A Review on Social Bot Detection Techniques and Research Directions

Arzum Karataş
Department of Computer Engineering
Izmir Institute of Technology
Izmir, Turkey
arzumkaratas@iyte.edu.tr

Serap Şahin
Department of Computer Engineering
Izmir Institute of Technology
Izmir, Turkey
serapsahin@iyte.edu.tr

Abstract— **The rise of web services and popularity of online social networks (OSN) like Facebook, Twitter, LinkedIn etc. have led to the rise of unwelcome social bots as automated social actors. Those actors can play many malicious roles including infiltrators of human conversations, scammers, impersonators, misinformation disseminators, stock market manipulators, astroturfers, and any content polluter (spammers, malware spreaders) and so on. It is undeniable that social bots have major importance on social networks. Therefore, this paper reveals the potential hazards of malicious social bots, reviews the detection techniques within a methodological categorization and proposes avenues for future research.**

Index Terms— **Social bots, OSN, Sybils, social bot detection.**

I. Introduction

Our world has been dominated by online social networks (OSN) like Facebook, Twitter, and LinkedIn and so on. They play a pivotal role in our lives as public communication channels. They provide a platform for their users to involve, interact, and share information. Therefore, they lead a great community with the value of attracting for advertisements. Due to the popularity and rich API of OSNs, they are attractive targets for exploitations of social bots [1] as well.

A social bot is software to automate user activities. These activities can be (i) generating pseudo posts which look like human generated to interact with humans on a social network, (ii) reposting post, photographs or status of the others, and (iii) adding comments or likes to posts, (iv) building connections with other accounts. Therefore, the level of the sophistication of the bots is diverge. A social bot [2, 3] could be dummy like bots aggregating information from news, weather news, blog posts and then reposts them in the social network. On the other hand, they also can be extremely sophisticated such as infiltrating human conversations. These capabilities have pros and cons for users of OSN and they can be used for good or bad intentions.

(i). One hand, bots can be designed for good intentions. They can use to protect anonymity of members as mentioned in related work or automate and perform tasks much faster than humans, like automatically pushing news, weather updates or adding a template in Wikipedia to all pages in a specific category[4], or sending a thank-you message to your new followers out of courtesy. They can be designed to be helpful like virtual assistants for individuals such as Siri[1] or serving a user-friendly customer service [5] for the companies and chatbots like Microsoft's Tay[6] artificial intelligence bot.

(ii). On the other hand, social bots can be designed for doing malicious activities such as spamming, malware dissemination, impersonation, Sybil attack launching and so on.

- One of the malicious functionality of social bots is the power of dissemination of misinformation. For example, Syrian Electronic Army hacks the Twitter account of Associated Press and announces the White House is under attack and Obama is injured. This fake news lead to a panic and huge loss in the stock market in 2013 [7].
- Another malicious functionality of social bots is that they are convenient way of propaganda. This malice activity is so-called *astroturfing* — an attempt to create a fake impression on real grassroots to support a policy, individual, product campaign [8]. Concerning this, Ratkiewicz et al.'s study [9] dissects how Twitter can be exploited by astroturfing campaigns during the 2010 U.S. midterm elections. According to Boshmaf et al.[10], as democratic communication platforms, OSN are one of the key enablers of the recent Arab Spring in the Middle East in 2011. Additionally, there is a concern whether automated propaganda sway or not 2016 elections between Trump and Clinton [11]. According to these possibilities, we can assume that social bots can be very powerful tools to fire social revolutions.
- Another obstacle is that the bots can be leveraged for getting fake rating and reviews. For example, there are influence bots that serve this purpose. Also, it is possible to find many web pages that serve fake

[1] http://www.apple.com/ios/siri/

followers and likes even for free by simply searching on any search engine. Subrahmanian et al. [3] state some politicians have been accused of buying influence on social media.

- In addition, a social bot can be malicious by impersonating actual person or an organization, i.e. identity fraud. One of the evil purpose of impersonation is to serve promoting ideologies. Via this promotion, attackers have a power to mislead the individuals on the networks or create real-looking fake identities. Next, they are able to use them in malicious activities such as follower fraud [3, 13] or Sybil attacks consisting of large-scale bot armies(botnets) with simple OSN accounts[12].

Hence, the main question is focused on “How we separately detect malicious activities on OSN”. Many techniques are proposed to detect social bots on OSN in the literature. We review these techniques within a methodological categorization and unveil possible research avenues for each category for the social bot detection. For this purpose, Section II is reserved for the literature review on the detection techniques. Then, open problems for the social bot detection techniques are presented to envision and motivate possible researchers in Section III. Finally, the work is concluded with a small discussion on current research directions in Section IV.

II. Related Work

For all reasons outlined above (malicious usages of social bots), computing community has been developing advanced techniques to detect social bots accurately. Broadly, it is possible to classify these detection techniques into three classes: (A) bot detection systems based on social network topology (i.e. structure-based) information, (B) systems based on crowdsourcing on user posts and profile analysis, and (C) systems based on feature-based machine learning methods.

A. Structure-Based (Social Network-Based) Bot Detection

Sybil accounts are the multiple accounts controlled by an adversary. The naming of “Sybil” term is coming from the subject of the book Sybil (a woman diagnosed with dissociative identity disorder [14]). Structure-based detection techniques focus on detecting Sybil accounts. These accounts are used to infiltrate OSN, steal private data, disseminate misinformation and malware. That’s why, Sybil attacks are fundamental threat for social networks [15-17] . For instance, it was reported in 2015 that around 170 million fake Facebook accounts are detected as Sybil accounts, then they are deleted [18]. Whereas Sybils can be generated intentionally by users for benign purposes such as preserving anonymity; we consider solely malicious ones as Sybils from this point.

Knowing how Sybil accounts spread on the network is crucial to detect them especially for this type of detection techniques. Fundamental assumption underlying the structure-based Sybil detection is that the social networks generally shows a *homophily tendency [17]*. That is, two connected accounts in OSN have a tendency of having similar attributes. Therefore, this assumption grounds the intuition in Fig. 1. Here the honest and Sybil regions of graph are sparsely connected and Sybils have small number of connections to legitimate (honest) users. By large connections the Sybil communities create a fake trustworthy impression on honest members of the OSN. It may be useful to note that trusted nodes in Fig. 1 are already honest and they are specified at initialization as reference members.

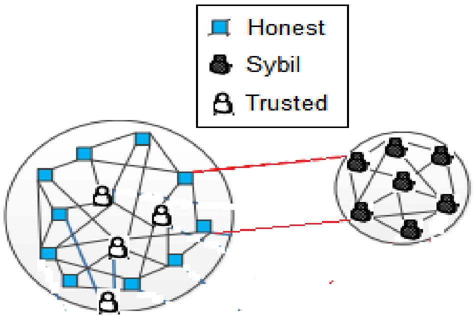

Figure 1. The social network with honest, trusted and Sybil nodes

There are many works to solve Sybil detection problem by using topology (structure) of the network. The analysis of network topology is a way for the detection of local communities. Let’s summarize their works with respect to the methods that they employ:

- Random Walk [19]

Generally, the intuition behind leveraging random walks is that social networks are fast mixing that helps to recognize Sybils from honest accounts. Fast mixing in this context implies that short random walks starting from an honest account quickly reach other honest accounts, whereas it is hard for random walks starting from Sybils to reach the honest accounts [20]. At a high level, it can be said that the works that employ random walks label the nodes as Sybil or honest in the network from the perspective of a trusted node.

As one of the random walk-based method SybilInfer [21] , uses a combination of Bayesian inference and Monte-Carlo sampling techniques to estimate the set of honest and Sybil users. It detects a bottleneck cut between honest and Sybil regions. SybilGuard [22] adopts the assumption that malicious user can create many Sybils, but the Sybils can have few connections to honest accounts like in Fig. 1. That is, the number and sizes are bounded of the honest account. Similarly, SybilLimit [23] attempts the isolate Sybils based on random walks. It adopts the same insight with SybilGuard but offers improved and near-optimality guarantees. SybilRank [24] ranks the accounts according to their perceived likelihood (landing probability of short random walks) of being Sybil. Because, there is limited probability of escaping to Sybil region for a short random walk starting from a trusted node.

- Markov Random Field[25] and Loopy Belief Propagation [26] .

The assumption that social networks are fast-mixing presumes one big community or cluster to be valid. However, Mohaisen et al. [27] show that OSN are not fast-mixing generally. Similarly, Leskovec et al. [28] demonstrate that OSN have many small periphery communities that do form small communities instead of constructing one big cluster (community). Therefore, Viswanath et al. [29] state that the Sybil detection problem can be regarded as a community detection problem. Besides, Boshmaf et al. [30] point out that structure-based Sybil detection algorithms should be designed to find local community structures around known honest (non-Sybil) identities, while incrementally tracking changes in the network by adding or deleting some nodes and edges dynamically in some period for better detection performance.

Additionally, Viswanath et al. [29] discover that dependency on community detection makes more vulnerable to Sybil attacks where honest identities conform strong communities. Because Sybils can infiltrate honest communities by carefully targeting honest accounts. That is, Sybils can be hidden as just another community on OSN by setting up a small number of the targeted links. The targeted links are the links given to the community which contains the trusted node. They make an experiment by allowing Sybils to place their links closer to the trusted node instead of random nodes, where closeness is defined by ranking used by the community detection algorithm they employ. Hence, Sybil nodes are high ranked in the defence scheme. Naturally, it leads to Sybils being less likely to be detected for that attack model because Sybils are appeared as part of the local community of the trusted node.

Due to the limitations on the fast-mixing assumption, other studies are done to handle. SybilBelief [17] and SybilFrame [15] do not use random walks, instead they rely on the Markov Random Fields and Loopy Belief Propagation to estimate probabilities of users being honest. While SybilBelief can incorporate information about known honest and known Sybil nodes, SybilFrame uses a multi-stage classification mechanism using local information of users and edges with global graph structure. In this category, SybilFrame shows the best social bot detection rate with maximum 68.2% [15].

Additionally, some structure-based Sybil detection systems like SybilRank also employ “innocent by association” paradigm [31]: if an identity has an interaction with an innocent identity, then itself is innocent as well. This is a vulnerable approach for a smart attacker mimics the structure of legitimate community. The effectiveness of this paradigm is limited by the refusal of innocent users to interact with unknown identities as in the case of LinkedIn. Nevertheless, some real-world social networks like Twitter and Renren (largest OSN in China) do not represent strong trust network. Therefore, the detection schemes employed the paradigm produce high false-negative rate.

B. *CrowdSourcing-Based Bot Detection*

The success of structure-based Sybil detection schemes has decreased over time whereas Sybils exploit the vulnerability by which Viswanath et al. reveal (dependency on community detection vulnerability), which is stated above. For example, Jiang et al. [32] show that Sybils occasionally connect to other Sybils. Instead, they target to infiltrate communities of trusted users [33].

Wang et al. [34] proposed a new approach of applying human effort (crowdsourcing) like Amazon’s Mechanical Turk [35] to label accounts. Their insight is that careful users can detect even slight inconsistencies in account profiles and posts. They propose a two-layered system containing filtering and crowdsourcing layer. They offer to use prior automation techniques such as community detection and network-based feature selection, and user reports in filtering layer to obtain suspicious profiles. Then, they apply crowdsourcing for final decision on classifying accounts either legitimate or Sybil. According to the authors, their strategy exhibits false positive and negative rates both below %1 for their simulated system that contains 2000 profiles combination of 1000 legitimate and 1000 Sybil profiles.

There are three fundamental issues related to leverage of this strategy. First, privacy of the users should be considered and personal information of the OSN users should be hidden before sharing the information with the crowd. Second, large OSN companies need to hire expert analysts additionally and small companies cannot afford it. Third, the strategy is not easy to implement for large OSN because of the existence of huge number of members. Crowd may need too much time during decision process when labelling the accounts either bot or human.

C. *Machine Learning-Based Bot Detection*

The more social bots are sophisticated with the rise of Artificial Intelligence (AI), the more they pose risk to even political issues. That’s why, detecting the bots on OSN become a challenge. For this reason, DARPA organized a competition and social structure-based detection techniques are found useful by none of the contestants [3]. The rise of AI leads to transcendent machine learning methods as social bot detection techniques as well. The main idea behind them is to find out key characteristics of social bots to draw the border between a human actor and a machine actor. The summary of some selected works can be found below.

Chu et al. [2] make a study on profiling human, bot, and cyborgs[2]. They observe the difference among them in terms of tweet content, tweeting behaviour, and account properties like external URL ratio. Lee et al. [36] present a study for social honeypots for profiling and filtering of content polluters in social media by using their profile features. Yang et al. [33] collect Sybil accounts from Renren as ground-truth data set. Then, they analyse it by using network-based and structured-based features such as network clustering coefficient, incoming and outgoing request rate.

SentiBot [37] is a framework for addressing the classification of human versus social bots. It relies on tweet syntax like average number of hashtags, semantics like average topic sentiment, user behaviour like tweet spread, and network-centric user features like in-degree. The authors of it regard the

[2] Cyborg accounts: human assisted bots or bot assisted human accounts

number of sentiment related features as key to the identification of the bots. Therefore, they also employ sophisticated sentiment analysis techniques.

"Bot or Not?" [38] is the first social bot detection framework publicly available for Twitter. Its first release is published in 2014 which is similar to other feature based detection systems. However, it analyses more than 1000 features and grouped them into 6 classes: network, user, friends, temporal, content, and sentiment. The authors of the work implement a detection algorithm heavily depends on these core features. They state that the overall all accuracy of "Bot or Not?" is 86% for simple and sophisticated social bots in 2017 [39].

It is useful to note that machine learning and the structure information of OSN together give this detection result. The best detection rate is achieved by "Bot or Not" with 86% success rate for this category.

III. OPEN PROBLEMS

Detection of the bots on OSN are challenging issue. That's why, there are some research avenues for peculiar to each category mentioned in related work.

Social networks contain big data within itself and they dynamically grow in their nature. Structure-based detection schemes usually have high running time cost even within a static (i.e., non-real time) environment. The known best computational cost for leveraging random-walk is $O(nlogn)$. That's why, developing a computationally more efficient real-time graph algorithm for big data processing can be a good research avenue. Other issue is that the schemes give high false positive rates with relatively low accuracy, yet. For example, SybilFrame gives 4.2% false positives (FP), with a classification accuracy of 95.4%, and the social bot detection rate is maximum 68.2% as mentioned just above section. False positives are detrimental to user experience because real users can respond very negatively. That's why, a new learning approach can be employed algorithms to decrease FP and false negatives (FN) rates on the graph topology. As for community based-schemes, new approaches for determining trusted nodes on-the fly is another open area for the researchers. Since, structure of the network and trusted nodes are in the heart of the success of structure-based approaches.

Crowdsourcing-based detection schemes leverage human intelligence against sophisticated social bots equipped with AI power. Protecting user privacy is a challenge for crowdsourced detection techniques. That's why, a work can be done for increasing ethical awareness of the crowd. In addition, privacy preserving data mining techniques can be employed for user privacy. However, the schemes are neither effective nor applicable in terms of both time and money costs for the crowd when we regard that OSN are dynamic environments and that they contain big data.

Bots are continuously evolving by gaining new human-like behaviours with the rise of AI. As for feature-based machine learning schemes, some additional features can be explored employing the-state-of-art machine learning techniques like deep learning to distinguish a human from a bot. Another issue is if the Sybils are just controlled bots by an adversary, who the master is. That' is the big question: what is the source of these Sybils? That is, source detection of the Sybils is one of the big deals.

IV. CONCLUDING DISCUSSION

Social networks are powerful tools that connect the millions of people over the world. Therefore, they are attractive for social bots as well. Since the possible harm of social bots such as identity theft, astroturfing, content polluter, follower fraud, misinformation dissemination etc., there is a need of recognition of bots and humans each other to avoid undesirable situations based on false assumptions.

In Table 1, the detection techniques, related works, limitations for each techniques and contingent research areas are summarized. Since OSN are dynamic environment and contains big data itself, possible future solutions need to handle efficiently both big data processing and dynamic detection. Besides, the solutions should decrease FN and FP of the existing solutions while increasing accuracy as much as possible. Since structure-based techniques needs at least one trusted node, determining the trusted nodes on-the-fly can be a possible research direction. The best success rate for these techniques is maximum 68.2 %, which is achieved by SybilFrame.

As it is seen from the table, no research avenue is proposed since crowdsourcing-based techniques are expensive in both time and cost of the crowd workforce. As for machine learning-based techniques, the limitations of them are AI-boosted bots. However, the remedy of those limitations is advanced AI techniques like deep learning to determine the features to draw a line between innocent accounts and Sybils as well. Also, these techniques can be used to source detection of Sybils as a research direction. The best success rate for these techniques on the overall 86%, which is succeeded by "Bot or Not?".

With the progress of the social bot detection techniques, it is seen that the higher social bot detection rates (over 80%) are obtained with the combination of the structure-based properties of OSN and unsupervised machine learning methods. It is useful to conduct research on some possible approaches to increase the detection rate. The approaches may be (i) use of autonomous-intelligent agent based and (ii) identification-based approaches as the future directions of researches.

i. *Use of autonomous - intelligent agent based approaches:* For example, detection and identification of community members within the community should be performed in a decentralized environment. That is, the detection and analysis tasks are distributed to the community according to the topology of OSN. In the

Table 1. Summary of detection techniques and research directions

Detection Approaches	**Related Work**	**Accuracy**	**Limitations**	**Open Research Areas**
Structure-Based	● SybilInfer ● SybilGuard ● SybilLimit ● SybilRank ● SybilBelief ● SybilFrame	● 68.2 %	● OSN contains big data inside. ● OSN are a dynamic environment ● Need of decreasing FN, FP rates ● High running time cost ● Need of at least one trusted node	● Developing more efficient big data processing algorithms ● Dynamic or real-time detection methods ● Considering new methods to decrease FN and FP rates ● Determining trusted nodes on-the fly
Crowdsourcing-Based	● Wang et al.'s work [34]		● Limited size analysis possibility with sampling method ● Privacy issues ● High running time cost of the crowd ● Cost of crowd workforce	● Privacy can be preserved via privacy preserving data mining algorithms.
Machine Learning-Based	● Chu et al. 's work[2] ● Lee et al. 's work [36] ● Yang et al. 's work [33] ● Bor or Not? ● SentiBot	● 86 %	● OSN contains big data inside. ● OSN are a dynamic environment ● AI-powered bots ● Need of decreasing FN, FP rates ● Unknown source of Sybils	● Dynamic or real-time detection methods ● Employing popular AI techniques (like deep learning) to detect features to distinguish a bot from an innocent account holder and handle big data. ● Considering new methods to decrease FN and FP rates ● Source detection of Sybils

environment, intelligent agents aware of their community boundaries and members should be present to monitor community activities based on identified characteristics.

ii. *Identification-based approaches:* If identification component of any type of entity is not present; any technologically and methodologically developed method to solve this problem can be exploited by attackers. For example, intelligent agents developed for automatic and dynamic detection of Sybils should be supported by trusted identification mechanisms. If this does not happen, intelligent agents will be targeted and the attacker will not be detected, and innocent accounts might be declared as Sybil. This possibly result in another attack problem.

In this paper, three classes of social bot detection techniques (i.e., structure-based, crowdsourcing-based and feature-based machine learning detection techniques) on OSN, their limitations and detection rates are reviewed. After examination, it is seen that the most effective and popular one is feature-based machine learning techniques among them. However, the rise of AI for development of sophisticated bot creations, the bottlenecks of real-time big data processing and the need of source detection for a global identification system lead us to find out a novel solution. Therefore, research avenues on social bot detection techniques are reviewed, and prospective methods to be able to increase the social bot detection rates are proposed with the intention of opening the doors for researchers to exploit.

REFERENCES

[1] S. K. Dehade and A. M. Bagade, "A review on detecting automation on Twitter accounts," *Eur. J. Adv. Eng. Technol,* vol. 2, pp. 69-72, 2015.

[2] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?," *IEEE Transactions on Dependable and Secure Computing,* vol. 9, pp. 811-824, 2012.

[3] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, *et al.*, "The darpa twitter bot challenge," *arXiv preprint arXiv:1601.05140,* 2016.

[4] (2016, September 12). *Wikipedia:Creating a bot*. Available: https://en.wikipedia.org/wiki/Wikipedia:Creating_a_bot

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

[6] E. Ferrera, "The Rise of Social Bots," ed, 2016.

[7] D. Mail, "Syrian Electronic Army linked to hack attack on AP Twitter feed that 'broke news' Obama had been injured in White House blast and sent Dow Jones plunging," ed, 2013.

[8] A. Bienkov, "Astroturfing: what is it and why does it matter?," in *The Guardian*, ed, 2012.

[9] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, "Detecting and Tracking Political Abuse in Social Media," *ICWSM,* vol. 11, pp. 297-304, 2011.

[10] 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*, 2011, pp. 93-102.

[11] B. Schreckinger. (2016, September 30,2016) Inside Trump's 'cyborg' Twitter army. Available: http://www.politico.com/story/2016/09/donald-trump-twitter-army-228923

[12] Abokhodair, N., Yoo, D., & McDonald, D. W. (2015, February). Dissecting a social botnet: Growth, content and influence in Twitter. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 839-851). ACM.

[13] O. Goga, G. Venkatadri, and K. P. Gummadi, "The doppelgänger bot attack: Exploring identity impersonation in online social networks," in *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, 2015, pp. 141-153.

[14] *Sybil attack*. Available: https://en.wikipedia.org/wiki/Sybil_attack

[15] P. Gao, N. Z. Gong, S. Kulkarni, K. Thomas, and P. Mittal, "Sybilframe: A defense-in-depth framework for structure-based sybil detection," *arXiv preprint arXiv:1503.02985,* 2015.

[16] D. Mulamba, I. Ray, and I. Ray, "SybilRadar: A Graph-Structure Based Framework for Sybil Detection in On-line Social Networks," in *IFIP International Information Security and Privacy Conference*, 2016, pp. 179-193.

[17] N. Z. Gong, M. Frank, and P. Mittal, "Sybilbelief: A semi-supervised learning approach for structure-based sybil detection," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 976-987, 2014.

[18] J. Parsons. (2015, September 12). *Facebook's War Continues Against Fake Profiles and Bots*. Available: http://www.huffingtonpost.com/james-parsons/facebooks-war-continues-against-fake-profiles-and-bots_b_6914282.html

[19] K. Pearson, "The problem of the random walk," *Nature,* vol. 72, p. 294, 1905.

[20] B. Carminati, E. Ferrari, and M. Viviani, "Security and trust in online social networks," *Synthesis Lectures on Information Security, Privacy, & Trust,* vol. 4, pp. 1-120, 2013.

[21] G. Danezis and P. Mittal, "SybilInfer: Detecting Sybil Nodes using Social Networks," in *NDSS*, 2009.

[22] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, "Sybilguard: defending against sybil attacks via social networks," in *ACM SIGCOMM Computer Communication Review*, 2006, pp. 267-278.

[23] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "Sybillimit: A near-optimal social network defense against sybil attacks," in *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 2008, pp. 3-17.

[24] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro. (2016). *SybilRank.* Available: http://www.tid.es/research/areas/sybil-rank

[25] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 25-39, 1983.

[26] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, 1999, pp. 467-475.

[27] A. Mohaisen, A. Yun, and Y. Kim, "Measuring the mixing time of social graphs," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, 2010, pp. 383-389.

[28] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," *Internet Mathematics,* vol. 6, pp. 29-123, 2009.

[29] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove, "An analysis of social network-based sybil defenses," *ACM SIGCOMM Computer Communication Review,* vol. 40, pp. 363-374, 2010.

[30] Y. Boshmaf, K. Beznosov, and M. Ripeanu, "Graph-based sybil detection in social and information systems," in *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, 2013, pp. 466-473.

[31] Y. Xie, F. Yu, Q. Ke, M. Abadi, E. Gillum, K. Vitaldevaria*, et al.*, "Innocent by association: early recognition of legitimate users," in *Proceedings of the 2012 ACM conference on Computer and communications security,* 2012, pp. 353-364.

[32] J. Jiang, C. Wilson, X. Wang, W. Sha, P. Huang, Y. Dai, *et al.*, "Understanding latent interactions in online social networks," *ACM Transactions on the Web (TWEB),* vol. 7, p. 18, 2013.

[33] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai, "Uncovering social network sybils in the wild," *ACM Transactions on Knowledge Discovery from Data (TKDD),* vol. 8, p. 2, 2014.

[34] G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. Metzger, H. Zheng, *et al.*, "Social turing tests: Crowdsourcing sybil detection," *arXiv preprint arXiv:1205.3856,* 2012.

[35] (2016). *Overview of Mechanical Turk*. Available: http://docs.aws.amazon.com/AWSMechTurk/latest/RequesterUI/OverviewofMturk.html

[36] K. Lee, B. D. Eoff, and J. Caverlee, "Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter," in *ICWSM*, 2011.

[37] J. P. Dickerson, V. Kagan, and V. Subrahmanian, "Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?," in *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, 2014, pp. 620-627.

[38] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "Botornot: A system to evaluate social bots," in *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 273-274.

[39] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," *arXiv preprint arXiv:1703.03107,* 2017.