User Behavior to Identify Malicious Activities in Large-Scale Social Networks.

Miss. Punam A Bane.
Dept of Electronics & Telecommunication
KIT college of engg. Kolhapur ,India

Prof. S.S.Nagtilak
Dept of Electronics & Telecommunication
KIT college of engg. Kolhapur ,India

Abstract- The enormous growth and volume of online social networks and their features, along with the vast number of socially connected users, it has become difficult to explain the true semantic value of published content for the detection of user behaviors. Without understanding the contextual background, it is impractical to differentiate among various groups in terms of their relevance and mutual relations, or to identify the most significant representatives from the community at large. In this paper, we propose an integrated social media network content analysis platform that leverages three levels of features, i.e., user-generated content, social sensing network, and user profile activities, to analyze and detect anomalous behaviors that deviate significantly from the norm in large-scale social networks. Several types of analyses have been conducted for a better understanding of the different user behaviors in the detection of highly adaptive malicious users.

Keywords: Malicious activity, social network, user behaviors.

1 Introduction

Online Social Network activities has greatly expanded in both scope and volume, opening new opportunities for public exposure can be fully expected that this tendency will continue to accelerate, there by facilitating the possibility of a more immersive examination of social behaviors and attitudes than ever before[1]. In addition to their increasingly impressive volume, social networks consist of context-sensitive and relational data while also including a considerable amount of malicious content. Taken together, these factors are forming a completely new social field, [5] suitable for observing and classifying many fascinating phenomena With an increase in the use and benefits of online social network comes an increase in various challenges.

One of the major challenges facing such networks today is the creation of false online identities.[2] Malicious behaviors can be described in general terms as the sum of all activities conducted by a platform user that break or circumvent the official terms and conditions, usually for the purposes of material benefit of the perpetrator. This type of activity has a decidedly detrimental effect on the performance of the entire system, as well as the personal experience of individual users Malicious users are financially harmful to the OSN platform, [10]and are, therefore, being actively suppressed by all social networks Most of the previously tested methods from this group suffer from serious deficiencies. On most platforms, establishing a difference between ill-intentioned users who represent a danger to the community, [14] and inactive users who rarely interact with others, is not easy. Because intruders are keenly aware of this blind spot, they are able [8] to plant numerous bot fake profiles that cannot be immediately spotted and removed. To as certain the reliability of online personalities, we have to introduce a mechanism that helps detect and differentiate between malicious users and in frequent user.

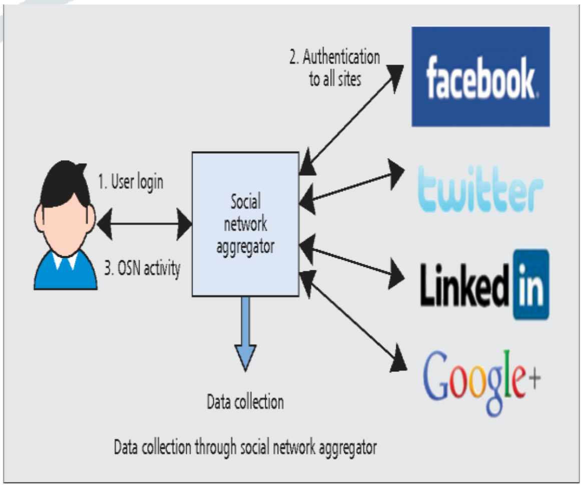

2. Authentication to all sites
facebook
1. User login
Social network aggregator
twitter
3. OSN activity
LinkedIn
Google+
Data collection
Data collection through social network aggregator

Fig 1 Data Collection through a social network aggregator.

1.1 User Behavior Datasets

The Twitter dataset consists of a random sample of 100K out of the 19M Twitter users who joined before September 2018 [4]. Previous work [4] identified topical experts in Twitter and the topics of interests of users were inferred (e.g., technology, fashion, health, etc) by analyzing the profile of topical experts followed by users. In this dataset, each expert's profile is associated with a set of topics of expertise. We construct a spatial histogram by randomly grouping multiple topics (34,334 of them) into 687 topic-groups and counting the number of experts a user is following in a given topic-group. The Twitter dataset does not have temporal features.

2. Related Work

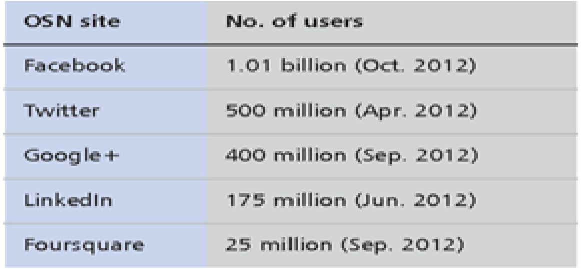
We survey approaches to detecting misbehaving identities along two axes.

2.1 Supervised Learning:

Most existing work on detecting malicious user identities in social networks leverage supervised learning techniques propose a scheme that deploys honey pots in OSNs to attract spam, trains a machine learning (ML) classifier over the captured spam, and then detects fake user using the classifier. Rahman et al. [5] propose a spam and malware detection scheme for Twitter using a Support Vector Machines-based classifier trained using the detected malicious URLs. The COMPA scheme [10] creates statistical behavioral profiles for Twitter users, trains a statistical model with a small manually labeled dataset of both benign and misbehaving users, and then uses it to detect compromised identities in Twitter.

While working with large Social networking systems, supervised learning approaches have inherent limitations. Specifically they are attack-specific and vulnerable to adaptive attacker strategies. Given the adaptability of the attacker strategies, to maintain efficiency, supervised learning approaches require labeling, training, and classification to be done periodically. In this cat-and-mouse game, they will always lag behind attackers who keep adapting to make a classification imprecise.

Table 1:Information about five popular OSNs



OSN site	No. of users
Facebook	1.01 billion (Oct. 2012)
Twitter	500 million (Apr. 2012)
Google+	400 million (Sep. 2012)
LinkedIn	175 million (Jun. 2012)
Foursquare	25 million (Sep. 2012)

2.2 Unsupervised Learning

Unsupervised learning-based anomaly detection has been found to be an effective alternative to non-adaptive supervised learning strategies [12]. For example, Li et al. [4] propose a system to detect volume anomalies in network traffic using unsupervised PCA-based methods. Auto RE [6] automatically extracts spam URL patterns in email spam based on detecting the bursty and decentralized nature of botnet traffic as anomalous. In crowd sourcing scenarios, Wang et al. [15] proposed a Sybil detection technique using server-side click stream models (based on user behavior defined by click through events generated by users during their social network browsing sessions). While the bulk of the paper presents supervised learning schemes to differentiate between Sybil and non-Sybils based on their click stream behavior, they also propose an unsupervised approach 14 that builds click stream behavioral clusters that capture normal behavior and users that are not part of normal clusters are flagged as Sybil. However, their approach still requires some constant amount of ground-truth information to figure out clusters that represent normal click-stream behavior. Tan Xn Raun [6] use a user-link graph along with the OSN graph to detect some honest users with supervised ML classifier and then perform an unsupervised analysis to detect OSN spam. Copy Catch [3] detects fraudulent likes by looking for a specific attack signature groups of users liking the same page at around the same time (lockstep behavior). Copy Catch is actively used in Twitter to detect fraudulent likes, however as evidenced in Table 2, it is not a silver bullet. While we welcome the push towards focusing more on unsupervised learning strategies for misbehavior detection, most of the current techniques are quite ad hoc and complex. Our approach using Principal Component Analysis provides a more systematic and general framework for modeling user behavior in social networks, and in fact, our PCA-based approach could leverage the user behavior features (e.g., user click-stream models [21]) used in existing work for misbehavior detection.

Classified As	Number of users
Black-market	470
Compromised	109
Colluding	345
Unclassified (no consensus)	484

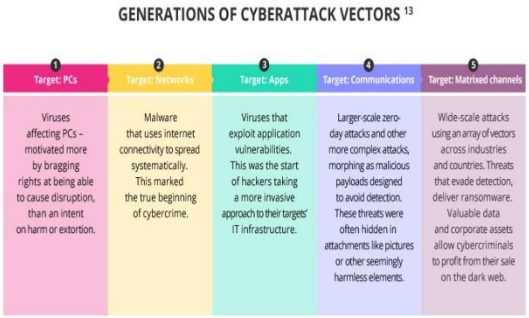
Table 2: Anomaly class predicted for the ad users that are flagged.

2.3 Prevention Mechanism

Prevention mechanisms place the core weight on protecting the user profiles in a social network. The strategy here is to make the creation of user profiles on social networks a difficult work, rather than trying to find such profiles and close them down once they have joined the social network. This strategy works well against certain types of maliceous, particularly in those scenarios in which the users may want to create many different fake profiles within a very short period of time in a bid to undertake malicious activities in a particular social network. Prevention mechanisms rely on strong verification processes that can go as far as requesting users to send copies of their identification before they can create a profile [17], [20], [5]. This type of technique is otherwise known as a Sybil prevention technique, and though quite powerful, such techniques are not commonly used given that they are not very popular among modern web users. The more popular approach among Sybil prevention mechanisms is the use of automated systems aimed at verifying whether a request has been sent by a real user or not [7]. This comes in the form of a CAPTCHA, where the user is asked to feed in a string of characters or resolve a given logical challenge [8], [15], [17],[20]. Hackers are becoming smarter, however, and are finding their way around such automated systems. What is even worse is that users can still be able to harvest a large number of bot profiles if legally registered profiles can collude together with fake profiles to undertake malicious activities. This means that Sybil prevention techniques are quite inefficient when it comes to system abuses conducted by legally registered users.

2.3 User Behavior-Based Mechanism

Solutions to this form compile and use historical data for certain users by keeping tabs on their activities over a given period of the time [3], [14], [22], [20]. Tracking is made possible given that the interactions between a profile and the elements of a online social network can be logged. In other words, users form connections with the elements as they interact in the network, e.g. through communication with other users. Researchers in this category use data and machine-learning methods to find users that do not conform to certain rules; such techniques usually compare the user activities to a predefined set of activities [1], [11], [23], [8]. In most cases, history-based algorithms need to have a set of predefined standards that describe legitimate users, hence ensuring that the systems adopt the strategy of finding Sybil profiles. Some algorithms look through all user activities to check for any malicious actions or content, such as texts, images, or videos. The algorithms then use these past online activities as baseline indicators to future traits. Such algorithms factor in several aspects of a user’s online behaviors, such as the linguistic aspects of the content, e.g., the language style [5], [21], [18], [10].All three categories used against Sybil attackers rely on the Sybil profiles depicting significant anomalous traits, which do not always exist. Sybil profiles have evolved in terms of their disruption models, meaning that there is a need for defensive methods to keep evolving along the same lines. There is also a need for algorithms that can find hijacked profiles and profiles colluding to subvert the reputation of the system. In this regard, we present and outline such a solution in the following section.



GENERATIONS OF CYBERATTACK VECTORS [13]

1 Target: PCs	2 Target: Networks	3 Target: Apps	4 Target: Communications	5 Target: Matrixed channels
Viruses affecting PCs – motivated more by bragging rights at being able to cause disruption, than an intent on harm or extortion.	Malware that uses internet connectivity to spread systematically. This marked the true beginning of cybercrime.	Viruses that exploit application vulnerabilities. This was the start of hackers taking a more invasive approach to their targets' IT infrastructure.	Larger-scale zero-day attacks and other more complex attacks, morphing as malicious payloads designed to avoid detection. These threats were often hidden in attachments like pictures or other seemingly harmless elements.	Wide-scale attacks using an array of vectors across industries and countries. Threats that evade detection, deliver ransomware. Valuable data and corporate assets allow cybercriminals to profit from their sale on the dark web.

Fig3. Generation of CyberAttack Vectors.

2 Evaluation

We now evaluate the effectiveness of our anomaly detection technique using real-world ground-truth data about normal and anomalous user behavior on Twitter. Our goal with anomaly detection in this section is to detect Twitter like spammers. Anomalous User Ground Truth We collected data for three types of anomalous behaviors: fake (Sybil) accounts that do not have any normal user activity, compromised accounts where the attacker's anomalous activity interleaves with the user's normal activity, and collusion networks where users collectively engage in undesirable behavior. We used the methods described below to collect data for over 6.8K users. We then used Selenium to crawl the publicly visible data for these users, covering 2.16M publicly-visible like sand an additional 1.19M publicly-visible Timeline posts including messages, URLs, and photos. We acquired all activity data for these users from their join date until end of August 2013.

3.1 Black-Market Services:

We searched on Google for websites offering paid Twitter likes (query:"buy Twitter likes"). We signed up with six services among the top search results and purchased the (standard) package for 1,000 likes; we paid on average $27 to each service. We created a separate Twitter page for each service to like so we could track their performance. Four of the services [18–21] delivered on their promise (3,438 total users), while the other two [22, 23] did not result in any likes despite successful payment. As mentioned, we crawled the publicly visible malicious user behavior of the black-market users who like dour pages. We discovered 1,555,535 likes (with timestamps at day granularity) by these malicious users. We further crawled the fake users publicly visible Timeline for public posts yielding an additional 89,453 Timeline posts.

3.2 Collusion Networks:

We discovered collaborative services [7, 8] where users can collaborate (or collude) to boost each other's likes. Users on these services earn virtual credits for liking Twitter pages posted by other users. Users can then encash these credits for likes on their own Twitter page. Users can also buy credits (using real money) which they can then encash for likes on their page. We obtained 2,259 likes on three Twitter pages we created, obtaining a set of 2,210 users, at an average cost of around $25 for 1,000 likes. The price for each like (in virtual credits) is set by the user requesting likes; the higher the price, the more likely it is that other users will accept the offer. We started getting likes within one minute of posting (as compared to more than a day for black-market services). As with black-market users, we crawled the user activity of the users we found through collusion networks. We collected 359,848 likes and 186,474 Timeline posts.

3.3 Compromised Accounts:

We leveraged the browser malware Febipos. A [35] that infects the user's browser and (silently) performs actions on Twitter and Twitter using the credentials/cookies stored in the browser. The malware consists of a browser plug in, written in (obfuscated) Java script, for all three major browsers: Chrome, Firefox and Internet Explorer [28, 29]. We installed the malware in a sandbox and de-obfuscated and analyzed the code. The malware periodically contacts a CnC (command-and-control) server for commands, and executes them. We identified 9 commands supported by the version of the malware we analyzed: (1) like a Twitter page, (2) add comments to a Twitter post, (3) share a wall post or photo album,(4) join a Twitter event or Twitter group, (5) post to the user's wall, (6) add comments to photos, (7) send Twitter chat messages, (8) follow a Twitter user, and (9) inject third-party ads into the user's Twitter page. We reversed engineer the application level protocol between the browser component and the CnC server, which uses HTTP as a transport. We then used curl to periodically contact the CnC to fetch the commands the CnC would have sent, logging the commands every 5 minutes. In so doing, we believe were able to monitor the entire activity of the malware for the time we measured it (August 21–30, 2018). Identifying which other Twitter users are compromised by Febipos. A requires additional data. Unlike in the black market services and collusion networks where we were able to create Twitter pages and give to the service to like we can only passively monitor the malware and cannot inject our page for the other infect edusers to like (since we do not control the CnC server). To identify other Twitter users compromised by 6 Febipos.

3.4 monitored the malware:

One which instructed the malware to like a specific Twitter page, and second, to join a specific Twitter event. We use Twitter's graph search [26] to find other users that liked the specific page and accepted the specific event directed by the CnC. From this list we sampled a total of 4,596 users. Note, however, that simply because a user matched the two filters does not necessarily mean they are compromised by Febipos. A. To improve our confidence in compromised users, we clustered the posts (based on content similarity) made to these users walls and manually inspected the top 20 most common posts. Among these 20 posts, two posts looked suspicious. Upon further investigation, we found out that one of the post was also found on pages the malware was directed to like. The other user post was present in the CnC logs we collected. The first was posted by 1,173 users while the second was posted by 135 users. We considered users from both these clusters and obtained a set of 1,193 unique users.6 We collected 247,589 like sand 916,613 Timeline posts from their profile.

4 Proposed Methodology

The proposed system architecture is realized through four separate layers, which are mutually related in a structured manner, with every module having direct communication with every other module along with an outlet to an open database. Our proposed system is based on multiple layers that facilitate simple scaling and upgrading to fit any need. The purpose, conceptual foundation, and practical application of each layer are described as follows.

A]Social Sensing Layer

B]Data Acquisition and Preparation Layer

C]Data Storage Management Layer

D] Analysis Representation Layer

A] Social Sensing Layer

Its role is to formulate and execute precise requests to the selected social system, classify the returned data, and sort the data based on their relatedness to the subject of the request with respect to the parameters that come from the request parameters handler. It also passes the response of the requests to the request-response manager, which helps manage the collected responses and extract the data

B] Data Acquisition and Preparation Layer

In this layer, we describe the steps that are involved in gathering and cleaning data as a part of the acquisition and preparation processes for analysis

C] Data Storage Management Layer

This layer is closely coordinated with the previous one, teaming up to properly utilize the information originally collected from the selected social media (e.g. Twitter) and Stored in file system

D] Analysis Representation Layer

During the procedure of social media content exploration and analysis, our proposed platform actively seeks relevant trends in any of the dimensions of the collected data that could be taken as illustrative of the general behavior on the network.

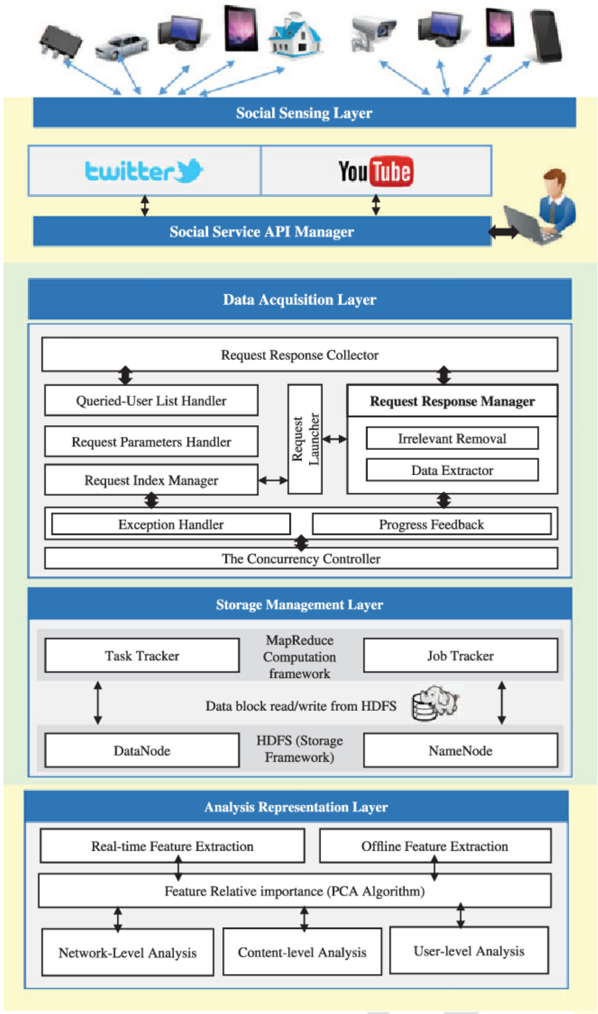


Fig 3 Layers of system Architecture.

5 Conclusion

From this research paper it's to present an integrated system with analytic abilities to detect malicious activities in OSNs. The system is expected to carefully examine and track social interactions on data consisting of textual content before reaching to an assumption of the activity by the user account detect to be real or malicious.

6 References

1] T.S. Behrend, D.J.Sharek, A.W.Meade, and E.N.Wiebe, "The viability of crowd sourcing for survey research," Behavior research methods, vol. 43, no. 3, pp.800–813, 2011.

[2] A.Kittur, E.H.Chi, and B.Suh, "Crowd sourcing user studies with mechanical turk," In Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 2008, pp. 453–456.

[3] C.C. Marshall and F.M.Shipman, "Experiences surveying the crowd: Reflections on methods, participation, and reliability", In Proceedings of the 5th Annual ACM Web Science Conference, ser. WebSci 13.New York, NY, USA: ACM, 2013, pp. 234–243..

[4] Y.Baba, H. Kashima, K. Kinoshita, G. Yamaguchi, and Y. Akiyoshi, "Leveraging crowd sourcing to detect improper tasks in crowd sourcing marketplaces," In Twenty-Fifth IAAI Conference, 2013.

[5] S. Dow, A. Kulkarni, B. Bunge, T. Nguyen, S. Klemmer, and B. Hartmann, "Shepherding the crowd: managing and providing feedback to crowd workers," In CHI'11 Extended Abstracts on Human Factors in Computing Systems. ACM, 2011, pp. 1669–1674.

[6] W.Mason and D. J.Watts, "Financial incentives and the performance of crowds," ACM SigKDD Explorations Newsletter, vol. 11, no. 2, pp. 100–108, 2010.

[7] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality management on amazon mechanical turk," in Proceedings of the ACM SIGKDD workshop on human computation. ACM, 2010, pp. 64–67.

[8] Xn Ruan, Zhenyu Wu, Member, IEEE, Haining Wang, Senior Member, IEEE, and Sushil Jajodia, Fellow, IEEE."Profiling Online Social Behaviors for Compromised Account Detection"JANUARY 2016.

[9] Michael Fire, Member,IEEE, Roy Goldschmidt, and Yuval Elovici, Member, IEEE "Online Social Networks : Threats and Solutions" 2014

[10] Bandar Alghamdi, Jason Watson, Yue Xu Faculty of science and Engineering "Toward detecting Malicious Links in Online Social Networks through User Behavior" 2016

[11] Stefanco Cresci, Member, IEEE, Roberto Di pietro, Marinella Petrocchi, Angelo Spognardi,and Maurizio Tesconi "Social Fingerprinting: Detection of Spambot GroupsnThrough DNA-Inspired Behavioral Modeling" August 2018.

[12] Amira Soliman, Sarunas Girdzijauskas "DLSAS: Distributed Large-Scale Anti-Spam Framework For Decentralized Online Social Networks" 2016

[13] Qiang Cao, Xiaowei Yang "Uncovering Large Groups of Active malicious Accounts in Online Social Networks" 2014.

[14] George W. Kibirige, "Big data Analysis on Multiple Social Network"2017

[15] Francesco Buccafurri, Gianluca Lax, Serena Nicolazzo, Antonino Nocera "Comparing twitter and Facebook User behavior: Privacy and other" 2016

[16] Muhammad Al-Qurishi, , M. Shamim Hossain ,Majed Alrubaian, Sk Md Mizanur Rahman, and Atif Alamri, "Leveraging Analysis of User Behavior to Identify Malicious Activities in Large-Scale Social Networks" ieee transactions on industrial informatics, vol. 14, no. 2, february 2018

[17] Mohd Fazil and Muhammad Abulaish "A hybrid approach for detecting automated spammers in Twitter " ieee transactions on industrial informatics, vol. 13, no.11 november 2018

[18] Muhammad usman Shahid khan , mazhar ali Assad Abbas, Sameer. Khan, and Albert y. Zomaya "Segregating Spammers and Unsolicited Bloggers from Genuine Experts on Twitter " ACM SigKDD Explorations Newsletter, vol. 11, no. 2, pp. 100–108, 2010.

[19] Stefano Cresci, , Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi , and Maurizio Tesconi "Social Fingerprinting: Detection of Spambot Groups Through DNA-Inspired Behavioral Modeling"ACM SigKDD Explorations Newsletter, vol. 15, no. 4, pp. 100–108, 2018. Chao Chen, Yu Wang, Jun Zhang, Yang Xiang, Wanlei Zhou, and Geyong Min "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam"ieee transactions on industrial informatics, vol. 12, no. 4, April 2018

[20] Manuel Egele, Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna "Towards Detecting Compromised Accounts on Social Networks "ieee transactions on industrial informatics, vol. 14, no. 4, August 2017

[21] Yubao Zhang, , Xin Ruan,, Haining Wang, Hui Wang, and Su He "Twitter Trends Manipulation: A First Look Inside the Security of Twitter Trending" ieee transactions on industrial informatics, vol. 12, january 2017

[22] J. Zhang, R. Zhang, Y. Zhang, and G. Yan, "The rise of social botnets: Attacks and counter measures, "IEEE Trans. Depend. Sec. Comput.,2016.

[23] J. Zhang, R. Zhang, J. Sun, Y. Zhang, and C. Zhang, "Truetop: A Sybil resilient system for user influence measurement on twitter," IEEE/ACM Trans. Netw., vol. 24, no. 5, pp. 2834–2846, Oct. 2016.

[24] Y. Xuan, Y. Chen, H. Li, P. Hui, and L. Shi, "Lbsnshield: Malicious account detection in location-based social networks," in Proc. 19th ACM Conf. Comput. Supported Cooperative Work Social Comput. Companion, 2016, pp. 437–440.

[25] W. Wu, J. Alvarez, C. Liu, and H.-M. Sun, "Bot detection using unsupervised machine learning," Microsystem Technologies. New York, NY, USA: Springer-Verlag, 2016, pp. 1–9.

[26] G. Wang, X. Zhang, S. Tang, H. Zheng, and B. Y. Zhao, "Unsupervised click stream clustering for user behavior analysis," in Proc. CHI Conf. Hum. Factors Comput. Syst., 2016, pp. 225–236.