DOI:10.1145/2818717

Today's social bots are sophisticated and sometimes menacing. Indeed, their presence can endanger online ecosystems as well as our society.

BY EMILIO FERRARA, ONUR VAROL, CLAYTON DAVIS, FILIPPO MENCZER, AND ALESSANDRO FLAMMINI

The Rise of Social Bots

BOTS (SHORT FOR software robots) have been around since the early days of computers. One compelling example of bots is chatbots, algorithms designed to hold a conversation with a human, as envisioned by Alan Turing in the 1950s.[33] The dream of designing a computer algorithm that passes the Turing test has driven artificial intelligence research for decades, as witnessed by initiatives like the Loebner Prize, awarding progress in natural language processing.[a] Many things have changed since the early days of AI, when bots like Joseph Weizenbaum's ELIZA,[39] mimicking a Rogerian psychotherapist, were developed as demonstrations or for delight.

Today, social media ecosystems populated by hundreds of millions of individuals present real incentives—including economic and political ones—to design algorithms that exhibit human-like behavior. Such ecosystems also raise the bar of the challenge, as they introduce new dimensions to emulate in addition to content, including the social network, temporal activity, diffusion patterns, and sentiment expression. A social bot is a computer algorithm that automatically produces content and interacts with humans on social media, trying to emulate and possibly alter their behavior. Social bots have inhabited social media platforms for the past few years.[7,24]

a www.loebner.net/Prizef/loebner-prize.html

Engineered Social Tampering

What are the intentions of social bots? Some of them are benign and, in principle, innocuous or even helpful: this category includes bots that automatically aggregate content from various sources, like simple news feeds. Automatic responders to inquiries are increasingly adopted by brands and companies for customer care. Although these types of bots are designed to provide a useful service, they can sometimes be harmful, for example when they contribute to the spread of unverified information or rumors. Analyses of Twitter posts around the Boston marathon bombing revealed that social media can play an important role in the early recognition and characterization of emergency events.[11] But false accusations also circulated widely on Twitter in the

» key insights

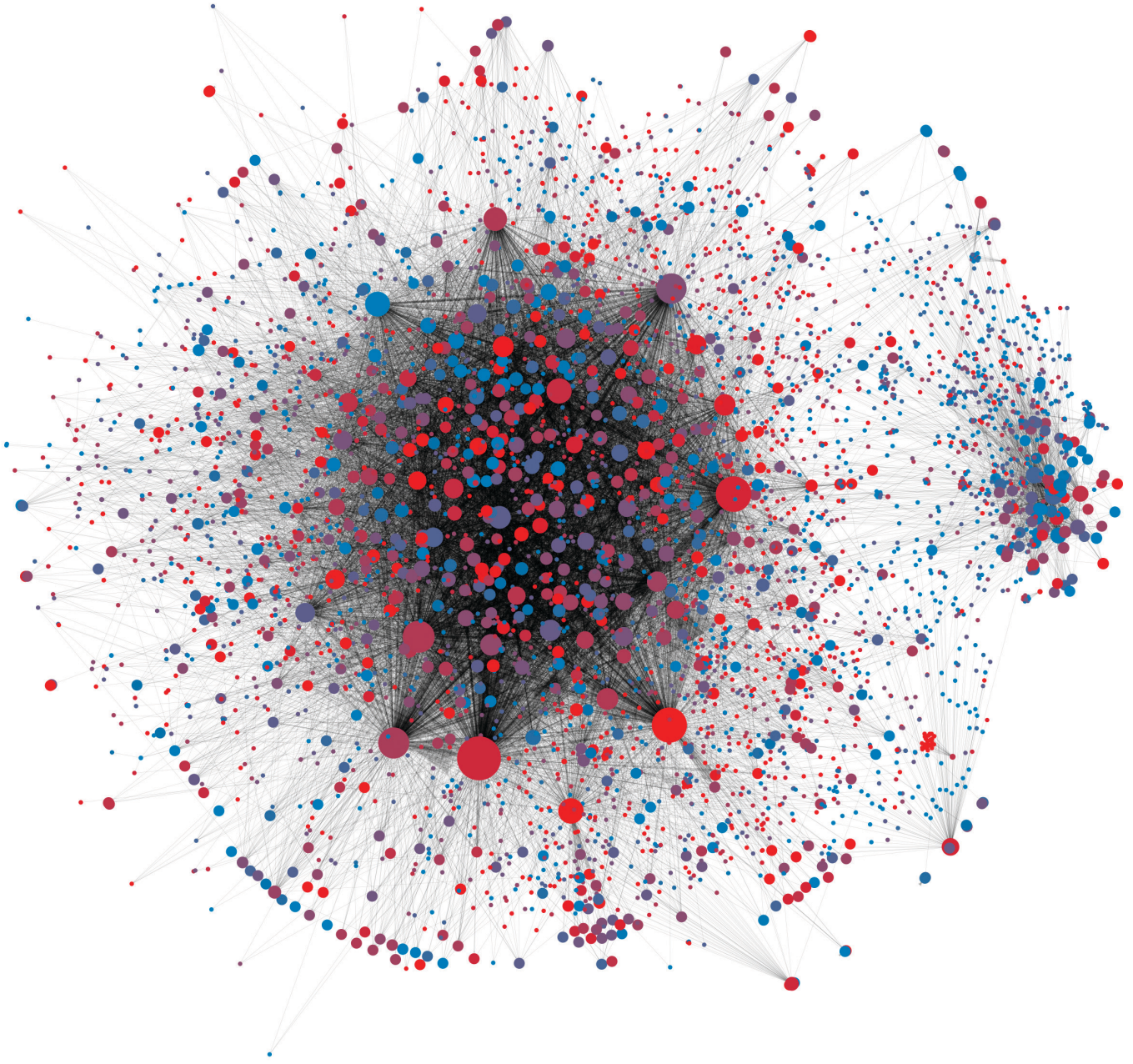
- **Social bots populate techno-social systems: they are often benign, or even useful, but some are created to harm, by tampering with, manipulating, and deceiving social media users.**
- **Social bots have been used to infiltrate political discourse, manipulate the stock market, steal personal information, and spread misinformation. The detection of social bots is therefore an important research endeavor.**
- **A taxonomy of the different social bot detection systems proposed in the literature accounts for network-based techniques, crowdsourcing strategies, feature-based supervised learning, and hybrid systems.**

ILLUSTRATION BY ARNO

aftermath of the attack, mostly due to bots automatically retweeting posts without verifying the facts or checking the credibility of the source.[20]

With every new technology comes abuse, and social media is no exception. A second category of social bots includes malicious entities designed specifically with the purpose to harm. These bots mislead, exploit, and manipulate social media discourse with rumors, spam, malware, misinformation, slander, or even just noise. This may result in several levels of damage to society. For example, bots may artificially inflate support for a political candidate;[28] such activity could endanger democracy by influencing the outcome of elections. In fact, this kind of abuse has already been observed: during the 2010 U.S. midterm elections, social bots were employed to support some candidates and smear their opponents, injecting thousands of tweets pointing to websites with fake news.[28] A similar case was reported around the Massachusetts special election of 2010.[26] Campaigns of this type are sometimes referred to as *astroturf* or Twitter bombs.

The problem is not just establishing the veracity of the information being promoted—this was an issue before the rise of social bots, and remains beyond the reach of algorithmic approaches. The novel challenge brought by bots is the fact they can give the false impression that some piece of information, regardless of



This network visualization illustrates how bots are used to affect, and possibly manipulate, the online debate about vaccination policy. It is the retweet network for the #SB277 hashtag, about a recent California law on vaccination requirements and exemptions. Nodes represent Twitter users, and links show how information spreads among users. The node size represents influence (times a user is retweeted), the color represents bot scores: red nodes are highly likely to be bot accounts, blue nodes are highly likely to be humans.

VISUALIZATION COURTESY OF INDIANA UNIVERSITY

its accuracy, is highly popular and endorsed by many, exerting an influence against which we haven't yet developed antibodies. Our vulnerability makes it possible for a bot to acquire significant influence, even unintentionally.[2] Sophisticated bots can generate personas that appear as credible followers, and thus are more difficult for both people and filtering algorithms to detect. They make for valuable entities on the fake follower market, and allegations of acquisition of fake followers have touched several prominent political figures in the U.S. and worldwide.

Journalists, analysts, and researchers increasingly report more examples of the potential dangers brought by social bots. These include the unwarranted consequences that the widespread diffusion of bots may have on the stability of markets. There have been claims that Twitter signals can be leveraged to predict the stock market,[5] and there is an increasing amount of evidence showing that market operators pay attention and react promptly to information from social media. On April 23, 2013, for example, the Syrian Electronic Army hacked the Twitter account of the Associated Press and posted a false rumor about a terror attack on the White House in which President Obama was allegedly injured. This provoked an immediate crash in the stock market. On May 6, 2010 a flash crash occurred in the U.S. stock market, when the Dow Jones plunged over 1,000 points (about 9%) within minutes—the biggest one-day point decline in history. After a five-month-long investigation, the role of high-frequency trading bots became obvious, but it yet remains unclear whether these bots had access to information from the social Web.[22]

The combination of social bots with an increasing reliance on automatic trading systems that, at least partially, exploit information from social media, is ripe with risks. Bots can amplify the visibility of misleading information, while automatic trading systems lack fact-checking capabilities. A recent orchestrated bot campaign successfully created the appearance of a sustained discussion about a tech company called Cynk. Automatic trading algorithms picked up this conversation and started trading heavily in the company's stocks. This resulted in a 200-fold increase in market value, bringing the company's worth to $5 billion.[b] By the time analysts recognized the orchestration behind this operation and stock trading was suspended, the losses were real.

The Bot Effect

These anecdotes illustrate the consequences that tampering with the social Web may have for our increasingly interconnected society. In addition to potentially endangering democracy, causing panic during emergencies, and affecting the stock market, social bots can harm our society in even subtler ways. A recent study demonstrated the vulnerability of social media users to a social botnet designed to expose private information, like phone numbers and addresses.[7] This kind of vulnerability can be exploited by cybercrime and cause the erosion of trust in social media.[22] Bots can also hinder the advancement of public policy by creating the impression of a grassroots movement of contrarians, or contribute to the strong polarization of political discussion observed in social media.[12] They can alter the perception of social media influence, artificially enlarging the audience of some people,[14] or they can ruin the reputation of a company, for commercial or political purposes.[25] A recent study demonstrated that emotions are contagious on social media[23]: elusive bots could easily infiltrate a population of unaware humans and manipulate them to affect their perception of reality, with unpredictable results. Indirect social and economic effects of social bot activity include the alteration of social media analytics, adopted for various purposes such as TV ratings,[c] expert findings,[40] and scientific impact measurement.[d]

b The Curious Case of Cynk, an Abandoned Tech Company Now Worth $5 Billion; mashable.com/2014/07/10/cynk

c Nielsen's New Twitter TV Ratings Are a Total Scam. Here's Why; defamer.gawker.com/nielsens-new-twitter-tv-ratings-are-a-total-scam-here-1442214842

d altmetrics: a manifesto; altmetrics.org/manifesto/

Act Like a Human, Think Like a Bot

One of the greatest challenges for bot detection in social media is in understanding what modern social bots can do.[6] Early bots mainly performed one type of activity: posting content automatically. These bots were naive and easy to spot by trivial detection strategies, such as focusing on high volume of content generation. In 2011, James Caverlee's team at Texas A&M University implemented a honeypot trap that managed to detect thousands of social bots.[24] The idea was simple and effective: the team created a few Twitter accounts (bots) whose role was solely to create nonsensical tweets with gibberish content, in which no human would ever be interested. However, these accounts attracted many followers. Further inspection confirmed that the suspicious followers were indeed social bots trying to grow their social circles by blindly following random accounts.

In recent years, Twitter bots have become increasingly sophisticated, making their detection more difficult. The boundary between human-like and bot-like behavior is now fuzzier. For example, social bots can search the Web for information and media to fill their profiles, and post collected material at predetermined times, emulating the human temporal signature of content production and consumption—including circadian patterns of daily activity and temporal spikes of information generation.[19] They can even engage in more complex types of interactions, such as entertaining conversations with other people, commenting on their posts, and answering their questions.[22] Some bots specifically aim to achieve greater influence by gathering new followers and expanding their social circles; they can search the social network for popular and influential people and follow them or capture their attention by sending them inquiries, in the hope to be noticed.[2] To acquire visibility, they can infiltrate popular discussions, generating topically appropriate—and even potentially interesting— content, by identifying relevant keywords and searching online for information fitting that conversation.[17] After the

appropriate content is identified, the bots can automatically produce responses through natural language algorithms, possibly including references to media or links pointing to external resources. Other bots aim at tampering with the identities of legitimate people: some are identity thieves, adopting slight variants of real usernames, and stealing personal information such as pictures and links. Even more advanced mechanisms can be employed; some social bots are able to "clone" the behavior of legitimate users, by interacting with their friends and posting topically coherent content with similar temporal patterns.

A Taxonomy of Social Bot Detection Systems

For all the reasons outlined here, the computing community is engaging in the design of advanced methods to automatically detect social bots, or to discriminate between humans and bots. The strategies currently employed by social media services appear inadequate to contrast this phenomenon and the efforts of the academic community in this direction just started.

Here, we propose a simple taxonomy that divides the approaches proposed in literature into three classes: bot detection systems based on social network information; systems based on crowdsourcing and leveraging human intelligence; and, machine-learning methods based on the identification of highly revealing features that discriminate between bots and humans. Sometimes a hard categorization of a detection strategy into one of these three categories is difficult, since some exhibit mixed elements: we present also a section of methods that combine ideas from these three main approaches.

Graph-Based Social Bot Detection

The challenge of social bot detection has been framed by various teams in an adversarial setting.[3] One example of this framework is represented by the Facebook Immune System:[30] An adversary may control multiple social bots (often referred to as *sybils* in this context) to impersonate different identities and launch an attack or infiltration. Proposed strategies to detect sybil accounts often rely on examining the structure of a social graph. SybilRank,[9] for example, assumes that sybil accounts exhibit a small number of links to legitimate users, instead connecting mostly to other sybils, as they need a large number of social ties to appear trustworthy. This feature is exploited to identify densely interconnected groups of sybils. One common strategy is to adopt off-the-shelf community detection methods to reveal such tightly knit local communities; however, the choice of the community detection algorithm has proven to crucially affect the performance of the detection algorithms.[34] A wise attacker may counterfeit the connectivity of the controlled sybil accounts to mimic the features of the community structure of the portion of the social network populated by legitimate accounts; this strategy would make the attack invisible to methods solely relying on community detection.

The computing community is engaging in the design of advanced methods to automatically detect social bots, or to discriminate between humans and bots.

To address this shortcoming, some detection systems, for example SybilRank, also employ the paradigm of innocent by association: an account interacting with a legitimate user is considered itself legitimate. Souche[41] and Anti-Reconnaissance[27] also rely on the assumption that social network structure alone separates legitimate users from bots. Unfortunately, the effectiveness of such detection strategies is bound by the behavioral assumption that legitimate users refuse to interact with unknown accounts. This was proven unrealistic by various experiments:[7,16,31] A large-scale social bot infiltration on Facebook showed that over 20% of legitimate users accept friendship requests indiscriminately, and over 60% accept requests from accounts with at least one contact in common.[7] On other platforms like Twitter and Tumblr, connecting and interacting with strangers is one of the main features. In these circumstances, the innocent-by-association paradigm yields high false-negative rates. Some authors noted the limits of the assumption of finding groups of social bots or legitimate users only: real platforms may contain many mixed groups of legitimate users who fell prey of some bots,[3] and sophisticated bots may succeed in large-scale infiltrations making it impossible to detect them

solely from network structure information. This brought Alvisi et al.[3] to recommend a portfolio of complementary detection techniques, and the manual identification of legitimate social network users to aid in the training of supervised learning algorithms.

Crowdsourcing Social Bot Detection

Wang et al.[38] have explored the possibility of human detection, suggesting the crowdsourcing of social bot detection to legions of workers. As a proof-of-concept, they created an Online Social Turing Test platform. The authors assumed that bot detection is a simple task for humans, whose ability to evaluate conversational nuances like sarcasm or persuasive language, or to observe emerging patterns and anomalies, is yet unparalleled by machines. Using data from Facebook and Renren (a popular Chinese online social network), the authors tested the efficacy of humans, both expert annotators and workers hired online, at detecting social bot accounts simply from the information on their profiles. The authors observed the detection rate for hired workers drops off over time, although it remains good enough to be used in a majority voting protocol: the same profile is shown to multiple workers and the opinion of the majority determines the final verdict. This strategy exhibits a near-zero false positive rate, a very desirable feature for a service provider.

Three drawbacks undermine the feasibility of this approach: first, although the authors make a general claim that crowdsourcing the detection of social bots might work if implemented since the early stage, this solution might not be cost effective for a platform with a large pre-existing user base, like Facebook and Twitter. Second, to guarantee that a minimal number of human annotators can be employed to minimize costs, "expert" workers are still needed to accurately detect fake accounts, as the "average" worker does not perform well individually. As a result, to reliably build a ground-truth of annotated bots, large social network companies like Facebook and Twitter are forced to hire teams of expert analysts,[30] however such a choice might not be suitable for small social networks in their early stages (an issue at odds with the previous point). Finally, exposing personal information to external workers for validation raises privacy issue.[15] While Twitter profiles tend to be more public compared to Facebook, Twitter profiles also contain less information than Facebook or Renren, thus giving a human annotator less ground to make a judgment. Analysis by manual annotators of interactions and content produced by a Syrian social botnet active in Twitter for 35 weeks suggests that some advanced social bots may no longer aim at mimicking human behavior, but rather at misdirecting attention to irrelevant information.[1]

Such smoke screening strategies require high coordination among the bots. This observation is in line with early findings on political campaigns orchestrated by social bots, which exhibited not only peculiar network connectivity patterns but also enhanced levels of coordinated behavior.[28] The idea of leveraging information about the synchronization of account activities has been fueling many social bot detection systems: frameworks like CopyCatch,[4] SynchroTrap,[10] and the Renren Sybil detector[37,42] rely explicitly on the identification of such coordinated behavior to identify social bots.

Classes of features employed by feature-based systems for social bot detection.

Class	Description
Network	Network features capture various dimensions of information diffusion patterns. Statistical features can be extracted from networks based on retweets, mentions, and hashtag co-occurrence. Examples include degree distribution, clustering coefficient, and centrality measures.[29]
User	User features are based on Twitter meta-data related to an account, including language, geographic locations, and account creation time.
Friends	Friend features include descriptive statistics relative to an account's social contacts, such as median, moments, and entropy of the distributions of their numbers of followers, followees, and posts.
Timing	Timing features capture temporal patterns of content generation (tweets) and consumption (retweets); examples include the signal similarity to a Poisson process,[18] or the average time between two consecutive posts.
Content	Content features are based on linguistic cues computed through natural language processing, especially part-of-speech tagging; examples include the frequency of verbs, nouns, and adverbs in tweets.
Sentiment	Sentiment features are built using general-purpose and Twitter-specific sentiment analysis algorithms, including happiness, arousal-dominance-valence, and emotion scores.[5,19]

Feature-Based Social Bot Detection

The advantage of focusing on behavioral patterns is that these can be easily encoded in features and adopted with machine learning techniques to learn the signature of human-like and bot-like behaviors. This allows for classifying accounts later according to their observed behaviors. Different classes of features are commonly employed to capture orthogonal dimensions of users' behaviors, as summarized in the accompanying table.

One example of a feature-based system is represented by Bot or Not?. Released in 2014, it was the first social bot detection interface for Twitter to be made publicly available to raise awareness about the presence of social bots.[13,e] Similarly to other feature-based systems,[29] Bot or Not? implements a detection algorithm relying upon highly predictive features that capture a variety of suspicious behaviors and well separate social bots from humans. The system employs off-the-shelf supervised learning algorithms trained with examples of both humans and bots behaviors, based on the Texas A&M dataset[24] that contains 15,000 examples of each class and millions of tweets. Bot or Not? scores a detection accuracy above 95%,[f] measured by AU-

e As of the time of this writing, Bot or Not? remains the only social bot detection system with a public-facing interface: http://truthy.indiana.edu/botornot

f Detecting more recent and sophisticated social bots, compared to those in the 2011 dataset, may well yield lower accuracy.

Figure 1. Common features used for social bot detection. (a) The network of hashtags co-occurring in the tweets of a given user. (b) Various sentiment signals including emoticon, happiness and arousal-dominance-valence scores. (c) The volume of content produced and consumed (tweeting and retweeting) over time.

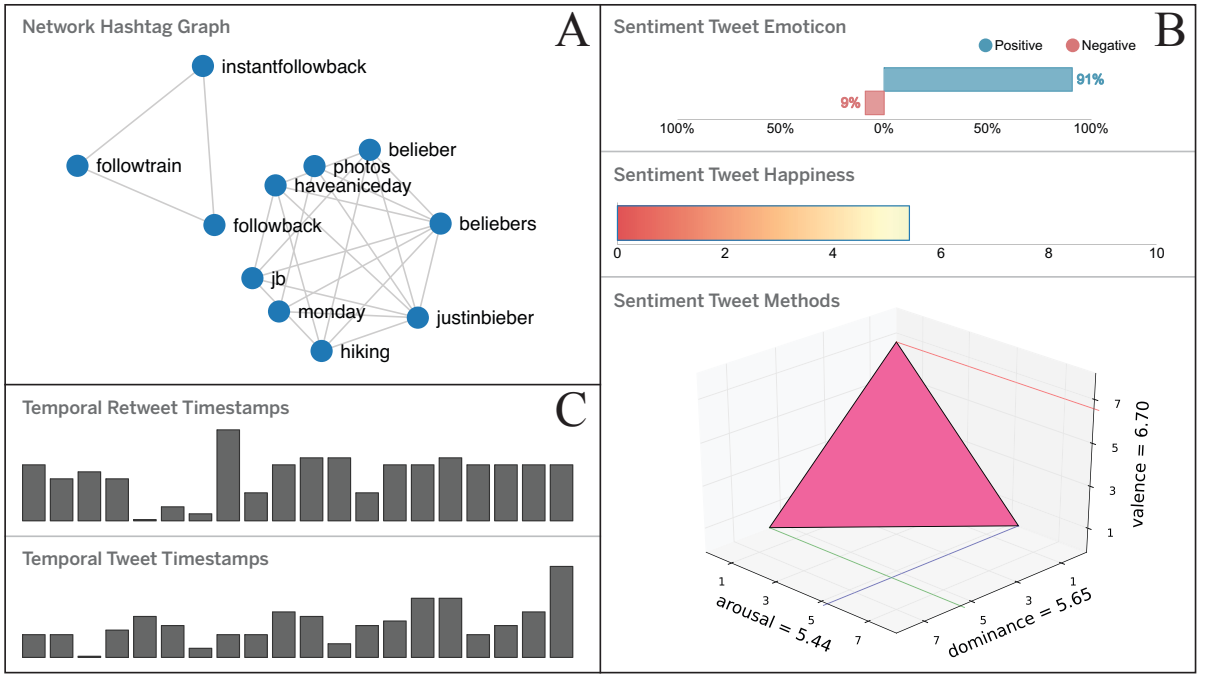
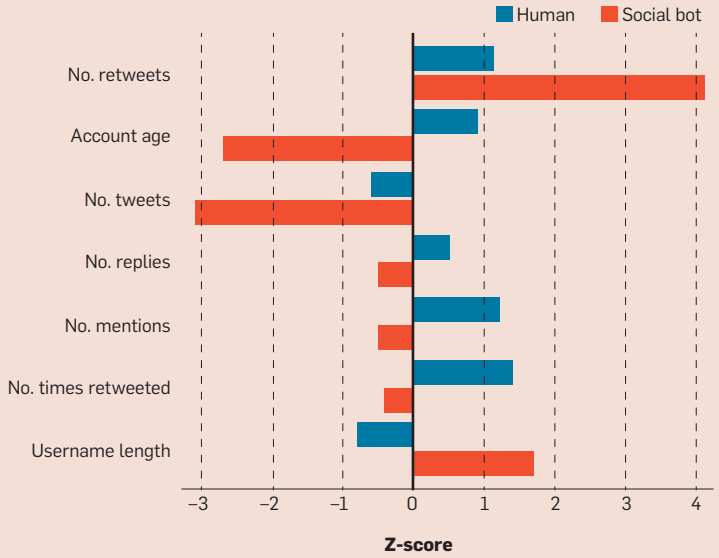


Figure 2. User behaviors that best discriminate social bots from humans.

Social bots retweet more than humans and have longer user names, while they produce fewer tweets, replies and mentions, and they are retweeted less than humans. Bot accounts also tend to be more recent.



ROC via cross validation. In addition to the classification results, Bot or Not? features a variety of interactive visualizations that provide insights on the features exploited by the system (see Figure 1 for examples).

Bots are continuously changing and evolving: the analysis of the highly predictive behaviors that feature-based systems can detect may reveal interesting patterns and provide unique opportunities to understand how to discriminate between bots and humans. User meta-data is considered among the most predictive feature and the most interpretable ones.[22,38] We can suggest a few rules of thumb to infer whether an account is likely a bot, by comparing its meta-data with that of legitimate users (see Figure 2). Further work, however, will be needed to detect sophisticated strategies exhibiting a mixture of humans and social bots features (sometimes referred to as cyborgs). Detecting these bots, or hacked accounts,[43] is currently impossible for feature-based systems.

Combining Multiple Approaches

Alvisi et al.[3] recognized first the need of adopting complementary detection techniques to effectively deal with sybil attacks in social networks. The Renren Sybil detector[37,42] is an example of system that explores multiple dimensions of users' behaviors like activity and timing information. Examination of ground-truth click-stream data shows that real users spend comparatively more time messaging and looking at other users' contents (such as photos and videos),

whereas Sybil accounts spend their time harvesting profiles and befriending other accounts. Intuitively, social bot activities tend to be simpler in terms of variety of behavior exhibited. By also identifying highly predictive features such as invitation frequency, outgoing requests accepted, and network clustering coefficient, Renren is able to classify accounts into two categories: bot-like and human-like prototypical profiles.[42] Sybil accounts on Renren tend to collude and work together to spread similar content: this additional signal, encoded as content and temporal similarity, is used to detect colluding accounts. In some ways, the Renren approach[37,42] combines the best of network- and behavior-based conceptualizations of Sybil detection. By achieving good results even utilizing only the last 100 click events for each user, the Renren system obviates to the need to store and analyze the entire click history for every user. Once the parameters are tweaked against ground truth, the algorithm can be seeded with a fixed number of known legitimate accounts and then used for mostly unsupervised classification. The "Sybil until proven otherwise" approach (the opposite of the innocent-by-association strategy) baked into this framework does lend itself to detecting previously unknown methods of attack: the authors recount the case of spambots embedding text in images to evade detection by content analysis and URL blacklists. Other systems implementing mixed methods, like CopyCatch[4] and SynchroTrap,[10] also score comparatively low false positive rates with respect to, for example, network-based methods.

Master of Puppets

If social bots are the puppets, additional efforts will have to be directed at finding their "masters." Governments[g] and other entities with sufficient resources[h] have been alleged to use social bots to their advantage.

If social bots are the puppets, additional efforts will have to be directed at finding their "masters."

Assuming the availability of effective detection technologies, it will be crucial to reverse engineer the observed social bot strategies: who they target, how they generate content, when they take action, and what topics they talk about. A systematic extrapolation of such information may enable identification of the puppet masters.

Efforts in the direction of studying platforms vulnerability have already started. Some researchers,[17] for example, reverse-engineer social bots reporting alarming results: simple automated mechanisms that produce contents and boost followers yield successful infiltration strategies and increase the social influence of the bots. Other teams are creating bots themselves: Tim Hwang's[22] and Sune Lehmann's[i] groups continuously challenge our understanding of what strategies effective bots employ, and help quantify the susceptibility of people to their influence.[35,36] Briscoe et al.[8] studied the deceptive cues of language employed by influence bots. Tools like Bot or Not? have been made available to the public to shed light on the presence of social bots online.

Yet many research questions remain open. For example, nobody knows exactly how many social bots populate social media, or what share of content can be attributed to bots—estimates vary wildly and we might have observed only the tip of the iceberg. These are important questions for the research community to pursue, and initiatives such as DARPA's SMISC bot detection challenge, which took place in the spring of 2015, can be effective catalysts of this emerging area of inquiry.[32]

Bot behaviors are already quite sophisticated: they can build realistic social networks and produce credible content with human-like temporal patterns. As we build better detection systems, we expect an arms race similar to that observed for spam in the past.[21] The need for training instances is an intrinsic limitation of supervised learning in such a scenario; machine learning techniques such as active learning might help respond to newer threats. The race will be over

g Russian Twitter political protests 'swamped by spam'; www.bbc.com/news/technology-16108876

h Fake Twitter accounts used to promote tar sands pipeline; www.theguardian.com/environment/2011/aug/05/fake-twitter-tar-sands-pipeline

i You are here because of a robot; sunelehmann.com/2013/12/04/youre-here-because-of-a-robot/

only when the effectiveness of early detection will sufficiently increase the cost of deception.

The future of social media ecosystems might already point in the direction of environments where machine-machine interaction is the norm, and humans navigate a world populated mostly by bots. We believe there is a need for bots and humans to be able to recognize each other, to avoid bizarre, or even dangerous, situations based on false assumptions of human interlocutors.[j]

Acknowledgments

The authors are grateful to Qiaozhu Mei, Zhe Zhao, Mohsen JafariAsbagh, Prashant Shiralkar, and Aram Galstyan for helpful discussions.

This work is supported in part by the Office of Naval Research (grant N15A-020-0053), National Science Foundation (grant CCF-1101743), DARPA (grant W911NF-12-1-0037), and the James McDonnell Foundation (grant 220020274). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. C

j That Time 2 Bots Were Talking, and Bank of America Butted In; www.theatlantic.com/technology/

References

1. Abokhodair, N., Yoo, D. and McDonald, D.W. 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* (2015). ACM.
2. Aiello, L.M., Deplano, M., Schifanella, R. and Ruffo, G. People are strange when you're a stranger: Impact and influence of bots on social networks. In *Proceedings of the 6th AAAI International Conference on Weblogs and Social Media* (2012). AAAI, 10–17.
3. Alvisi, L., Clement, A., Epasto, A., Lattanzi, S. and Panconesi, A. Sok: The evolution of sybil defense via social networks. In *Proceedings of the 2013 IEEE Symposium on Security and Privacy*. IEEE, 382–396.
4. Beutel, A., Xu, W., Guruswami,V., Palow, C. and Faloutsos, C. Copy-Catch: stopping group attacks by spotting lockstep behavior in social networks. In *Proceedings of the 22nd International Conference on World Wide Web* (2013), 119–130.
5. Bollen, J., Mao, H. and Zeng, X. Twitter mood predicts the stock market. *J. Computational Science 2*, 1 (2011), 1–8.
6. Boshmaf, Y., Muslukhov, I., Beznosov, K. and Ripeanu, M. Key challenges in defending against malicious socialbots. In *Proceedings of the 5th USENIX Conference on Large-scale Exploits and Emergent Threats, Vol. 12* (2012).
7. Boshmaf, Y., Muslukhov, I., Beznosov, K. and Ripeanu, M. 2013. Design and analysis of a social botnet. *Computer Networks 57*, 2 (2013), 556–578.
8. Briscoe, E.J., Appling, D.S. and Hayes, H. Cues to deception in social media communications. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (2014). IEEE, 1435–1443.
9. Cao, Q., Sirivianos, M., Yang, X. and Pregueiro, T. Aiding the detection of fake accounts in large scale social online services. *NSDI* (2012). 197–210.
10. Cao, Q., Yang, X., Yu, J. and Palow, C. Uncovering large groups of active malicious accounts in online social networks. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 477–488.
11. Cassa, C.A., Chunara, R., Mandl, K. and Brownstein, J.S. Twitter as a sentinel in emergency situations: Lessons from the Boston marathon explosions. *PLoS Currents: Disasters* (July 2013); http://dx.doi.org/10.1371/currents.dis.ad70cd1c8bc585e9470046cde334ee4b
12. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F. and Flammini, A. Political polarization on Twitter. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (2011), 89–96.
13. Davis, C.A., Varol, O., Ferrara, E., Flammini, A. and Menczer, F. BotOrNot: A system to evaluate social bots. In *Proceedings of the 25th International World Wide Web Conference Companion* (2016); http://dx.doi.org/10.1145/2872518.2889302 Forthcoming. Preprint arXiv:1602.00975.
14. Edwards, C., Edwards, A., Spence, P.R. and Shelton, A.K. Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior 33* (2014), 372–376.
15. Elovici, Y., Fire, M., Herzberg, A. and Shulman, H. Ethical considerations when employing fake identities in online social networks for research. *Science and Engineering Ethics* (2013), 1–17.
16. Elyashar, A., Fire, M., Kagan, D. and Elovici, Y. Homing socialbots: Intrusion on a specific organization's employee using Socialbots. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM, 1358–1365.
17. Freitas, C.A. et al. Reverse engineering socialbot infiltration strategies in Twitter. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM 2015.
18. Ghosh, R., Surachawala, T. and Lerman, K. Entropy-based classification of "retweeting" activity on Twitter. In *Proceedings of the KDD Workshop on Social Network Analysis* (2011).
19. Golder, S.A. and Macy, M.W. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science 333*, 6051 (2011), 1878–1881.
20. Gupta, A., Lamba, H. and Kumaraguru, P. $1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on Twitter. eCrime Researchers Summit. IEEE (2013), 1–12.
21. Heymann, P., Koutrika, G. and Garcia-Molina, H. Fighting spam on social web sites: A survey of approaches and future challenges. *Internet Computing 11*, 6 (2007). IEEE, 36–45.
22. Hwang, T., Pearce, I. and Nanis, M. Socialbots: Voices from the fronts. *ACM Interactions 19*, 2 (2012), 38–45.
23. Kramer, A.D. Guillory, J.E. and Hancock, J.T. Experimental evidence of massive-scale emotional contagion through social networks. In *Proceedings of the National Academy of Sciences* (2014), 201320040.
24. Lee, K., Eoff, B.D., and Caverlee, J. Seven months with the devils: A long-term study of content polluters on Twitter. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (2011), 185–192.
25. Messias, J., Schmidt, L., Oliveira, R. and Benevenuto, F. You followed my bot! Transforming robots into influential users in Twitter. *First Monday 18*, 7 (2013).
26. Metaxas, P.T. and Mustafaraj, E. Social media and the elections. *Science 338*, 6106 (2012), 472–473.
27. Paradise, A., Puzis, R. and Shabtai, A. Anti-reconnaissance tools: Detecting targeted socialbots. *Internet Computing 18*, 5 (2014), 11–19.
28. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A. and Menczer, F. Detecting and tracking political abuse in social media. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (2011). 297–304.
29. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A. and Menczer, F. Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th International Conference on the World Wide Web* (2011), 249–252.
30. Stein, T., Chen, E. and Mangla, K. Facebook immune system. In *Proceedings of the 4th Workshop on Social Network Systems* (2011). ACM, 8.
31. Stringhini, G., Kruegel, C. and Vigna, G. Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference* (2010), ACM, 1–9.
32. Subrahmanian, VS, Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A., Menczer, F. and others. The DARPA Twitter Bot Challenge. *IEEE Computer* (2016). In press. Preprint arXiv:1601.05140.
33. Turing, A.M. Computing machinery and intelligence. *Mind 49*, 236 (1950), 433–460.
34. Viswanath, B., Post, A., Gummadi, K.P. and Mislove, A. An analysis of social network-based sybil defenses. *ACM SIGCOMM Computer Communication Review 41*, 4 (2011), 363–374.
35. Wagner, C., Mitter, S., Körner, S. and Strohmaier, M. When social bots attack: Modeling susceptibility of users in online social networks. In *Proceedings of the 21st International Conference on World Wide Web* (2012), 41–48.
36. Wald, R., Khoshgoftaar, T.M., Napolitano, A. and Sumner, C. Predicting susceptibility to social bots on Twitter. In *Proceedings of the 14th IEEE International Conference on Information Reuse and Integration*. IEEE, 6–13.
37. Wang, G., Konolige, T., Wilson, C., Wang, X., Zheng, H. and Zhao, B.Y. You are how you click: Clickstream analysis for sybil detection. *USENIX Security* (2013), 241–256.
38. Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M., Zheng, H. and Zhao, B.Y. Social turing tests: Crowdsourcing sybil detection. NDSS. The Internet Society, 2013.
39. Weizenbaum, J. ELIZA—A computer program for the study of natural language communication between man and machine. *Commun. ACM 9*, 1 (Sept. 1966), 36–45.
40. Wu, X., Feng, Z., Fan, W., Gao, J. and Yu, Y. Detecting marionette microblog users for improved information credibility. *Machine Learning and Knowledge Discovery in Databases*. Springer, 2013, 483–498.
41. Xie, Y., Yu, F., Ke, Q., Abadi, M., Gillum, E., Vitaldevaria, K., Walter, J., Huang, J. and Mao, Z.M. Innocent by association: Early recognition of legitimate users. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*. ACM, 353–364.
42. Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B.Y. and Dai, Y. 2014. Uncovering social network sybils in the wild. ACM Trans. Knowledge Discovery from Data 8, 1 (2014), 2.
43. Zangerle, E. and Specht, G. 'Sorry, I was hacked' A classification of compromised Twitter accounts. In *Proceedings of the 29th Symposium On Applied Computing* (2014).

Emilio Ferrara (emiliofe@usc.edu) is a research assistant professor at the University of Southern California, Los Angeles, and a computer scientist at the USC Information Sciences Institute. He was a postdoctoral fellow at Indiana University when this work was carried out.

Onur Varol (ovarol@indiana.edu) is a Ph.D. candidate at Indiana University, Bloomington, IN.

Clayton Davis (claydavi@indiana.edu) is a Ph.D. candidate at Indiana University, Bloomington, IN.

Filippo Menczer (fil@indiana.edu) is a professor of computer science and informatics at Indiana University, Bloomington, IN.

Alessandro Flammini (aflammin@indiana.edu) is an associate professor of informatics at Indiana University, Bloomington, IN.

Watch the authors discuss their work in this exclusive *Communications* video. http://cacm.acm.org/videos/the-rise-of-social-bots