficacy of a socialbot's profile picture and inferred age in luring and persuading other users to follow them back. In Twitter, only 0.45% users provide date of birth². Therefore users infer age of other users based on their profile pictures, descriptions, alphanumeric used in twitter handle, and so on. In our experiment, all socialbots, except 21, have profile picture. On analysis, we found that socialbots with young profile picture were much successful in attracting followers, which is apparent in figure 3. Average number of followers for socialbots younger than 50 years is 33, whereas for older than 50 years, it is only 28. Follower rate for older is moderate due to few socialbots. One such profile used Hillary Clinton as profile picture hiding bot's age and alluring democrat workers and followers, advancing him/her as highest followers gainer. Cumulative distribution of number of followers trapped by socialbots is shown in figure 4.

2) *Gender-Based Analysis*: In Twitter, there is no option for gender exhibition. In our experiment, gender of a profile is unveiled by profile picture and there are equal number of male and female socialbots. Gender-based cumulative distribution of the followers gained by socialbots is shown in figure 5, which displays an interesting observation that gender of a profile play no role in trapping followers, except the case of profiles with seductive profile picture and description enticing the users. We avoid using seductive profile picture for female profiles to imitate them as normal profiles except for two profiles having obscene profile picture. We find that both the profiles gained large number of followers with number stood at 61 and 40, respectively that are much higher than average follower rate 33 for the female profiles.

²<http://www.beevolve.com/twitter-statistics/#a2>

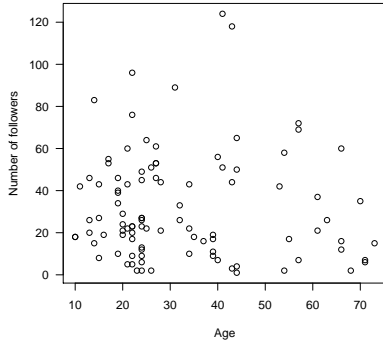


Fig. 3: Number of followers vs age of socialbots

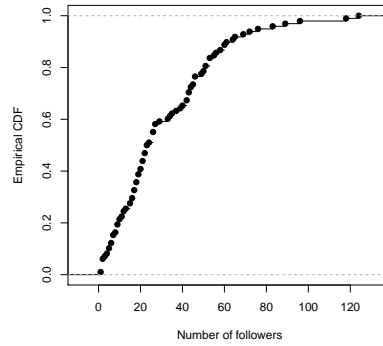


Fig. 4: CDF of socialbots' followers

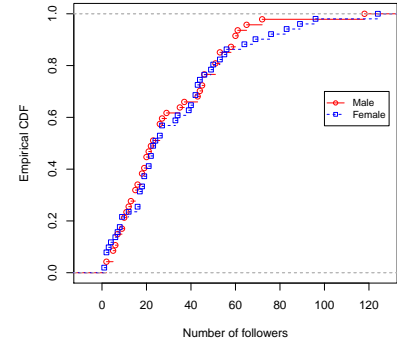


Fig. 5: CDF of gender-wise socialbots' followers

B. Country-Based Analysis

This section demonstrates the effect of socialbots' feature on intrusion using the results grouped by the country of socialbots, though in Twitter there is no regional splitting of network on the basis of continent or country. Country assignment of socialbots means location, time-zone, and other profile features that were adjusted as per their assigned country. Twitter indexes trending topics and recommend friends to a user based on the user time-zone. Average number of followers for socialbots associated to each country is shown in figure 6, which represents that socialbots linked to India were the most damaging (i.e., users being most vulnerable), whereas Indonesian socialbots were least. This exposes users' vulnerabilities towards such hostile attacks.

1) *Profile Feature*: In OSNs, users are acknowledged and known by profiles. Therefore, profile attributes are vital, and accordingly we adjusted socialbots' attributes and analyzed their efficacy. Normal users, generally use their original descent and charming picture. We have used real looking morphed pictures. In this analysis, all results are grouped on the basis of country, and it is found that socialbots with profile picture were more successful in enticing users with average 34 followers, whereas socialbots without profile picture lured only on average 25 followers. We have also analyzed seductiveness of profile pictures by two profiles having exposed profile picture. It is found that each of the two profiles entices more users than other socialbots. It concludes that users are more vulnerable towards pornographic and seductive profiles, fascinating the users but most of the followers of such profiles are themselves obscene and suspicious.

2) *Age*: This subsection presents an analysis of the effect of socialbots' age on infiltration performance. There are socialbots without profile pictures, but they are assigned an age group albeit difficult for others to guess age for such profiles, nullifying the age factor in their infiltration efficacy. Socialbots are divided into two age groups – socialbots younger than 30 years, and socialbots older than 30 years. We have compared the infiltration performance of the both age-group socialbots

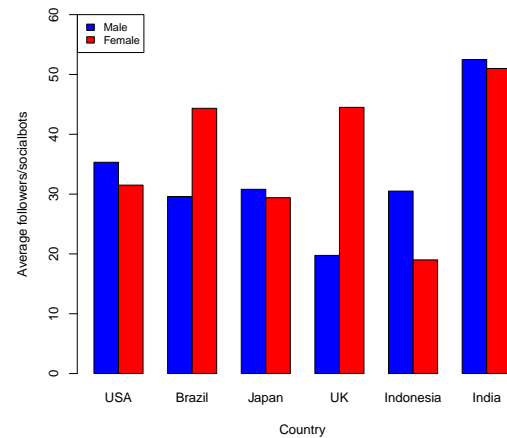


Fig. 10: Average no. of followers for the two gender groups of socialbots across the country

of each country in acquiring average number of followers. Figure 7 shows that there is minor difference between the two age groups in reaping followers, except UK, where younger socialbots were more successful. Here age groups include profiles without profile picture and it might have lead biasness in result. In order to observe exact age impact, results on only those profiles having profile picture are shown in figure 8. It is interesting that Indian socialbots who are older than 30 years have significantly more average followers than average followers without age consideration. UK is found to be the only country where older bots have lower average number of followers than the average follower rate for all socialbots from UK, which is apparent in figure 9. Another interesting observation is that except USA, other country users are significantly attracted by captivating profile pictures. Thus, it can be inferred that Twitter users are influenced by the age of an account for following a user, with old age profiles considered as more reliable.

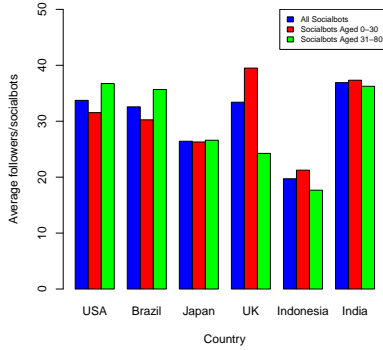


Fig. 7: Average no. of followers grouped by country for two different age groups of socialbots

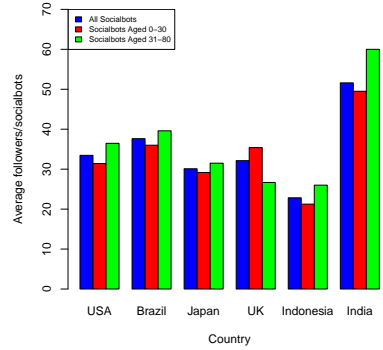


Fig. 8: Average no. of followers across the country for the socialbots having profile pictures

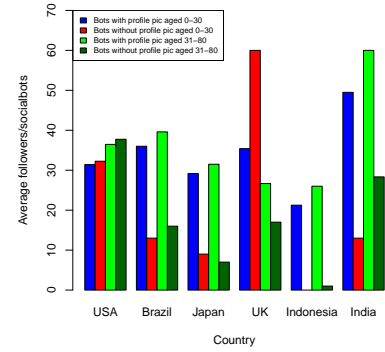
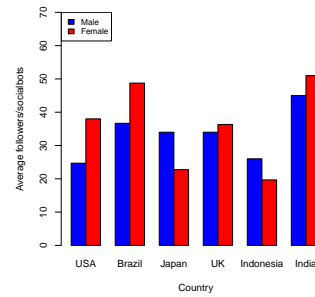


Fig. 9: Average no. of followers for the two age groups socialbots with and without profile pictures across the country

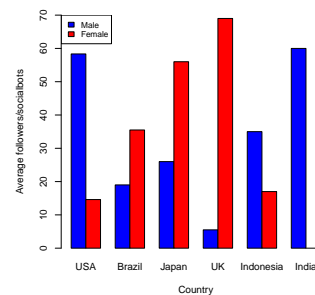
3) *Gender*: This section presents an analysis of the effect of gender of a socialbot on trapping users with the results grouped by socialbots' countries. Here, we have considered only those socialbots that have a profile picture. Figure 10 portrays the average followers per socialbots grouped by gender for each country. On analysis, we found that female profiles from UK and Brazil are much dominating to male counterpart neutralizing the effect of male profile's performance of the remaining four countries. Critical analysis reveals a very interesting fact that female profiles from these two countries were exposing and alluring, whereas socialbots from Indonesia have not used profile picture for female profiles except one, and consequently failed to grab users attention. We have also analyzed collaborative effect of age and gender on infiltration, which is shown in figure 11. It can be observed that younger female socialbots are dominating to their male counterpart, whereas older male socialbots are more appealing than their female counterpart.

4) *Activity*: So far, we have analyzed the effect of only profile features like age and gender on infiltration, but in cyber space, like the physical world, engagement of a user with network regulates the user's position in the network. In case of OSNs, active users are frequently chosen for recommendation as well as recommended to others on the basis of homophily and some other measures. Active users are assets for OSN service providers and they try to monetized these users. In this section, we have investigated the effect of different socialbots activities and their efficacy to infiltrate a network of trust and tried to answer the following questions: (a) Is there exist any casual relationship between activity and infiltration? (b) If yes, then what is the level of the relationship? In Twitter, people follow users, post tweets, retweet tweets, reply to tweets, mark tweets as favourite, etc. Therefore, we have considered the following activity set including four activitiest: $\{following, retweeting, tweeting, favourite marking\}$.

Socialbots were not made capable to reply as there is no



(a) Socialbots younger than 30 years



(b) Socialbots older than 30 years

Fig. 11: Average followers gained by the socialbots for the two age groups, grouped on the basis of gender and country

technique that can generate intelligent sentences at par with human-generated sentences or that can not be detected by detection systems [3]. In OSNs, activities of the friends of a user work as input for the user, impacting and regulating the user's activities and responses. So, users are like computational devices and their activities are affected by their neighbours

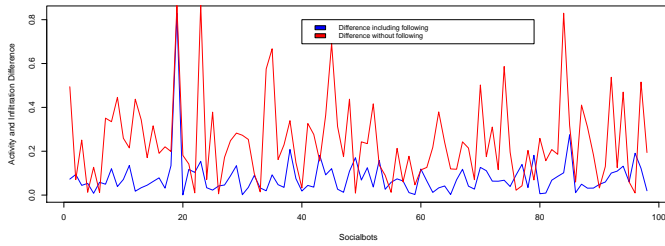


Fig. 12: Activity set and associated infiltration performance difference for two variant of activity set

and surroundings. We found correlation coefficient between activity set and infiltration performance as 0.85, and this high value shows strong bonding between activity set and followers gained. Z-normalized difference between activity values and followers count is plotted and shown in figure 12 using blue line, where low value discloses high association between two entities. This Figure also plots difference between activity set, excluding following, and followers gained as shown using red line, which generally has high value, revealing that except following other activities do not had any significant impact on infiltration. It is also clear from the correlation coefficient values which sharply declined at 0.22 from 0.85 when following is excluded from the activity set. We have analyzed the correlation between individual activity component and followers, and it is found that ‘following’ is highly dominated activity and it has highest correlation with infiltration performance. It may be due to the case that ‘following’ is directly associated with the user followed and it is more eye-catching activity comparing to retweet, tweet, and other activities.

V. ARE SOCIALBOTS FOLLOWERS OF OTHER SOCIALBOTS, BOTNETS OR CONTENT POLLUTERS?

OSNs are very fascinating platform for criminals, spammers, and fraudsters starting from its evolution. Social media platforms are proved to be heaven for malicious users and they are continuously creating fake profiles and bots in large number to carry out ill-activities. Facebook and Twitter are no exception from this hazard. Following the experiment, we tried different tactics like direct messaging, following from new accounts, but only few responded. It was also observed that number of socialbots followers have been suspended by the Twitter. During manual scrutiny of profiles, many of them appeared to be pornographic, spammers, etc. It is also observed that number of USA-based followers were advertisers. One socialbot grabbed 124 followers with all of them, excluding 16, having no profile description, location, and date of birth. All of these were part of a community of thousand profiles with alike tweeting pattern, account creation date, profile pictures, etc., but they did not seem to posse hostile or malicious behaviour.

VI. CONCLUSION

In this paper, we have presented a statistical analysis of profile features and their impact on infiltration behaviour of the socialbots. We have considered top six Twitter using countries and analyzed the infiltration efficacy of the socialbots at different levels of granularity. Some of the key findings in this study are as follows: (i) Socialbots linked to India are found to be most infiltrating and captivating for the users, (ii) Gender of a socialbot do not play any significant role, except those with young and exposing profiles, (iii) Profile attribute setting is impressive, when profiles are continuously functional, and among the activity set *following* is found to be most effective activity, (iv) Among the grabbed follower set, footprints of various form of fake and malicious profiles such as spammers, bots, content polluters are observed.

REFERENCES

- [1] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, “Design and analysis of a social botnet,” *Computer Networks*, vol. 57, no. 2, pp. 556–578, 2013.
- [2] N. Abokhodair, d. Yoo, and D. W. McDonald, “Dissecting a social botnet: Growth, content and influence in twitter,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*. Vancouver, BC, Canada: ACM, 2015, pp. 839–851.
- [3] C. A. 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*. Paris, France: IEEE Computer Society/ACM, 2015, pp. 25–32.
- [4] A. Elyashar, M. Fire, D. Kagan, and Y. Elovici, “Homing socialbots: intrusion on a specific organizations employee using socialbots,” in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Niagara Falls, Canada: IEEE Computer Society/ACM, 2013, pp. 1358–1365.
- [5] J. Ratkiewicz, M. Conover, M. Meiss, B. Goncalves, S. Patil, A. Flammini, and F. Menczer, “Detecting and tracking political abuse in social media,” in *Proceedings of the 5th Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, Spain: AAAI Press, 2011.
- [6] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, “Using sentiment to detect bots on twitter: Are humans more opinionated than bots?” in *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Beijing, China: IEEE Computer Society/ACM, 2014, pp. 620–627.
- [7] L. P. Bilge, T. Strufe, D. Balzarotti, and E. Kirde, “All your contacts are belong to us: Automated identity theft attacks on social networks,” in *Proceedings of the 2014 IEEE/ACM International Conference on World wide web*. Madrid, Spain: ACM, 2009, pp. 551–560.
- [8] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, “The socialbot network: When bots socialize for fame and money,” in *Proceedings of the 27th Annual Computer Security Applications Conference*. Orlando, Florida, USA: ACM, 2011, pp. 93–102.
- [9] L. M. Aiello, M. Deplano, R. Schifanella, and G. Ruffo, “People are strange when youre a stranger: impact and influence of bots on social networks,” in *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*. Dublin, Ireland: AAAI Press, 2012, pp. 10–17.
- [10] R. Wald, T. M. Khoshgoftaar, A. Napolitano, and C. Sumner, “Which users reply to and interact with twitter social bots?” in *Proceedings of the 25th International Conference on Tools with Artificial Intelligence*. Washington DC, USA: IEEE, 2013, pp. 135–144.
- [11] J. Zhang, R. Zhang, Y. Zhang, and G. yan, “On the impact of social botnets for spam distribution and digital influence manipulation,” in *Proceedings of the 6th IEEE International Conference on ommunications and Network Security*. National Harbour, MD, USA: IEEE Communications Society, 2012, pp. 46–54.
- [12] J. Messias, L. Schmidt, R. A. R. Oliveira, and F. Benevenuto, “You followed my bot! transforming robots into influential users in twitter,” *First Monday*, vol. 18, no. 7, 2013.