Automatic Recognition of Fake Profiles on Social Media: A Survey

Fazel[1], Ritika Sharma[2],
[1]*M.Tech Student, Desh Bhagat University,Mandi Gobindgarh*
[2] *Assistant Professor, Desh Bhagat University,Mandi Gobindgarh*
(E-mail: fazelshah1@gmail.com)

***Abstract*—**At present social network sites are part of the life for most of the people. Every day several people are creating their profiles on the social network platforms and they are interacting with others independent of the user's location and time. The social network sites not only providing advantages to the users and also provide security issues to the users as well their information .Fake accounts are a preferred means for malicious users of online social networks to send spam, commit fraud, or otherwise abuse the system. A single malicious actor may create dozens to thousands of fake accounts in order to scale their operation to reach the maximum number of legitimate members. Detecting and taking action on these accounts as quickly as possible is imperative in order to protect legitimate members and maintain the trustworthiness of the network. However, any individual fake account may appear to be legitimate on first inspection, for example by having a real-sounding name or a believable profile. In this literature survey, we review the existing research work done on Facebook, and study the techniques used to identify and analyze poor quality content on Facebook, and other social networks. We also attempt to understand the limitations posed by Facebook in terms of availability of data for collection, and analysis, and try to understand if existing techniques can be used to identify and study poor quality content on Facebook.

Keywords*—*Conceptual Similarity, Trademark Infringement, Trademark Retrieval, Trademark Similarity.

I. INTRODUCTION

Online social networks(OSNs), such as Facebook, Twitter, RenRen, LinkedIn, Google+, and Tuenti, have become increasingly popular over last few years. People use OSNs to keep in touch with each others, share news, organize events, and even run their own e-business. For the period between 2014 and 2018 around 2.53 million U.S. dollars have been spent on sponsoring political ads on Facebook by non-profits [1]. The open nature of OSNs and the massive amount of personal data for its subscribers have made them vulnerable to Sybil attacks [2].

In 2012, Facebook noticed an abuse on their platform including publishing false news, hate speech, sensational and polarizing, and some others [3]. However, online Social Networks (OSNs) have also attracted the interest of researchers for mining and analyzing their massive amount of data, exploring and studying user's behaviors as well as detecting their abnormal activities [4].

In [5] researchers have made a study to predict, analyze and explain customers loyalty towards a social media-based online brand community, by identifying the most effective cognitive features that predict their customers attitude. Facebook community continues to grow with more than 2.2 billion monthly active users and 1.4 billion Daily active users, with an increase of 11% on a year-over-year basis [6].

In the second quarter of 2018 alone, Facebook reported that its total revenue was $13.2 billion with $13.0 billion from ads only [6]. Similarly, in second quarter of 2018 Twitter has reported reaching about one billion of Twitter subscribers, with 335 million monthly active users [7]. In 2017 twitter reported a steady revenue growth of 2.44 billion U.S. dollars, with 108 million U.S. dollars lower profit compared to the previous year [7].

In 2015 Facebook estimated that nearly 14 million of its monthly active users are in fact undesirable, representing malicious fake accounts that have been created in violation of the websites terms of service [8]. Facebook, for the first time, shared a report in the first quarter of 2018 that shows their internal guidelines used to enforce community standards covering their efforts between October 2017 to March 2018, this report illustrates the amount of undesirable content that has been removed by Facebook, and it covers six categories: graphic violence, adult nudity and sexual activity, terrorist propaganda, hate speech, spam, and fake accounts [9].

837 million posts containing spam have been taken down, and about 583 million fake accounts have been disabled, Facebook also has removed around 81 million undesirable content in terms of the rest violating content types. However, even after preventing millions of fake accounts from Facebook, it was estimated that, around 88 million accounts, are still fake [9].

For such OSNs, the existence of fake accounts lead advertisers, developers, and inventors to distrust their reported user metrics, which would negatively impacts their revenues as recently, banks and financial institutions in U.S. have started to analyze Twitter and Facebook accounts of loan applicants, before actually granting the loan [10]. Attackers follow the concept of having OSNs user accounts are "keys to walled gardens" [11], so they deceive themselves off as somebody else, by using photos and profiles that are either snatched from a real person without his/her knowledge, or are generated

artificially, to spread fake news, and steal personal information. These fake accounts are generally called imposters [8], [12].

In both cases, such fake accounts have a harmful effect on users, and their motives would be anything other than good intentions as they usually flood spam messages, or steal private data. They are keen to phish individual naive users to phony relationships that lead to sex scam, human trafficking, and even political astroturfing [13], [14], [15], [8], and [12]. Statistics show that 40% of parents in the United States and 18% of teens have a great concern about the use of fake accounts and bots on social media to sell or influence products [16].

Another example, during the 2012 US election campaign, the Twitter account of challenger Romney experienced a sudden jump in the number of followers. The great majority of them were later claimed to be fake followers [17]. To enhance their effectiveness, these malicious accounts are often armed with stealthy automated tweeting programs, to mimic real users, known as bots [14]. In December 2015, Adrian Chen, a reporter for the New Yorker, noted that he had seen a lot of the Russian accounts that he was tracking switch to pro-Trump efforts, but many of those were accounts that were better described as troll's accounts managed by real people that were meant to mimic American social media users [18].

Similarly, before the general Italian elections of February 2013, online blogs and newspapers reported statistical data over a supposed percentage of fake followers of major candidates [19]. Detecting those threatening accounts in OSNs has become a must to avoid various malicious activities, insure security of user's accounts and protect personal information. Researchers attempt to come up with automated detection tools for identifying fake accounts, which would be labor-intensive and costly if done manually. The implications of researchers attempt may allow an OSN operator detecting fake accounts efficiently and effectively, it would improve the experience of its users by preventing annoying spam messages and other abusive content. The OSN operator can also increase the credibility of its user metrics and enable third parties to consider its user accounts [20].

Information security and privacy are among the primary requirements of social network users, maintaining and providing those requirements increases network credibility and subsequently its revenues. OSNs are employing different detecting algorithms and mitigation approaches to address the growing threat of fake/malicious accounts. Researchers focus on identifying fake accounts through analyzing user level activity by extracting features from recent users e.g. number of posts, number of followers, profiles. They apply trained machine learning technique for real/fake accounts classification [8], [21].

Another approach is using graph level structure where the OSN is modeled as a graph essentially presented as a collection of nodes and edges. Each node represents an entity (e.g. account), and each edge represents as a relationship (e.g. friendship) [20], [22]. Though Sybil accounts find a way to cloak their behavior with patterns resembling real accounts [14], [23], [24], they manifest numerous profile features and activity patterns. Thus, automated Sybil detection are not always robust against adversarial attacks, and does not yield desirable accuracy.

II. RELATED WORK

Prevention of Fake Profile Proliferation in Online Social Networks (2015) Today, Online Social Networks (OSNs) are the most common platforms on the Internet, on which millions of users register to share personal facts with their friends. Online social network users are unaware of the numerous security risks that exist in these networks, like privacy violation, identity theft and sexual harassment etc. Many users disclose their personal information like phone no., date of birth, address etc. Leakage of personal information is a significant Concern for social network users. Fake profiles are being created in all the sites and one's information is becoming more and more vulnerable in the past decade. Nowadays the Identity Clone Attack (ICA) is increased in the many social networking websites that causes the frustration between the peoples and social networking websites too. This attack is done by retrieving the information of the individuals profile by anonymous person i.e. individual information is leaked and clone or fake profile is created which shows as real one. Thus this leads to the ambiguity between the owner of the profiles and the person associated to their profile i.e. we cannot have control to create over creation of clone profiles in the OSN and impacts it to the person having his or her own profiles. Hence, a new way of protecting personal information on online social sites is being proposed in this paper.

Implications of Various Fake Profile Detection Techniques in Social Networks In the recent years, the fast development and the exponential utilization of social networks has prompted an expansion of social Computing. In social networks users are interconnected by edges or links. Facebook, twitter, LinkedIn are most popular social networks websites. In this paper focus is made on Facebook for detection of fake profile. Facebook is most used social networking site in which user can share messages, images and videos also users may add number of friends in their personal profiles. But it is difficult to find out whether the new person is genuine or not. May be it could be a malicious user. To detect malicious users or fake profiles different techniques has been proposed. In this paper an attempt has been made to analysis various existing techniques that includes comparison in perspective of various applications mapping various performance parameters.

Automatic detection of fake profiles (2015) this paper presents the study of various methods for detection of fake profiles. In this paper a study of various papers is done, and in the reviewed paper we explain the algorithm and methods for detecting fake profiles for security purpose. The main part of this paper covers the security assessment of security on social networking sites.

An IAC Approach for Detecting Profile Cloning in Online Social Networks (2014) Nowadays, Online Social Networks (OSNs) are popular websites on the internet, which millions of users register on and share their own personal information with others. Privacy threats and disclosing personal information are the most important concerns of OSNs' users. Recently, a new attack which is named Identity Cloned Attack is detected on

OSNs. In this attack the attacker tries to make a fake identity of a real user in order to access to private information of the users' friends which they do not publish on the public profiles. In today OSNs, there are some verification services, but they are not active services and they are useful for users who are familiar with online identity issues. In this paper, Identity cloned attacks are explained in more details and a new and precise method to detect profile cloning in online social networks is proposed. In this method, first, the social network is shown in a form of graph, then, according to similarities among users, this graph is divided into smaller communities. Afterwards, all of the similar profiles to the real profile are gathered (from the same community), then strength of relationship (among all selected profiles and the real profile) is calculated, and those which have the less strength of relationship will be verified by mutual friend system. In this study, in order to evaluate the effectiveness of proposed method, all steps are applied on a dataset of Facebook, and finally this work is compared with two previous works by applying them on the dataset.

Towards Detecting Compromised Accounts on Social Networks Social Network accounts has become a profitable course of action for cybercriminals. By hijacking control of a popular media or business account, attackers can distribute their malicious messages or disseminate fake information to a large user base. The impacts of these incidents range from a tarnished reputation to multi-billion dollar monetary losses on financial markets. In our previous work, we demonstrated how we can detect large-scale compromises (i.e., so-called campaigns) of regular online social network users. In this work, we show how we can use similar techniques to identify compromises of individual high-profile accounts. High-profile accounts frequently have one characteristic that makes this detection reliable – they show consistent behavior over time. We show that our system, were it deployed, would have been able to detect and prevent three real-world attacks against popular companies and news agencies. Furthermore, our system, in contrast to popular media, would not have fallen for a staged compromise instigated by a US restaurant chain for publicity reasons.

Fake Identities in Social Media: A Case Study on the Sustainability of the Facebook Business Model Social networks such as Facebook, Twitter and Google+ have attracted millions of users in the last years. One of the most widely used social networks, Facebook, recently had an initial public offering (IPO) in May 2012, which was among the biggest in Internet technology. For profit and nonprofit organizations primarily use such platforms for target- oriented advertising and large-scale marketing campaigns. Social networks have attracted worldwide attention because of their potential to address millions of users and possible future customers. The potential of social networks is often misused by malicious users who extract sensitive private information of unaware users. One of the most common ways of performing a large-scale data harvesting attack is the use of fake profiles, where malicious users present themselves in profiles impersonating fictitious or real persons.

III. DETECTION OF FAKE PROFILES

Fake identities in social media are often used in APT cases, both to gather intelligence prior the attack, and to establish trust and deliver malware or a link to it. Such fake identities are also used in other types of malicious activities. To combat these activities, a significant body of research to date has focused on the timely and accurate detection of the presence of a fake identity in social media. Generally, following the taxonomy in Song et al. (2015), the approaches to detecting false social media accounts can be classified into the approaches aimed analyzing individual accounts (profile-based techniques as well as graph-based methods), and the approaches capturing the coordinated activities spanning a large group of accounts.

For instance, the paper Nazir et al. (2010) describes detecting and characterizing phantom profiles in online social gaming applications. The article analyses a Facebook application, the online game "Fighters club", known to provide incentives and gaming advantage to those users who invite their peers into the game. The authors argue that by providing such incentives the game motivates its players to create fake profiles. By introducing those fake profiles into game, the user would increase incentive value for him/herself. At first, the authors extract 13 features for each game user, and then perform classification using support vector machines (SVMs). The paper concludes that these methods do not suggest any obvious discriminants between real and fake users.

Adikari and Dutta (2014) describe identification of fake profiles in LinkedIn. The paper shows that fake profiles can be detected with 84% accuracy and 2.44% false negative, using limited profile data as input. Methods such as neural networks, SVMs, and principal component analysis are applied. Among others, features such as number of languages spoken, education, skills, recommendations, interests, and awards are used. Characteristics of profiles, known to be fake, posted on special web sites are used as a ground truth.

Chu et al. (2010) aim at differentiating Twitter accounts operated by human, bots, or cyborgs (i.e., bots and humans working in concert). As a part of the detection problem formulation, the detection of spamming accounts is realized with the help of an Orthogonal Sparse Bigram (OSB) text classifier that uses pairs of words as features. Accompanied with other detecting components assessing the regularity of tweets and some account properties such as the frequency and types of URLs and the use of APIs, the system was able to accurately distinguish the bots and the human-operated accounts.

Detecting spamming accounts in Twitter as well as in My Space, was also the objective of the study by Lee et al. (2010). As compared with the study by Chu et al., the set of features here was expanded to cover also the number and type of connections. A number of classifiers available in Weka machine learning suite were tried, and the Decorate meta classifier was found to provide the best classification accuracy. In addition to, or instead of analyzing the individual profiles, another stream of approaches rely on graph-based features when distinguishing the fake and legitimate accounts. For instance, Stringhini et al. (2010) describe methods for spam detection in Facebook and Twitter. The authors created 900

honeypot profiles in social networks, and performed continuous collection of incoming messages and friend requests for 12 months. User data of those who performed these requests were collected and analyzed, after which about 16K spam accounts were detected. Authors further investigated the application of machine learning for further detection of spamming profiles. On top of the features used in the studies above, the authors were also using the message similarity, the presence of patterns behind the search of friends to add, and the ratio of friend requests, and then used Random Forest as a classifier.

Seeking robust features to detect spamming Twitter accounts was also the focus of the work by C. Yang et al. (2011). Graph based features and neighbor-based features were combined with automation-based features and timing-based features in order to construct four different classifiers. A similar approach, although with a much smaller set of features were employed by Z. Yang et al. (2011) to detect fake accounts in Renren.

Clustering coefficient was used as a metric reflecting the properties of the social graphs. These features were used to build a SVMs classifier that resulted in 99% correct classifications. Papers by Cao et al. (2011) and Conti et al. (2012) likewise propose an application of graph features for the detection of fake profiles. Cao et al. (2011) base their detection on the observation that fake (Sybil) profiles typically connect to other fake profiles, rather than the legitimate ones. Thus, there is a cut between fake and non-fake subgraphs in the graph. Conti et al. (2012) base their detection method on analysis of distribution of number of friends over time. Boshmaf et al. (2016), however, claim that the hypothesis that fake accounts mostly befriend other fake accounts does not hold, and propose a new detection method, which is based on analysis features of victim accounts, i.e. those accounts, which were befriended by a fake account.

Finally, Zang et al. (2013), under the assumption that the user of a Sybil account is unable to establish a large number of friendship relationships to non-Sybil nodes, proposed the use of a generative probabilistic block model to model the growth of the social network graph and identify latent groups within this graph. Often times, the profile-based approaches overviewed above are aimed at detecting the accounts involved in spamming. Traditional spamming, however, targets a large audience of receivers, as opposed to the spear phishing campaigns common in advanced persistent threats where a single individual or a small group of recipients is targeted instead. It is therefore unclear whether these techniques, unmodified, would perform equally well when detecting fake accounts involved in an advanced persistent threat.

This limitation is partially addressed in a work by Egele et al. (2015) who, instead of characterizing the profiles of spamming accounts, attempt to detect the cases when a high-profile legitimate account is (temporarily) subverted and acts maliciously. To this end, the authors are seeking for behavioral anomalies in these accounts, by monitoring the timing and the origin of the messages, language and message topic, URLs, use of direct interaction, and geographical proximity. These are used to construct a SVM classifier based on sequential minimal optimization algorithm. The dataset was semiannually labelled: the messages with malicious URLs within messages, abruptly changed topics, or malicious URLs within application description pages were seen as indications of compromised profiles. The idea of detecting (dis)similarities in user behavior was also explored in the work by Egele et al. (2015). Albeit focusing on interaction over email messages rather than through social networks, the authors nevertheless strive to detect spear phishing by profiling individual email writers and then recognizing whether a new coming email does really originate from the same profile.

Instead of analyzing individual profiles and their connections, many researchers focus on characterizing malicious activities involving a coordinated use of numerous accounts – for instance, in the context of black markets of bots and fake accounts for online social networks. Stringhini et al. (2013) analyses Twitter follower markets. They describe the characteristics of Twitter follower markets and classify the customers of the markets. The authors argue that there are two major types of accounts who follow the "customer": fake accounts ("sybils"), and compromised accounts, owners of which do not suspect that their followers' list is increasing. Customers of follower markets may be celebrities or politicians, aiming to give the appearance of having a larger fan base, or may be cyber criminals, aiming at making their account look more genuine, so they can quickly spread malware and spam.

Thomas et al. (2013) investigate black market accounts used for distributing Twitter spam. De Cristofaro et al. (2014) analyses Facebook like farms by deploying honeypot pages. Viswanath et al. (2014) detect black-market Facebook accounts based on the analysis of anomalies in their like behavior. Farooqi et al. (2015) investigate two black-hat online marketplaces, SEO Clerks and My Cheap Jobs.

Fayazi et al. (2015) study manipulation in online reviews. A specific type of large-scale fake account creation campaigns is referred to as crowdturfing, the term representing a merger of two other terms, astroturfing (i.e., sponsored information dissemination campaigns obfuscated to appear spontaneous movements) and crowdsourcing. Thus, WEBIST 2017 - 13th International Conference on Web Information Systems and Technologies 366 crowdturfing is malicious crowdsourcing. Song et al. (2015) study how to detect objects of crowdturfing tasks in Twitter. In particular,

Wang et al. (2012) describe the operational structure of crowdturfing systems, by both crawling the websites used for coordinating crowdturfing campaigns, and by executing a similar, though benign campaign of their own. The authors have found these campaigns to be highly effective in hiring users, and, given the growth in their popularity, they thus pose a serious threat to security. In a subsequent paper,

Wang et al. (2014) study the applicability of machine learning approaches to detect crowdturfing campaigns, and the robustness of these approaches to being evaded by the adversaries. The paper suggests that traditional machine learning can be used to detect crowdturfing workers with the accuracy of 95-99%, albeit the detection can be relatively easily evaded if the workers adjust their behavior.

Lee et al. (2014, 2015) likewise aim at developing a method for detecting crowdturfing campaigns. The classifier built by the authors was able to achieve crowdturfing task detection accuracy of 97.35%. Further, based on comparing the profiles of crowdturfing workers at Twitter against the generic Twitter user profiles, the authors constructed a classifier that detected Twitter crowdturfing users with 99.29% accuracy. The distinguishing features used by this classifier included, among others, the variability of the number of followers over time, the graph density of the worker accounts, tweeting activity, and ratio of friends and followers.

Song et al. (2015) has proposed another method for detecting crowdturfing, CrowdTarget. Rather than aiming at detecting workers, the authors focus on detecting the target objects of crowdturfing tasks (e.g., post, page, and URL). The proposed method can successfully distinguish between crowdturfing and benign tweets with the true positive rate up to 98%, even when they both come from the same account, thus making it more robust to detection evasion techniques. The following features were proven to be discriminative: (i) retweet time distribution, (ii) the ratio of the most dominant application, (iii) the number of unreachable retweeters, and (iv) the number of received clicks. Alas, similarly to the approaches above targeting the detection of spamming campaigns, the crowdturfing detection techniques also assume the presence of a large scale activity, and are therefore hardly able to detect a small-footprint activity carried out as a part of a targeted attack.

Krombholz et al. (2015) proposes classification of social engineering attacks into physical methods (such as dumpster diving), social approaches (relying on socio-psychological techniques), reverse social engineering (attacker attempts to make victim believe that she is a trustworthy entity, and the goal is to make the victim approach attacker e.g. for help), technical approaches, and socio-technical approaches (combining approaches above).

Kontaxis et al. (2011) describe prototype of the software which aims at finding whether profile of particular user was cloned from one online social network into another by comparing characteristics of the profiles having similar characteristics among several online social networks.

Krombholz et al. (2012) propose the raising of users' awareness as the most efficient countermeasure against social media identity theft, and describes the methods for it. Authors perform focus groups research, and suggest that the users are mostly unaware of fake profiles occurrence and its consequences.

Jiang et al. (2016) surveyed more than 100 advanced techniques for detecting suspicious behaviors that have existed over the past 10 years and presented several experimentally successful detection techniques (i.e. CopyCatch, which was described in (Beutel et al., 2013)).

IV. CONCLUSION

False identities in the form of compromised or fake email accounts, accounts in social media, fake or cracked websites, fake domain names, and malicious Tor nodes, are heavily used in APT attacks, especially in their initial phases, and in other malicious activities. Using these fake identities, the attacker(s) aim at establishing trust with the target and at crafting and mounting a spear phishing or another attack. Based on research evidence, information gathering for a spear phishing attack heavily relies on the use of social media and fake accounts therein. It is therefore important to detect, as early as possible, the presence of a fake social media account. A number of recent research works have focused on detecting such fake accounts, either by analyzing the characteristics of individual profiles and their connections, or – in case of coordinated activities, by multiple fake social media accounts, Detection of Fake Profiles in Social Media - Literature Review 367 such as in the case of crowdturfing – by analyzing the commonality of these activities, too. The main shortcoming of the majority of these research works is their implicit assumption that the owners of the fake social media accounts target a large audience of followers. While such an assumption may be valid in case of traditional spamming campaigns or in case of crowdturfing, the spear phishing commonly used in APT exhibits a different pattern of targeting only a small subset of individuals, and otherwise keeping a low profile to evade detection. As a result, the proposed detection techniques often expect, e.g., a high ratio of accepted friend requests, which is unlikely in APT. This invalid assumption, as well as the availability of other evading techniques, makes it relatively easy for the attacker behind an APT to circumvent detection the contribution of this paper consists of the literature review of current research aimed at detecting fake profiles in social media from an advanced persistent threats point of view

REFERENCES

[1] Political advertising spending on facebook between 2014 and 2018. Internet draft. [Online]. Available: https://www.statista.com/statistics/891327/political-advertisingspending-facebook-by-sponsor-category/

[2] J. R. Douceur, "The sybil attack," in International workshop on peerto-peer systems. Springer, 2002, pp. 251–260.

[3] Cbc.facebook shares drop on news of fake accounts. Internet draft. [Online]. Available: http://www.cbc.ca/news/technology/facebook-shares-drop-onnews-of-fake-accounts-1.1177067

[4] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," Egyptian informatics journal, vol. 17, no. 2, pp. 199–216, 2016.

[5] L. M. Potgieter and R. Naidoo, "Factors explaining user loyalty in a social media-based brand community," South African Journal of Information Management, vol. 19, no. 1, pp. 1–9, 2017.

[6] Quarterly earning reports. Internet draft. [Online]. Available: https://investor.fb.com/home/default.aspx

[7] Statista.twitter: number of monthly active users 2010-2018. Internet draft. [Online]. Available: https://www.statista.com/statistics/282087/number-of-monthlyactive-twitter-users/

[8] Y. Boshmaf, M. Ripeanu, K. Beznosov, and E. Santos-Neto, "Thwarting fake osn accounts by predicting their victims," in Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security. ACM, 2015, pp. 81–89.

[9] Facebook publishes enforcement numbers for the first time. Internet draft. [Online]. Available: https://newsroom.fb.com/news/2018/05/enforcement-numbers/

[10] Banque populaire dis-moi combien damis tu as sur facebook, je te dirai si ta banque va taccorder un prłt. Internet draft. [Online]. Available:

http://bigbrowser.blog.lemonde.fr/2013/09/19/popularitedis-moi-combien-damis-tu-as-sur-facebook-je-te-dirai-si-ta-banqueva-taccorder-un-pret/

[11] S.-T. Sun, Y. Boshmaf, K. Hawkey, and K. Beznosov, "A billion keys, but few locks: the crisis of web single sign-on," in Proceedings of the 2010 New Security Paradigms Workshop. ACM, 2010, pp. 61–72.

[12] S. Fong, Y. Zhuang, and J. He, "Not every friend on a social network can be trusted: Classifying imposters using decision trees," in Future Generation Communication Technology (FGCT), 2012 International Conference on. IEEE, 2012, pp. 58–63.

[13] K. Thomas, C. Grier, D. Song, and V. Paxson, "Suspended accounts in retrospect: an analysis of twitter spam," in Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011, pp. 243–258.

[14] 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. ACM, 2011, pp. 93–102.

[15] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonc¸alves, S. Patil, A. Flammini, and F. Menczer, "Truthy: mapping the spread of astroturf in microblog streams," in Proceedings of the 20th international conference companion on World wide web. ACM, 2011, pp. 249–252.

[16] How concerned are you that there are fake accounts and bots on social media platforms that are used to try to sell you things or influence you? Internet draft. [Online]. Available: https://www.statista.com/statistics/881017/fakesocial-media-accounts-bots-influencing-selling-purchases-usa/

[17] Buying their way to twitter fame. Internet draft. [Online]. Available: https://investor.fb.com/home/default.aspxhttp://www.nytimes.com/2012/08/23/fashion/twitterfollowers-for-sale.html?smid=pl-share

[18] Welcome to the era of the bot as political boogeyman. Internet draft. [Online]. Available: https://www.washingtonpost.com/news/politics/wp/2017/06/12/welcometo-the-era-of-the-bot-as-political-boogeyman/?utmterm = .2271ba8db710

[19] Human or 'bot'? doubts over italian comic beppe grillo's twitter followers. Internet draft. [Online]. Available: https://www.telegraph.co.uk/technology/twitter/9421072/Human-orbot-Doubts-over-Italian-comic-Beppe-Grillos-Twitter-followers.html

[20] Y. Boshmaf, D. Logothetis, G. Siganos, J. Ler´ıa, J. Lorenzo, M. Ripeanu, K. Beznosov, and H. Halawa, "´Integro: Leveraging victim prediction for robust fake account detection in large scale osns," Computers & Security, vol. 61, pp. 142–168, 2016.

[21] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, 2012, pp. 15–15.

[22] L. Alvisi, A. Clement, A. Epasto, S. Lattanzi, and A. Panconesi, "Sok: The evolution of sybil defense via social networks," in Security and Privacy (SP), 2013 IEEE Symposium on. IEEE, 2013, pp. 382–396.

[23] P. Patel, K. Kannoorpatti, B. Shanmugam, S. Azam, and K. C. Yeo, "A theoretical review of social media usage by cyber-criminals," in Computer Communication and Informatics (ICCCI), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[24] M. Tsikerdekis and S. Zeadally, "Multiple account identity deception detection in social media using nonverbal behavior," IEEE Transactions on Information Forensics and Security, vol. 9, no. 8, pp. 1311–1321, 2014.