Detecting Malicious Social Bots: Story of a Never-Ending Clash

Stefano Cresci(✉)

Institute for Informatics and Telematics (IIT-CNR), Pisa, Italy
stefano.cresci@iit.cnr.it

Abstract. Recently, studies on the characterization and detection of social bots were published at an impressive rate. By looking back at over ten years of research and experimentation on social bots detection, in this paper we aim at understanding past, present, and future research trends in this crucial field. In doing so, we discuss about one of the nastiest features of social bots – that is, their *evolutionary* nature. Then, we highlight the switch from supervised bot detection techniques – focusing on feature engineering and on the analysis of one account at a time – to unsupervised ones, where the focus is on proposing new detection algorithms and on the analysis of groups of accounts that behave in a coordinated and synchronized fashion. These unsupervised, group-analyses techniques currently represent the state-of-the-art in social bot detection. Going forward, we analyze the latest research trend in social bot detection in order to highlight a promising new development of this crucial field.

Keywords: Social bots · Bot evolution · Reactive detection · Proactive detection · Adversarial machine learning · Generalizability

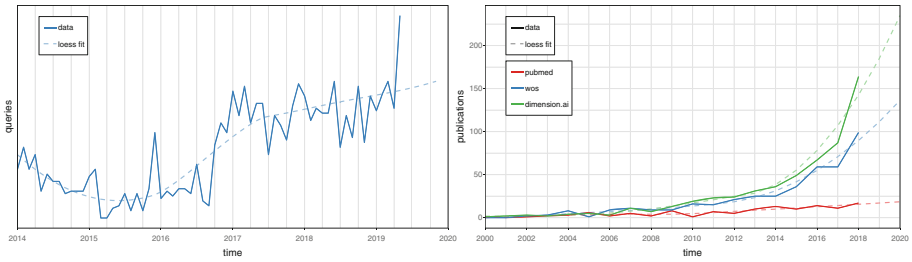
1 Introduction

Social media and Online Social Networks (OSNs) are having a profound impact on our everyday life, giving voice to the crowds and reshaping the information landscape. Indeed, the deluge of real-time data spontaneously shared in OSNs already proved valuable in many different domains, spanning tourism [7], safety and security [3,4], transportation and politics [14,23], to name but a few notable cases.

However, the democratizing effect of OSNs does not come without costs [6]. In 2016, "post-truth" was selected by the Oxford dictionary as the word of the year, and in 2017 "fake news" was selected for the same purpose by Collins dictionary. Still in 2017, the World Economic Forum raised a warning on the potential distortion effect of OSNs on user perceptions of reality[1]. Moreover, the same openness of OSNs that favored the democratization of information

[1] http://reports.weforum.org/global-risks-2017.

C. Grimme et al. (Eds.): MISDOOM 2019, LNCS 12021, pp. 77–88, 2020.
https://doi.org/10.1007/978-3-030-39627-5_7



(a) Normalized Google queries per month. (b) Publications per year.

Fig. 1. Trends in search queries and publications regarding social bots.

(e.g., the support for programmatic access via APIs and the support for anonymity), also inevitably favored the proliferation of *social bots*. Indeed, previous studies report that social bots are as old as OSNs themselves [18]. With the term social bot, we broadly refer to computer programs capable of automatically producing, re-sharing, and liking content in OSNs, or even capable of establishing and maintaining social relations. In fact, any of our supposedly online friends may instead be a fake, automated account, part of large coordinated groups [18].

Not all social bots are malicious and dangerous, and some of them also serve beneficial purposes, such as contributing to gather accurate information in the aftermath of emergencies [2,25]. Unfortunately, however, the vast majority actually pursue malicious goals. These malicious bots try to hide their automated nature by imitating the behaviors of legitimate users. Moreover, they often act in a synchronized and coordinated fashion – a strategy that collectively allows them to increase their impact. Many recent studies concluded that social bots played a role in strategic information operations orchestrated in the run up to several major political elections, both in western and eastern countries [32,33]. As additional evidence for this claim, Twitter recently banned several thousands accounts, linked to many different malicious information operations perpetrated between 2016 and 2019[2]. Other recent studies also suggested that social bots were used to exacerbate online social discussions about controversial topics (e.g., vaccination and immigration debates), thus increasing polarization and fueling abusive and hateful speech [34]. Across the whole Twittersphere, it is reported that social bots account for 9 to 15% of total active platform users [35]. Even more worryingly however, when strong political or economical incentives are at stake, the presence of bots exponentially increases. As an example, a recent study reported that 71% of all users mentioning stocks traded in US financial markets, are likely to be bots [10].

Since social bots have a central role in the diffusion of disinformation, spam, and malware, both scholars and practitioners devoted much effort to the development of detection techniques. Nowadays, new studies on the characterization and detection of social bots are published at an impressive rate, as shown in

[2] https://about.twitter.com/en_us/values/elections-integrity.html.

Fig. 1. An analysis of a subset of publications from 2018 reports that more than 3 new papers were published (on average) every week on the topic of social bots[3]. The rapidly growing publication trend suggests that in the near future there will be one new paper published every day, which poses a heavy burden on researchers trying to keep pace with the evolution of this field. This issue is also emphasized by the lack of a thorough survey. Perhaps more importantly, the rate at which new studies on this topic are published implies that a huge effort is taking place worldwide in order to overcome the diffusion of social bots. Given this picture, an important question arises: *where is all this effort leading?*

In the remainder of this paper we try to answer this crucial question via a longitudinal analysis of ten years of research in the field of social bot detection.

2 Traditional Social Bot Detection

The first work that focused on the detection of misbehaving accounts in OSNs dates back to January 2010 [38]. Since then and until present days, the vast majority of attempts at bot detection have been based on heuristics (i.e., rule-based) or on supervised machine learning [9]. An important implication of the adoption of supervised machine learning is that each account is analyzed singularly. In other words, given a group of accounts to investigate (e.g., an OSN community), the detection technique is separately applied to each account of the group, to which it assigns a label (either bot or legitimate). In fact, the key assumption of this large body of work is that each bot/fake/spammer has peculiar features that make it clearly distinguishable from legitimate accounts. This approach to the task of social bot detection, which we call "traditional", thus revolves around the application of off-the-shelf machine learning algorithms on the accounts under investigation, rather than on developing new algorithms. Indeed, most of the works in this branch are focused on designing machine learning features – that is, they are focused on the task of feature engineering – capable of maximizing detection performances of well-known algorithms, such as SVM, decision trees, random forests, and more [9].

Regarding features to exploit for the detection, 3 classes have been mainly considered: (i) profile features [8,15]; (ii) features extracted from the posts, such as posting behavior and content of posted messages [5,28]; and (iii) features derived from the social or interaction graph of the accounts [22,26]. The classes of features exploited by the detection technique have a strong impact on both the performances of the detector as well as its efficiency. For instance, in Twitter it has been demonstrated that those features that mostly contribute towards the predictive power of bot detectors (e.g., graph-based features such as measures of centrality in the social graph), are also the most costly ones, in terms of needed data and computation [8].

Despite achieving promising initial results, the traditional approach – which still comprises the majority of papers published nowadays – has a number of drawbacks. The first challenge in developing a supervised detector is related

[3] Source: https://www.dimensions.ai/.

to the availability of a ground truth (i.e., labeled) dataset, to be used in the learning phase of the classifier. In most cases, a real ground truth is lacking and the labels are simply given by human operators that manually analyze the data. Critical issues arise since, as of 2019, we still lack a "standard" definition of what a social bot is [21,37]. Moreover, humans have been proven to suffer from several biases [29] and to largely fail at spotting modern, sophisticated bots, with only $\simeq 24\%$ bots correctly labeled as such by humans [9].

The biggest drawback of traditional approaches, however, is due to the evolutionary nature of social bots, which we discuss in the following section.

3 The Issue of Bot Evolution

Early success at social bot detection, in turn, inevitably inspired countermeasures by bot developers. Because of this, newer bots often feature advanced characteristics that make them way harder to detect with respect to older ones. This iterative process, that leads to the development of always more sophisticated social bots, is commonly referred to as *bot evolution*.

A noteworthy work published in 2011, and later extended in 2013 [36], provided the first evidence and the theoretical foundations to study social bot evolution. The first wave of social bots that populated OSNs until around 2011 was made of rather simplistic bots – mainly accounts with very low perceived reputation (e.g., few social connections and posted messages) and featuring clear signs of automation (e.g., repeated spam of the same URLs). On the contrary, the social bots studied in [36] appeared as more popular and credible, given the relatively large number of their social connections. In addition, they were no longer spamming the same messages over and over again, but they were instead posting several messages with the same meaning but with different words, in order to avoid detection techniques based on content analysis. Starting from these findings, authors of [36] also proposed a supervised machine learning classifier that was specifically designed for detecting *evolving* bots. Their classifier simultaneously leveraged features computed from the content of posted messages, social connections, and tweeting behaviors, and initially proved capable of accurately detecting the sophisticated bots. More recently, new studies provided evidence of a third generation of social bots that spread through OSNs from 2016 onwards [9,18]. Unfortunately, the classifier originally developed in [36] was no longer successful at detecting the third wave of social bots, as shown in [9].

The previous example serves as anecdotal evidence of bot evolution, and of the detrimental effect it has on bot detectors. Additional evidence is reported in [9], where authors evaluated the *survivability* of different bots, and the ability of humans in spotting bots in the wild. Specifically, authors of [9] showed that only $\simeq 5\%$ of evolved bots are removed from social platforms (i.e., high survivability), whilst "old" social bots are removed $\simeq 60\%$ of the times (i.e., low/moderate survivability). Moreover, in a large-scale crowdsourcing experiment, tech-savvy social media users proved unable to tell apart evolved bots and legitimate users, 76% of the times (i.e., 3 out of 4 evade detection by humans).

The same users were instead unable of spotting "old" social bots only 9% of the times (i.e., only 1 out of 10 evades detection) [9].

What results reported in [9,18] ultimately tell us, is that current sophisticated bots are practically indistinguishable from legitimate accounts, if analyzed one at a time. In other words, the results about bot evolution tell us that the assumption of traditional (i.e., supervised) bot detection approaches, according to which bots have features that allow to distinguish them from legitimate accounts, is no longer true.

4 Modern Social Bot Detection

The difficulties in detecting sophisticated bots with supervised approaches that are based on the analysis of individual accounts, recently gave rise to a new research trend that aims to analyze groups of accounts as a whole. This new research trend is also motivated by the interest of platform administrators in detecting what they typically refer to as "coordinated inauthentic behavior"[4,5].

Since 2013, several different research teams independently started to propose new techniques for social bot detection. Despite being based on different key concepts, all these new techniques – that collectively represent the "modern" approach to social bot detection – included important contributions also from the algorithmic point of view, thus shifting from general-purpose machine learning algorithms such as SVMs and decision trees, to ad-hoc algorithms that were specifically designed for detecting bots. Furthermore, the majority of these new algorithms considered groups of accounts as a whole, rather than single accounts, thus moving in the direction of detecting the coordinated and synchronized behavior that characterizes malicious botnets [9].

As a consequence of this paradigm-shift, modern bot detectors are particularly effective at detecting evolving, coordinated, and synchronized bots. For instance, the technique discussed in [13] associates each account to a sequence of characters that encodes its behavioral information. Such sequences are then compared between one another to find anomalous similarities among sequences of a subgroup of accounts. The similarity is computed by measuring the longest common subsequence shared by all the accounts of the group. Accounts that share a suspiciously long subsequence are then labeled as bots. Instead, the family of systems described in [22,26] build a bipartite graph of accounts and their interactions with content (e.g., retweets to some other tweets) or with other accounts (e.g., becoming followers of other accounts). Then, they aim to detect anomalously dense blocks in the graph, which might be representative of coordinated and synchronized attacks. Another recent example of an unsupervised, group-based technique is RTbust [27], which is tailored for detecting mass-retweeting bots. The technique leverages unsupervised feature extraction and clustering. An LSTM autoencoder converts the retweet time series of accounts

[4] https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/.

[5] https://help.twitter.com/en/rules-and-policies/platform-manipulation.

into compact and informative latent feature vectors, which are then clustered by a hierarchical density-based algorithm. Accounts belonging to large clusters characterized by malicious retweeting patterns are labeled as bots, since they are likely to represent retweeting botnets.

Given that bot detection techniques belonging to this modern approach still represent the minority of all published papers on social bot detection, we still lack a through and systematic study of the improvement brought by the modern approach to social bot detection. However, the first preliminary results that compared the detection performances of traditional and modern detectors on the same datasets, seem to support the increased effectiveness of the latter. In particular, the technique introduced in [13] outperformed several traditional detectors on two datasets, yielding an average $F1$ improvement of +0.37. Similarly, RTbust [27] improved on a widely used traditional bot detector by increasing $F1$ of +0.44. The promising results with modern bot detectors tell us that focusing on groups is advantageous. In fact, large groups of coordinated bots are more likely to leave traces of automation than a single bot, independently of how sophisticated the individual bots are [9]. By performing analyses at group level, this modern approach appears to be able to raise the bar for bot developers to evade detection. Furthermore, the majority of modern bot detectors are semi-supervised or unsupervised, which gives higher guarantees on the generalizability of the detector and mitigates challenges related to the acquisition of a reliable ground-truth.

5 The Way Ahead

So far, we highlighted that a shift is taking place in the development of bot detectors, in order to counter the evolutionary nature of social bots. Now, by looking at the latest advances in this thriving field, we aim at gaining some insights into the future of social bot detection.

Notably, both the traditional and the modern approach to social bot detection have always followed a *reactive* schema. Quite naturally, the driving factor for the development of new and better bot detectors have been bot mischiefs themselves. As soon as scholars and OSN administrators identified a new group of bots, possibly featuring new and advanced characteristics, they started the development of detectors capable of spotting them. A major implication of this reactive approach is that improvements in bot detection are possible only after having collected evidence of new bot mischiefs. In turn, this means that scholars and OSN administrators are constantly one step behind of bot developers, and that bots have a significant time span (i.e., the time needed to design, develop, and deploy a new detector) during which they are essentially free to tamper with our online environments.

However, another – radically different – approach to social bot detection is possible, and has just started being investigated by several researchers. This trailblazing direction of research involves the application of adversarial machine learning [19] to bot detection. Adversarial machine learning has already been

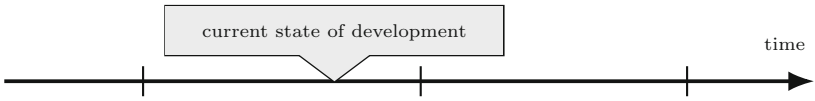
applied to a number of fields such as computer vision [24] and speech recognition [30], with exceptional results. In general, it is considered as a machine learning paradigm that can be profitably applied to all scenarios that are intrinsically adversarial (i.e., with adversaries interested in fooling machine learning models) [19], with social bots detection clearly being one of such scenarios [17]. In the so-called *adversarial social bot detection*, scholars try to find meaningful adversarial examples with which to test current bot detectors [11]. In other words, this branch of research aims at studying possible attacks to existing bot detectors, with the goal of building more robust and more secure detectors. In this context, adversarial examples might be sophisticated types of existing bots that manage to evade detection by current techniques [1], or even bots that do not exist yet, but whose behaviors and characteristics are simulated [12], or bots developed ad-hoc for the sake of experimentation [20]. Finding good adversarial examples can, in turn, help scholars understand the weaknesses of existing bot detection systems, before such weaknesses are effectively exploited by bot developers. As a result, bot hunters need not wait anymore for new bot mischiefs in order to adapt their techniques, but instead they can *proactively* test them, in an effort that could quickly make them more robust. Among the positive outcomes of adversarial approaches to bot detection, is a more rapid understanding of the drawbacks of current detectors and the opportunity to gain insights into new features for achieving more robust and more reliable detectors.

Despite the high hopes placed on adversarial social bot detection, this research direction is still in its infancy. The very first works in this field have in fact been published just in 2018 and 2019. Adversarial approaches to social bot detection thus represent a promising new development of this field. However, efforts at adversarial social bot detection can only be successful if the scientific community decides to rise to the many open challenges. Among the challenges opened up by proactive and adversarial approaches is the development of techniques for creating many different kinds of *adversarial examples*, with which to test existing bot detectors. A task that, to date, was only tackled by relying on the creativity of some researchers and only for a few limited cases [11,12,20]. Moreover, adversarial approaches have proved computationally and data intensive in some of the early tasks to which they were applied, with only few solutions proposed to date to boost their efficiency [31]. Another challenge thus revolves around assessing the *efficiency* of adversarial social bot detection, as well as its *coverage* of the possible types of attacks (i.e., how likely it is with the adversarial approach to anticipate a real future attack or a real future evolution of bots).

6 Conclusions

Our longitudinal analysis of the first decade of research in social bot detection revealed some interesting trends in the development of bot detectors. In particular, we identified 3 ages of bot detection: (i) the *traditional age*, characterized by the study of account features and by the adoption of off-the-shelf supervised machine learning algorithms; (ii) the *modern age*, characterized by the development of ad-hoc unsupervised algorithms for detecting groups of colluding bots;

Table 1. The analysis of more than a decade of research and experimentation in social bot detection allows to identify 3 main directions of research, corresponding to 3 different ages: the traditional, the modern, and the adversarial age. In turn, each age is characterized by a few distinctive features reported above. Furthermore, an analysis of recently published papers on social bot detection, positions current endeavors somewhere in between the traditional and the modern ages.



	traditional	**modern**	**adversarial**
key concept	features allow to tell apart bots and legitimate accounts	synchronization and coordination allow to detect botnets	improve bot detectors by finding their weaknesses
development focus †	features (e.g., via feature engineering)	detection algorithms	adversarial examples
method ‡	supervised, off-the-shelf ML (e.g., decision trees, SVMs)	unsupervised, ad-hoc algorithms	adversarial ML
target §	single accounts	groups of accounts	bot detectors

†: what scholars aim to optimize
‡: which machine learning (ML) paradigm scholars adopt
§: to what scholars apply their method

and (iii) the newborn *adversarial age*, whose promise is to apply the paradigm of adversarial machine learning to the task of bot detection. Given the considerable amount of work still needed to lay the foundations of adversarial social bot detection, the adversarial age has not really sparked yet. However, if it lives up to its expectations, it might blossom soon with a tremendous impact. Apart from the adversarial age, the characteristics of currently published works in social bot detection still highlight a majority of traditional detectors. However, the gap between newly proposed traditional and modern detectors is narrowing. Hence we can conclude that the peak of the traditional age is probably over, and that we are moving towards the peak of the modern age, as pictorially shown in Table 1.

The exponentially growing body of work on social bot detection shown in Fig. 1, somehow reassures us that much effort is bound to be devoted to the fight of this critical issue. However, at the same time it also poses some new challenges. Firstly, it is becoming more and more important to be able to organize this large body of work. Doing so would not only contribute to a better exploitation of this knowledge, but would also allow researchers in bot detection to more effectively and more efficiently provide new solutions (e.g., avoid wasting time and effort on solutions that have already proved unsuccessful). Unfortunately, thorough and comprehensive surveys on bot detection are still few and far between. To this regard, this paper aims to provide a contribution to the critical review and analysis of the vast literature in this field. Secondly, more papers on this topic inevitably imply that more bot detectors will

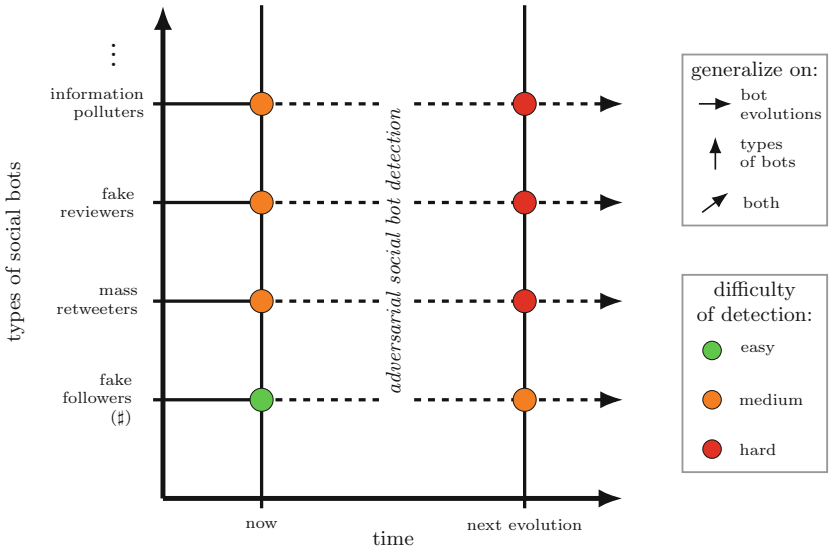


Fig. 2. A bi-dimensional theory of generalizability for social bot detectors. Let us consider a detector developed for a specific kind of bots (marked with ♯). The detector will likely achieve its best performances when used against the same bots it was developed for (green-colored scenario). However, it would be useful to also evaluate its detection performances against different kinds of bots, thus moving along the y axis. Furthermore, by exploiting adversarial social bot detection, it could also be possible to estimate its detection performances against evolved bots, thus moving along the x axis of the generalizability space. The hardest foreseeable evaluation scenario is the one where a detector is tested against evolved versions of bots for which it was not originally designed (red-colored). The vast majority of newly proposed bot detectors are only evaluated in the easiest scenario. (Color figure online)

be proposed. With the growing number of disparate detection techniques, it is thus becoming increasingly important to have standard tools (e.g., frameworks, reference datasets, methodologies) to evaluate and compare them. In particular, one facet of bot detectors that is often overlook is their *generalizability* – that is, their capability in maintaining good detection results also for types of bots that have not been originally considered. To this regard, the analyses carried out in this study lay the foundations for a bi-dimensional theory of generalizability, as shown in Fig. 2. A desirable scenario for the near future would involve the possibility to easily evaluate any new bot detector against many different types of social bots in order to assess its strengths and weaknesses, for instance by following the approach laid out in [16]. It would also be profitable to be able to evaluate detectors against possible evolved versions of current bots, by applying the adversarial approach previously described. In order to reach this ambitious goal, we must first create reference datasets that comprise several different kinds

of bots, thus significantly adding to the sparse resources existing as of today[6]. Then, as already anticipated, we should also devise additional ways for creating a broad array of diverse adversarial examples. These challenges currently stand as unsolved, and call for the highest effort of our scientific community.

Acknowledgments. This research is supported in part by the EU H2020 Program under the scheme `INFRAIA-1-2014-2015: Research Infrastructures` grant agreement #654024 *SoBigData: Social Mining & Big Data Ecosystem*.

References

1. Assenmacher, D., Adam, L., Frischlich, L., Trautmann, H., Grimme, C.: Openbots. arXiv preprint arXiv:1902.06691 (2019)
2. Avvenuti, M., Bellomo, S., Cresci, S., La Polla, M.N., Tesconi, M.: Hybrid crowdsensing: a novel paradigm to combine the strengths of opportunistic and participatory crowdsensing. In: ACM WWW Companion (2017)
3. Avvenuti, M., Cresci, S., Del Vigna, F., Fagni, T., Tesconi, M.: CrisMap: a big data crisis mapping system based on damage detection and geoparsing. Inf. Syst. Front. **20**(5), 993–1011 (2018)
4. Avvenuti, M., Cresci, S., Marchetti, A., Meletti, C., Tesconi, M.: Predictability or early warning: using social media in modern emergency response. IEEE Internet Comput. **20**(6) (2016)
5. Chavoshi, N., Hamooni, H., Mueen, A.: DeBot: Twitter bot detection via warped correlation. In: IEEE ICDM (2016)
6. Cresci, S.: Harnessing the social sensing revolution: challenges and opportunities. Ph.D. dissertation, University of Pisa (2018)
7. Cresci, S., D'Errico, A., Gazzé, D., Lo Duca, A., Marchetti, A., Tesconi, M.: Towards a DBpedia of tourism: the case of Tourpedia. In: ISWC (2014)
8. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: Fame for sale: efficient detection of fake Twitter followers. Decis. Support Syst. **80** (2015)
9. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: The paradigm-shift of social spambots: evidence, theories, and tools for the arms race. In: ACM WWW Companion (2017)
10. Cresci, S., Lillo, F., Regoli, D., Tardelli, S., Tesconi, M.: Cashtag piggybacking: uncovering spam and bot activity in stock microblogs on Twitter. ACM Trans. Web **13**(2), 11 (2019)
11. Cresci, S., Petrocchi, M., Spognardi, A., Tognazzi, S.: From reaction to proaction: unexplored ways to the detection of evolving spambots. In: ACM WWW Companion (2018)
12. Cresci, S., Petrocchi, M., Spognardi, A., Tognazzi, S.: Better safe than sorry: an adversarial approach to improve social bot detection. In: ACM WebSci (2019)
13. Cresci, S., Pietro, R.D., Petrocchi, M., Spognardi, A., Tesconi, M.: Social fingerprinting: detection of spambot groups through DNA-inspired behavioral modeling. IEEE Trans. Dependable Secure Comput. **15**(4), 561–576 (2018)
14. D'Andrea, E., Ducange, P., Lazzerini, B., Marcelloni, F.: Real-time detection of traffic from Twitter stream analysis. IEEE Trans. Intell. Transp. Syst. **16**(4), 2269–2283 (2015)

[6] https://botometer.iuni.iu.edu/bot-repository/datasets.html.

15. Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F.: BotOrNot: a system to evaluate social bots. In: ACM WWW Companion (2016)
16. De Cristofaro, E., Kourtellis, N., Leontiadis, I., Stringhini, G., Zhou, S., et al.: LOBO: evaluation of generalization deficiencies in Twitter bot classifiers. In: ACM ACSAC (2018)
17. Ferrara, E.: The history of digital spam. Commun. ACM **62**(8), 82–91 (2019)
18. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Communun. ACM **59**(7) (2016)
19. Goodfellow, I.J., McDaniel, P.D., Papernot, N.: Making machine learning robust against adversarial inputs. Communun. ACM **61**(7), 56–66 (2018)
20. Grimme, C., Assenmacher, D., Adam, L.: Changing perspectives: is it sufficient to detect social bots? In: Meiselwitz, G. (ed.) SCSM 2018. LNCS, vol. 10913, pp. 445–461. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91521-0_32
21. Grimme, C., Preuss, M., Adam, L., Trautmann, H.: Social bots: human-like by means of human control? Big Data **5**(4) (2017)
22. Jiang, M., Cui, P., Beutel, A., Faloutsos, C., Yang, S.: Catching synchronized behaviors in large networks: a graph mining approach. ACM Trans. Knowl. Discov. From Data **10**(4) (2016)
23. Kavanaugh, A.L., et al.: Social media use by government: from the routine to the critical. Gov. Inf. Q. **29**(4), 480–491 (2012)
24. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: IEEE ICCV (2017)
25. de Lima Salge, C.A., Berente, N.: Is that social bot behaving unethically? Commun. ACM **60**(9), 29–31 (2017)
26. Liu, S., Hooi, B., Faloutsos, C.: HoloScope: topology-and-spike aware fraud detection. In: ACM CIKM (2017)
27. Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W., Tesconi, M.: RTbust: exploiting temporal patterns for botnet detection on Twitter. In: ACM WebSci (2019)
28. Miller, Z., Dickinson, B., Deitrick, W., Hu, W., Wang, A.H.: Twitter spammer detection using data stream clustering. Inf. Sci. **260**, 64–73 (2014)
29. Pandey, R., Castillo, C., Purohit, H.: Modeling human annotation errors to design bias-aware systems for social stream processing. In: IEEE/ACM ASONAM (2019)
30. Pascual, S., Bonafonte, A., Serrà, J.: SEGAN: speech enhancement generative adversarial network. In: Interspeech (2017)
31. Sahay, R., Mahfuz, R., Gamal, A.E.: A computationally efficient method for defending adversarial deep learning attacks. arXiv preprint arXiv:1906.05599 (2019)
32. Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.C., Flammini, A., Menczer, F.: The spread of low-credibility content by social bots. Nat. Commun. **9**(1) (2018)
33. Starbird, K., Arif, A., Wilson, T.: Disinformation as collaborative work: surfacing the participatory nature of strategic information operations. In: ACM CSCW (2019)
34. Stella, M., Ferrara, E., De Domenico, M.: Bots increase exposure to negative and inflammatory content in online social systems. Proc. Nat. Acad. Sci. **115**(49) (2018)
35. Varol, O., Ferrara, E., Davis, C.A., Menczer, F., Flammini, A.: Online human-bot interactions: detection, estimation, and characterization. In: AAAI ICWSM (2017)
36. Yang, C., Harkreader, R., Gu, G.: Empirical evaluation and new design for fighting evolving Twitter spammers. IEEE Trans. Inf. Forensics Secur. **8**(8), 1280–1293 (2013)

37. Yang, K.C., Varol, O., Davis, C.A., Ferrara, E., Flammini, A., Menczer, F.: Arming the public with artificial intelligence to counter social bots. Hum. Behav. Emerg. Technol. **1**(1), 48–61 (2019)
38. Yardi, S., Romero, D., Schoenebeck, G., et al.: Detecting spam in a Twitter network. First Monday **15**(1) (2010)