Strangers Intrusion Detection - Detecting Spammers and Fake Profiles in Social Networks Based on Topology Anomalies

Michael Fire, Gilad Katz, Yuval Elovici
Telekom Innovation Laboratories and Information Systems Engineering Department,
Ben-Gurion University of the Negev, Beer-Sheva, Israel
Email: {mickyfi, katzgila, elovici}@bgu.ac.il

***Abstract*—Today's social networks are plagued by numerous types of malicious profiles which can range from socialbots to sexual predators. We present a novel method for the detection of these malicious profiles by using the social network's own topological features only. Reliance on these features alone ensures that the proposed method is generic enough to be applied on a range of social networks. The algorithm has been evaluated on several social networks and was found to be effective in detecting various types of malicious profiles. We believe this method is a valuable step in the increasing battle against social network spammers, socialbots, and sexual predictors.**

I. Introduction

In recent years we have seen a surge in the use of online social networks. Online social network such as Facebook[1], Twitter[2], Google+[3], MySpace[4], Bebo[5], Academia.edu[6], and AnyBeat[7] have been growing at exponential rates and now serve hundreds of millions of users on a daily bases. The Facebook social network, for example, was founded in 2004 and has more than 901 million monthly active users as of March 2012 [1]. Facebook users have an 130 average friends and create around 90 pieces of content each month. Due to the sharing nature of online social networks, users expose many personal details about themselves, either intentionally or unintentionall; details, such as date of birth, email address, high school name, and even phone numbers are frequently exposed [2], [3]. In recent years, users and their personal details have been a main target for many different online attacks which can threaten the well-being of users in both the virtual and the real world. These attacks include identity-theft [4], [5], user de-anonimzation [6], inference attacks [7], viruses [8], click-jacking [9], phishing [10], sybil attacks [11], reverse social engineering [12], and socialbots [3]. Cybercriminal attackers have a vault full of combination attacks in order to collect users' personal information and gain their trust. By using the user's collected personal information, an attacker can send personally crafted spam messages in an attempt to lure such users into malicious websites [13] or even blackmail them into transferring money to the attacker's account [14]. In some cases, an attacker can be an online "predator", who uses online attacks in order to gain information which will enable them to obtain the user's trust and convince the user to meet in real life [15], [16]. In many cases, social network attackers attempt to cover their tracks by using fake profiles. Moreover, the number of fake profiles in online social networks can be counted by the millions. Facebook, for example, estimates that around 5% of its users could be false or duplicate accounts [17].

In this paper we present an algorithm for the detection of spammers and fake profiles in social networks. Our algorithm, which is based solely on the topology of the social network, detects users who randomly connect to others by detecting the anomalies in that network's topology. According to previously conducted research, social network users who connect randomly to other users in the network may be fake profiles [3], spammers [13], or even online "predators". Our algorithm uses the fact that social networks are scale-free [18] and have a community structure [19]. This fact ensures that most of the users in the network have a small degree and are connected only to a small number of communities. Fake profiles, on the other hand, tend to establish friendship connections with users from different communities. In order to evaluate our algorithm, we ran it on different directed online social networks structures with different levels of anonymity: Academia.edu, AnyBeat, and Google+.

We first developed a code which simulates a single fake user's infiltration effort into the social network. Subsequently, we used supervised learning algorithms in order to detect our fake profiles and other real profiles with similar features in the social networks. Lastly, we evaluated our algorithm results with the expertise of a committee of experts. By using our algorithms, we successfully detected fake profiles in real online social networks which use the social networks as a platform for collecting users' data (also known as Friend Infiltrators [20]), spamming and even distributing of sexual content (also known as Pornographic Storytellers [20]). Our algorithm was also successful in detecting other particular types of users who use the social network to connect to random users. For example, we detected users who use the social

[1]http://www.facebook.com
[2]http://www.twitter.com
[3]http://plus.google.com
[4]http://www.myspace.com
[5]http://www.bebo.com
[6]http://academia.edu
[7]http://www.anybeat.com

networks only as dating platform on which they only connect with random users of a specific gender.

The remainder of this paper is organized as follows. In section II, we give a brief overview of previous studies on social networks threats and protection solutions. We also describe several algorithms and definitions from graph theory and social networks analysis. In section III, we describe the different social network datasets used throughout this study. In section IV we describe the methods and experiments used in the construction and evaluation of our classifiers. In section V we present the results of our experiments, and in section VI we analyze the results. Finally, in section VII, we present our conclusions from this study and offer future research directions.

II. Related Work

A. Social Networks Threats

In recent years online social networks usage has grown exponentially. In today's world, an average user spends more time on popular social networking sites than on any other site [21], [22]. With the increasing usage of online social networks, users have become fertile ground for spammers, cybercriminals, and many other potential threats. These threats put social networks users at risk because users of these networks tend to publish personal information about themselves. This information is sensitive and may cause serious harm if obtained by the wrong people. A research carried out by Acquisti and Gross [2] evaluated the amount of personal information exposed by users on Facebook. It concluded that many Facebook users disclose personal information about themselves; this data includes (but is not limited to) dates of birth, email addresses, relationship statuses, and even phone numbers.

Another disturbing fact which was uncovered that around 80% of Facebook users accept friend requests from people they do not know if they share more than 11 mutual friends [3]. By accepting these friend requests, users disclose their private information to strangers [3], [23]. Moreover, by accepting friend requests from strangers, users can expose their friends to inference attacks [7], [24].

In recent years, social networks users have been exposed to other types of attacks as well. These attacks include: a) *viruses* which use the social networks as convenient spreading platforms [8], b) *click-jacking attacks* that attempt to hijack the user's web sessions [9], c) *phising attacks* that aim at fraudulently acquire a user's sensitive information by impersonating a trustworthy third party [10], d) *spammers* using the user social network data in order to send tailored spam messages to the user [13], e) *user de-anonimization attacks* that attempt to expose the identity of the social network user [6], f) *sybil attack* where the attacker obtains multiple fake identities and pretends to be multiple, distinct nodes in the system (sybil nodes); the attacker uses these nodes in order to harm the reputation of honest users in the network [11], g) *socialbots*, an army of fake profiles which aim to harvest users' personal data [3], and h) *clone and identity theft attacks* where attackers duplicate a user's online presence in the same network or across different networks in order to fool the cloned user's friends into forming a trusting relation with the cloned profile [4]. Moreover, online "predators" can also use social networks as a platform for finding their next victim. They do so by collecting personal information, gaining trust, and arranging encounters in the real world [15].

According to previous studies [3], [13], in many cases social network attackers, such as spammers and socialbots choose their victims randomly or according to certain criteria. Stringhini et al. [13] observed that many spammers on Facebook seemed to follow criteria, like a names list, when picking their victims. Moreover, in their study, Boshmaf et al. [3] used an army of socialbots to connect to users which were chosen at random. For these reasons, we assume in this paper that fake social nodes behave in the manner described above and devise our methods accordingly.

B. Social Networks Protection Solutions

In recent years, security companies, social network operators, and academic researchers have tried to cope with the above mentioned threats through a variety of solutions. Social networks operators try to protect their users by adding authentication processes to ensure that the registered user is a real person [3]. Many social network operators also support a configurable user privacy setting that enables users to protect their personal data from other users in the network [25], [26]. Additional protection may include defense against spammers, fake profiles, scams, and other threats. For example, Facebook users have the option to report abuse or policy violations by other users in the network [27]. In some countries, social networks such as Facebook and Bebo also added a "Panic Button" to better protect young people from other users in its social network [28]. Security companies like Checkpoint[8] and UnitedParents[9] offer users of social networks tools for protecting themselves. For example, the Checkpoint SocialGuard software [29] aims to protect its users from cyberbullies, predators, dangerous link, and strangers on Facebook.

In recent years, several published studies have attempted to propose solutions to different social networks threats. DeBarr and Wechsler [30] used the graph centrality measure in order to predict whether a user is likely to send spam. Wang proposed a method to classify spammers on Twitter by using content and graph based features [31]. Bosma et al. proposed a spam detection framework based on the HITS web link analysis framework [32]. Stringhini et al. proposed a solution for detecting spammers in social networks and succeeded in detecting spammers on Twitter, Facebook and MySpace social networks by using "honey-profiles" [13]. In the same study, Stringhini et al. also proposed a method for detecting spammer profiles by using supervised learning algorithms. Lee et al. also presented a method for detecting social spammers of different types by using honeypots combined

[8]http://www.checkpoint.com

[9]http://www.unitedparents.com

with machine learning algorithms [20]. In recent years, many defense solutions against sybil attacks (sybil defense) were proposed [33]–[36]. Many of these solutions were based on social network topology [37]. In 2011, Kontaxis et al. proposed a methodology for detecting social network profile cloning. They designed and implemented a prototype which can be employed by users and assists in investigating whether users have fallen victim to clone attacks [5].

C. Social Network Topology

In this study we make use of the fact that social networks are scale-free [18] and have a community structure [19]. Scale free networks are networks that obey the power law degree distribution $P(k) \sim ck^{-\gamma}$, where some node degrees greatly exceed the average. In addition, the nodes of such networks can be grouped into sets such that each set of nodes is densely connected internally. There are many algorithms with different properties for finding communities in social networks. In this study, we use the Louvain method, a greedy algorithm that attempts to optimize the "modularity" of a partition of the network [38]. We use this algorithm in order to split each of the tested social networks into communities and extract relevant features from them. Once the splits are completed, we extract various attributes and use them to train our classifiers. The combination of network attributes and machine learning is not new and has been discussed in works, such as Liben-Nowell and Kleinberg [39], Lee et al. [20], Stringhini et al. [13], and and Fire et al. [40].

III. Social Networks Datasets

In this study we evaluate our fake profile detection algorithm on three different directed social networks datasets: Academia.edu, AnyBeat, and Google+. Each one of the data sets mentioned above has a different size and a different anonymity level. In the remainder of this section we describe each of the datasets in detail.

Academia.edu. Academia.edu is a platform for academics to share and follow research papers. Members upload and share their papers with other researchers on over 350,000 research interests. An Academia.edu social network members may choose to follow any of the network's members, hence the directed nature of the links. In this study, we evaluated our algorithms on a major part of the network topology, containing more than 200,000 users and almost 1.4 million links. We obtained the Academia.edu network topology by using a dedicated web crawler in the beginning of 2011. Due to the nature of the social network, many users provide their first and last name in addition to their academic affiliation. The level of user anonymity in this network is therefore low. In this social network, we focused on attempting to detect fake profiles containing spam messages, profiles that provide false details, and those that present a fake picture of the member. In Academia.edu, we mainly look for spammers and fake profiles.

AnyBeat. "AnyBeat is an online community, a public gathering place where you can interact with people from around your neighborhood or across the world" [41]. AnyBeat

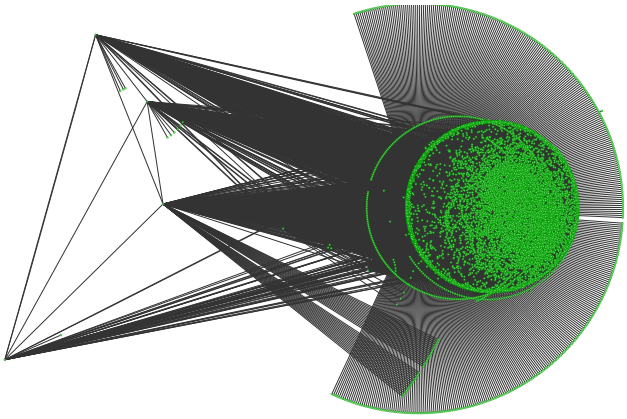

Fig. 1. AnyBeat Social Netwrok

is a relatively new social network where members can log in without using their real name, and where members can follow any other member in the network. In this study, we evaluated our algorithm on a major part of the network's topology, which was obtained using a dedicated web crawler. The topology contained 12,645 users and 67,053 links (see Figure 1). AnyBeat users enjoy a high level of anonymity and connections to strangers are common; therefore it relativity easy to activate fake profiles and connect to other users. In AnyBeat, we mainly look for fake profiles and pornographic storytellers.

Google+. Google+ is an online and directed social network with more than 170 million users[10]. Users can login using real or user names and can organize their contacts into circles, which are groups of information sharing. In this study we evaluate our algorithm on a subgraph of the network which contained more than 211,187 users and 1,506,896 links. All data was obtained by a dedicated crawler which collected publicly available data from each profile. Google+ users have a medium anonymity level where it is typical for a user to use his real name, but made-up names are common as well. In Google+, we mainly look for spammers, fake profiles, and pornographic storytellers.

TABLE I
Social Networks Datasets

	Academia.edu	**AnyBeat**	**Google+**
Nodes Num.	200K	12.6K	211K
Links Num.	1.4M	67K	1.5M
Anonymity	Low	High	Medium
Date	2011	2012	2012

IV. Methods and Experiments

The task of identifying fake profiles in social networks requires many non-trivial steps in order to be executed properly.

[10]http://googleblog.blogspot.com/2012/04/toward-simpler-more-beautiful-google.html

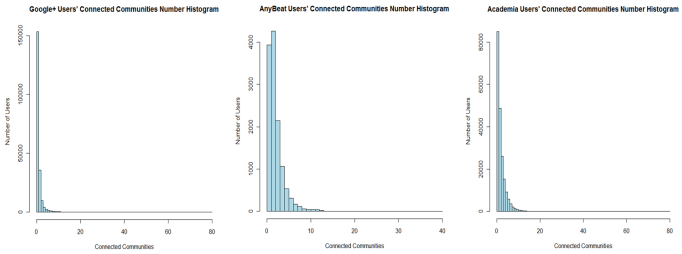

Fig. 2. Distribution of the number of communities each user is connected to in each one of the evaluated social networks.

In this study, we chose to apply methods from the domain of graph theory and supervised learning in order to achieve this goal. When using supervised learning techniques for detecting fake profiles, one of the main constraints is obtaining a set of positive examples in order to train our classifiers.

In our case, obtaining negative examples is a relatively easy task due the fact that in most cases, social network users are legitimate. However, obtaining positive examples of fake profiles is not an easy task due to the fact that many of them tend to camouflage themselves as legitimate ones. Therefore, in order to obtain positive examples of fake profiles, we developed a code which simulates the infiltration of fake users into the social networks by using random friend requests. Based on the guidelines and characteristics described in [3], [13]. We then used the simulated fake profiles as positive examples and chose random legitimate profiles from the network as negative examples. For each of the positive and negative examples, we extracted a features vector that was used as a train set for our fake profile detection classifiers. Next, we used the classifiers to detect other existing profiles in the social networks that had a high probability of being fake.

Lastly, we used a team of experts to manually evaluate a subset of our most likely fake profiles. The "control group" consisted of a set of randomly selected profiles. In the remainder of this section we describe each of these steps in detail.

A. Fake Profiles Infiltration Simulation

In order to create positive examples for our classifiers, we developed a code which simulates the infiltration of a single fake users (or a group of fake users) to direct social networks. For each social network, the simulator loaded the topology graph and inserted 100 new nodes which represented 100 fake users into the graph. The insertion process of each fake profile into the graph was done by simulating a series of "follow" requests sent to random users in the network. Each fake user had a limit on the number of friend requests in order to comply with a reality in which many social networks limit the number of user requests allowed for new members (exactly for the purpose of blocking spammers and socialbots).

In our case, the social networks were directed ones (Academia.edu, AnyBeat, and Google+), where each friend request was a "follow" request that did not need to be accepted in order to become active. Therefore, in order to create different types of fake users in directed social networks, we randomized the number of follow request of each user to be between 10 and 250.

B. Features Extraction

After obtaining a set of positive and negative user examples, we extracted a small set of features from each sample (user). For each user, we extracted the following four features: a) the degree of the user, b) the number of communities the user is connected to , c) how many connections exist among the friends of the user, and d) the average number of friends inside each of the user's connected communities.

The formal definition of the feature is as follows: Let $G =< V, E >$ be a directed graph which represents a social network topology. Let be C be the disjoint sets of all communities in G after G was partitioned into communities by the Louvian algorithm ($V = \bigcup_{C' \in C} \bigcup_{u \in C'} u$). We define the following features for each $u \in V$:

1) **The user degree**: $d(u) := |\Gamma(u)|$, where $\Gamma(u)$ is the neighborhood of u defined by:

$$\Gamma(u) := \{(v|(u,v) \in E \text{ or } (v,u) \in E\}.$$

2) **Users' connected communities number**:

$$cocommunities(u) := |\{C' \in C | v \in C' \text{ and } v \in \Gamma(u)\}|.$$

3) **The number of connections between u's friends**:

$$f - conn(u) := |(x,y) \in E | x \in \Gamma(u) \text{ and } y \in \Gamma(u)|.$$

4) **The average number of friends inside connected communities**:

$$avg\text{-}friends\text{-}comm(u) := \frac{d(u)}{cocommunities(u)}.$$

We calculated these four topological features based on methods and observations of previous studies [3], [13]. According to these studies, socialbots and other attackers tend to choose their victims randomly or according to specific criteria like age or gender. Therefore, we assume that malicious fake profiles are very likely to be connected to random users from different communities and have a high *Users' connected communities number*. Nevertheless, there are many celebrity profiles with a high user degree, like Britney Spears on Google+[11], who are also connected to many communities. However, due the fact that fake users choose to follow random users, we assume that the chances of the attacked users knowing each other are slim. Therefore, the value *the number of connections between user's friends* is predicted to be relatively low in such cases.

C. Constructing Classifiers

In order to construct fake profile detection classifiers, we created a subset of positive and negative examples with differing sizes from each social network. Similar to the study of Boshmaf et al. [3] which used 102 socialbots, we simulated an infiltration attack of 100 fake users on each tested social

[11] https://plus.google.com/100000772955143706751

network. We used these 100 profiles as a positive training set for each social network classifier. First, we used the fake profiles created by us as positive examples in a simulation. Then, for each social network, we filtered out some of the users, including some of the positive examples. The users who were removed had a relatively small number of friends, as they did not impose a serious threat to a large number of users in the networks. Moreover, a manual evaluation by experts of profiles with low degrees usually resulted in inconclusive results regarding the validity of these users. We then randomly chose negative examples from each social network. Due to the fact we randomly chose profiles as our negative trainset, some of these profiles may have actually been fake. Therefore, we preformed empirical test and evaluated the false-positive rates which were obtained by using negative trainsets of different sizes. We discovered that for the three evaluated social networks, a size of 500[12] negative examples for Anybeat network and a size of 3,000 negative examples for Google+ and Acadeima.edu networks performed well as negative trainset and returned low false-positive rates(see Table II).

After the completion of this process, we obtained the following train set for each social network:

- **Academia.edu**: In Academia.edu, we removed all the users with a degree less than 21, leaving us with 23,759 users (the absolute majority of the members in the network). Our train set was constructed from 93 positive and 2,999 negative examples.
- **AnyBeat**: In AnyBeat, which is a new and small social network, we removed users with a degree less than 6. This decision enabled us to have an overall of 3,208 users. Our train set was constructed from 100 positive and 499 negative examples
- **Google+**: In Google+, due the fact that we obtained only a small partition of the network, we did not remove any users. Our train set was constructed from 100 positive and 3,000 negative examples.

Lastly, we used the WEKA software [42] together with the train set extracted from each social network to construct classifiers for the different social networks. For each social network, we constructed both a decision tree (J48) and Naive Bayes classifiers. These simple classifiers were used in order to detect fake profiles inside the social networks.

D. Evaluation

After the classifiers were created, they had to be evaluated. This was done in two ways. First, we used a 10-fold cross validation in order to determine how well each classifier is capable of identifying our made-up fake profiles in the train sets. Secondly, we attempted to determine whether the classifier was right in flagging some "original" social network profiles as fake.

In addition to these "suspected" profiles, a list of randomly selected profiles was also chosen as a control group. These two lists, both of the same length, were combined into one list with random ordering. The final list was sent to a team of experts for evaluation.

Our team of experts consisted of four people with different backgrounds: a) two fourth year B.Sc. students from the information systems engineering department whose final project focused on protecting users in online social networks, b) one Ph.D. student with knowledge in the field of network security, and c) one Human Resource manager whose expertise is in hiring people for high-tech companies.

Each profile in our list was evaluated by three of these individuals, with each individual spending several minutes evaluating each profile. For every profile, the experts were asked to answer three questions: a) is the profile fake?, b) does the profile belong to a spammer? , and c) is the owner of the profile interested in following users from a specific gender? The possible answers for each of these questions were yes/no/maybe. The experts were also instructed not to be satisfied with a cursory examination but instead to "dig deep" whenever relevant (especially in Academia.edu). Names were fed to search engines in order to check the truthfulness of the information and profile pictures were searched in order to determine their originality (this was done using Google Search by Image service [43]).

V. RESULTS

We evaluated the classifiers' results in two ways. First, we evaluated each classifier on the train set using a 10 folds cross-validation. For each classifier, we measured false positive rates, f-measure, and AUC (area under the ROC curve) in order to evaluate the classifiers' performance (see Table II). Following this, we used the classifiers to identify other users in the social network who have a high probability of being either fake or spammer profiles. Using a decision tree (J48) classifier, we detected 19 profiles in Academia.edu, 23 profiles in AnyBeat, and 283 profiles in Google+.

The list of "suspected" profiles was combined with a list of an equal size of randomly selected profiles designed to act as a control group. The only limitation on the random profiles selection was that they must have a minimum number of friends (the exact number depended on the size of the network). In Academia.edu, each chosen random profile had to have at least 21 friends, in Anybeat the number was six, and in Google+ the number was one. The ordering of the lists was created using a random numbers generator and each user was evaluated by three of the experts mentioned above.

During the evaluation, we discovered that some of the profiles which were flagged by our algorithm had already been removed by the social network administrator. One such example was found in Academia.edu, where a user named "Bilbo Baggins" from Oxford University was removed prior to the evaluation. These profiles were not considered as successful detection of fake profiles, despite the high probability of them being so.

In the end of the evaluation process, the experts evaluated all 172 profiles, where each profile received a score for the

[12]Due to a minor off-by-one error, we only used 499 negative examples in AnyBeat and 2,999 negative examples in Academia.edu.

three questions presented in section IV-C. For every "yes" answer, the profile received one point. If the expert answered "maybe" the profile received 0.5 point, and "no" answers did not returned any points. The final profile score was the sum of the experts' responses.

The results were evaluated by comparing the number of profiles in each group (flagged and control) that received a "score" greater or equal to 1.5 (meaning that the majority of experts declared them as illegitimate). We will now go over each social network and describe the performance of the proposed method:

Acadeima.edu. The J48 decision tree classifier had indicated that 21 profiles had a high probability of being fake. Some of these profiles had been removed from the social network before the evaluation began, which left us with 15 valid profiles. The profiles indicated by the decision tree classifier received an average score of 1, while the profiles in the random group received an average score of 0.166. Moreover, 7 (46.6%) of the 15 flagged profiles received a score equal or higher than 1.5 points, compared to 0 in the control group.

AnyBeat. The J48 decision tree classifier flagged 23 profiles as having a high probability of being fake. One of these profiles has already been removed from the network, which left us with 22 profiles to analyze. The experts found that 7 (31.8%) of the 22 profiles received a score equal or higher than 1.5 points, compared to only 4 (20%) of the profiles in the control group. Moreover, 14 (63.6%) of the profiles in the group indicated by the J48 classifier were following other users of a specific gender, compared to only 7 (35%) in the control group.

Google+. In this network, we evaluated the performance of our algorithm on the top fifty flagged results. Three of these flagged profiles had already been removed or blocked before the evaluation began, leaving us with forty eight profiles to analyze. Of these profiles, 17 (35.4%) received a score higher or equal to 1.5 points, compared to only 10 (20.4%) of the control group. In addition, the experts concluded that 16 (33.3%) of the 48 flagged profiles may be spammers, compared to only 4 (8.1%) in the control group. With regard to the final research question, users who only follow users of a specific gender, the results were 2 (4.1%) and 0 for the flagged users and the control group, respectively.

VI. Discussion

Based on the experts' classifications, we can conclude that the proposed algorithm performs very well overall. However, the performance varies with each network (see Table III and Figures 3 and Figures 3-6). The differences are due to the special characteristics of each of the social networks and their users. In Academia.edu, 46.6% of the 15 profiles flagged by our algorithm were not "legitimate", while none of the profiles in the control group were flagged as such. In Google+, which has a medium anonymity level, we discovered that 35.4% in the flagged group and 20.4% of the control group were suspected of being fake. Moreover, almost 33.3% of the profiles returned by our algorithm were considered to be spammers, compared to only 8.1% in the control group.

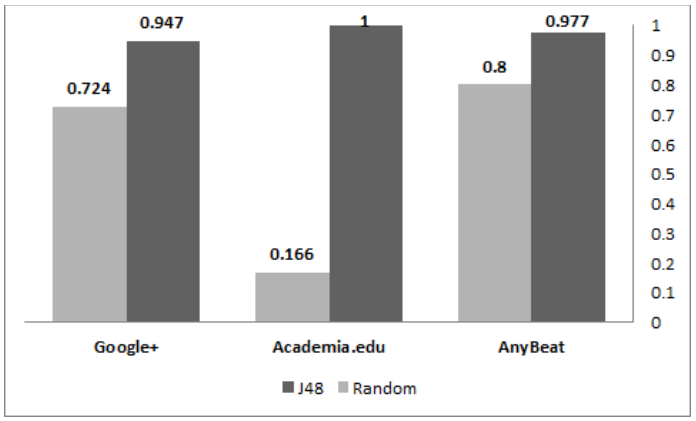

Fig. 3. Average Fake Profiles' Scores

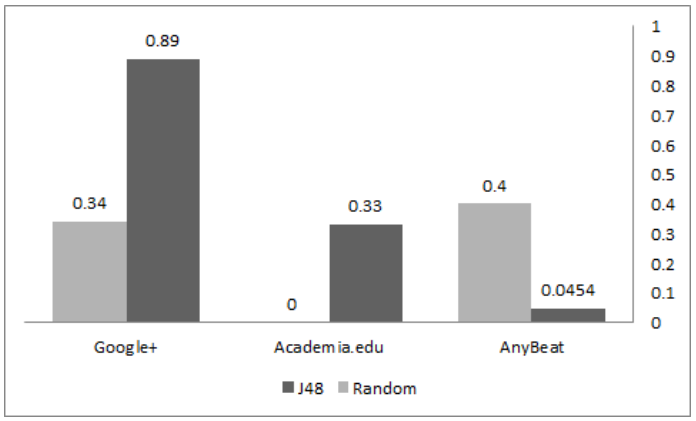

Fig. 4. Average Spam Profiles' Scores.

In AnyBeat, which is a relatively new network with a high level of anonymity, our algorithm succeeded in detecting fake profiles in 31.8% of the flagged profiles, while the control group contained only 20%. In this network, our algorithm was also successful in detecting users who were interested in a specific gender. However, in this network, our algorithm failed in detecting spammer profiles. We believe this can be attributed to the fact that the network is only a few months old. Because of its "young" age, the network has a small number of users and is therefore not a target for spammers. Other important issues are the high anonymity level and the type of the network. In AnyBeat users are encouraged to meet new people. In a sense, the users of this social network are encouraged to behave somewhat like socialbots, a fact that makes the detection task more difficult.

Another interesting and disturbing discovery is the high percentages of fake and spammer profiles in the various social networks. This is a strong indication of how widespread socialbots and spammers have become. This also highlights the urgent need for solution.

TABLE II
CLASSIFIERS 10 FOLDS CROSS VALIDATION RESULTS

Social Network	Classifier	False Positive	F-Measure	AUC
Academia.edu	J48	0.052	0.967	0.983
Academia.edu	Naive Bayes	0.063	0.995	0.999
AnyBeat	J48	0.026	0.99	0.992
AnyBeat	Naive Bayes	0.126	0.968	0.982
Google+	J48	0.01	0.999	0.995
Google+	Naive Bayes	0.01	0.993	1

TABLE III
THE SUMMARY OF THE EXPERTS' RESULTS FOR EACH SOCIAL NETWORK

Social Network	Question	Group	#Profiles	Scores Sum	Profiles with score $\geq$1.5
AnyBeat	Is Profile Fake?	J48	22	21.5	7
AnyBeat	Is Profile Fake?	Random	20	16	4
AnyBeat	Is Profile Spammer?	J48	22	1	0
AnyBeat	Is Profile Spammer?	Random	20	8	2
AnyBeat	Is Following Specific Gender?	J48	22	28	**14**
AnyBeat	Is Following Specific Gender?	Random	20	16.5	**7**
Academia.edu	Is Profile Fake?	J48	15	15	**7**
Academia.edu	Is Profile Fake?	Random	18	3	**0**
Academia.edu	Is Profile Spammer?	J48	15	5	2
Academia.edu	Is Profile Spammer?	Random	18	0	0
Google+	Is Profile Fake?	J48	48	45.5	**17**
Google+	Is Profile Fake?	Random	49	35.5	**10**
Google+	Is Profile Spammer?	J48	48	**43**	**16**
Google+	Is Profile Spammer?	Random	49	**17**	**4**
Google+	Is Following Specific Gender?	J48	48	18.5	2
Google+	Is Following Specific Gender?	Random	49	14.5	0

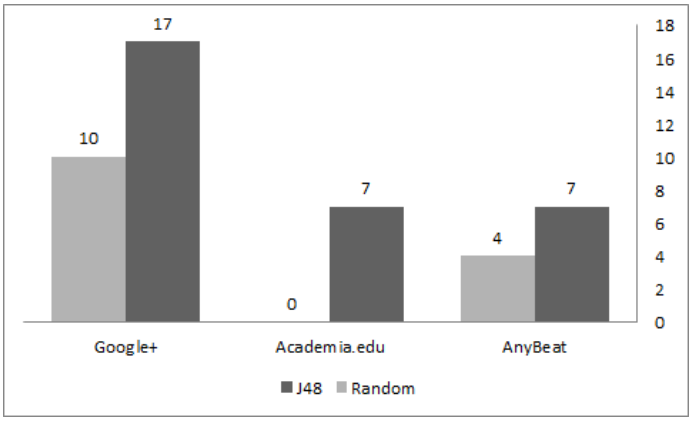

Fig. 5. Number of Profiles with High Fake Score (greater than or equal to 1.5 points).

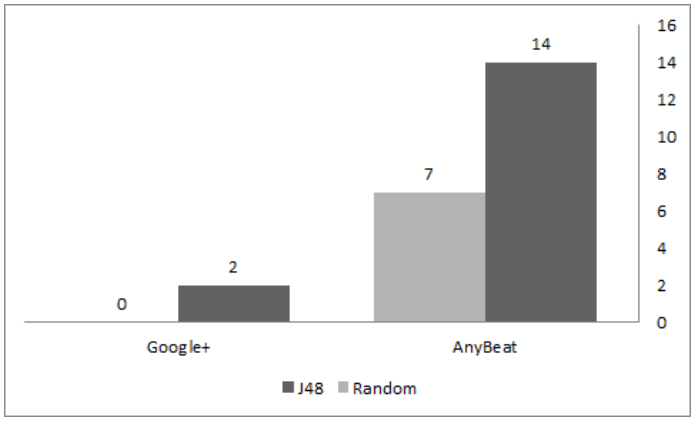

Fig. 6. Number of Profiles with High "Following Same Gender" Score (greater than or equal to 1.5 points).

VII. CONCLUSIONS AND FUTURE WORK

When we were children, our parents warned us not to talk to strangers. The streets may have become virtual, but the dangers remain the same. In this study, we tried to offer a solution to some of these pertinent threats by presenting an algorithm capable of detecting these "strangers in street". Our proposed algorithm uses a combination of graph theory algorithms and machine learning in order to detect these types of users and does so by only using the graph topology structure.

The proposed algorithm was tested on three different directed online social networks, each with a different level of anonymity. For each social network, our proposed method performed well when evaluated on both real and simulated profiles. A small number of profiles with a high probability of being fake or "spammer" profiles were extracted and analyzed. The analysis was performed by a team of experts with diverse backgrounds. According to the experts, evaluating a user's profile authenticity is a difficult task due to the fact that many fake profiles go to great lengths to appear legitimate. Only a thorough analysis of the few available details can reveal such a deception. For example, during the course of this research, we encountered a profile which appeared in several social

networks, was very active, and had many friends. Only through the use of "photo watermarks" were we able to uncover the fact that the picture actually belongs to a different user in a different country. In many ways, the identification of fake profiles can be described as a version of the Turing Test [44]. Our tested algorithms demonstrated false-positive rates of 0.01 to 0.052 on the different networks. These false-positive rates can provide sufficient protection for small and medium size online social networks with several millions of users. However, these false-positive rates are not low enough for large scale network, like Facebook, which has more than 901 million registered users. To protect networks of this magnitude, the proposed detection method will need to be extended by adding other features, possibly such as users activity features.

We intend to continue and develop our algorithm and others in order to better tackle this important problem. On primary research direction is the expansion and adaptation of this algorithm for additional types of networks. Possible types include (but are not limited to) emails, social networks, and mobile phones calls. Another research direction we intend to pursue is the integration of our algorithm together with content based methods and by this decrease the model's false-positive rates. We believe that using the user created content, such as private messages and wall posting, can be of great value when attempting to detect fake profiles and spammers

VIII. Availability

Anonymous version of Academia.edu, AnyBeat and Google+ social networks topologies are available for other researchers to use on our research group website http://proj.ise.bgu.ac.il/sns/.

References

[1] Facebook-Newsroom, http://www.facebook.com.

[2] A. Acquisti and R. Gross, *Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook*, 2006.

[3] 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.

[4] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: automated identity theft attacks on social networks," in *Proceedings of the 18th international conference on World wide web*. ACM, 2009, pp. 551–560.

[5] G. Kontaxis, I. Polakis, S. Ioannidis, and E. Markatos, "Detecting social network profile cloning," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 295–300.

[6] G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A practical attack to de-anonymize social network users," in *Security and Privacy (SP), 2010 IEEE Symposium on*. IEEE, 2010, pp. 223–238.

[7] A. Mislove, B. Viswanath, K. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 251–260.

[8] J. Baltazar, J. Costoya, and R. Flores, "The real face of koobface: The largest web 2.0 botnet explained," *Trend Micro Threat Research*, 2009.

[9] G. Rydstedt, E. Bursztein, D. Boneh, and C. Jackson, "Busting frame busting: a study of clickjacking vulnerabilities on popular sites a survey of frame busting," *Web 20 Security and Privacy 2010*, pp. 1–13, 2010.

[10] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Communications of the ACM*, vol. 50, no. 10, pp. 94–100, 2007.

[11] J. Douceur, "The sybil attack," *Peer-to-peer Systems*, pp. 251–260, 2002.

[12] D. Irani, M. Balduzzi, D. Balzarotti, E. Kirda, and C. Pu, "Reverse social engineering attacks in online social networks," *Detection of Intrusions and Malware, and Vulnerability Assessment*, pp. 55–74, 2011.

[13] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *Proceedings of the 26th Annual Computer Security Applications Conference*. ACM, 2010, pp. 1–9.

[14] S. Nelson, J. Simek, and J. Foltin, "The legal implications of social networking," *Regent UL Rev.*, vol. 22, pp. 1–481, 2009.

[15] J. Wolak, D. Finkelhor, K. Mitchell, and M. Ybarra, "Online predators and their victims," *Psychology of Violence*, vol. 1, pp. 13–35, 2010.

[16] M. Ybarra and K. Mitchell, "How risky are social networking sites? a comparison of places online where youth sexual solicitation and harassment occurs," *Pediatrics*, vol. 121, no. 2, p. e350, 2008.

[17] Facebook, http://www.sec.gov/Archives/edgar/data/1326801/000119312512101422/d287954ds1a.htm#toc287954_2.

[18] A. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, p. 509, 1999.

[19] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, p. 7821, 2002.

[20] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots+ machine learning," in *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2010, pp. 435–442.

[21] Alexa, http://www.alexa.com/.

[22] Nielsen, http://blog.nielsen.com/nielsenwire/online_mobile/august-2011-top-us-web-brands.

[23] F. Nagle and L. Singh, "Can friends be trusted? exploring privacy in online social networks," in *Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in*. IEEE, 2009, pp. 312–315.

[24] J. Lindamood, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Inferring private information using social network data," in *Proceedings of the 18th international conference on World wide web*. ACM, 2009, pp. 1145–1146.

[25] Y. Liu, K. Gummadi, B. Krishnamurthy, and A. Mislove, "Analyzing facebook privacy settings: User expectations vs. reality," 2011.

[26] S. Mahmood and Y. Desmedt, "Poster: preliminary analysis of google+'s privacy," in *Proceedings of the 18th ACM conference on Computer and communications security*. ACM, 2011, pp. 809–812.

[27] Facebook, "Report abuse or policy violations."

[28] S. Axon, "Facebook will add a panic button for uk teens."

[29] Checkpoint, http://www.zonealarm.com/security/en-us/zonealarm-socialguard-facebook-parental-control.htm.

[30] D. DeBarr and H. Wechsler, "Using social network analysis for spam detection," *Advances in Social Computing*, pp. 62–69, 2010.

[31] A. Wang, "Don't follow me: Spam detection in twitter," in *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*. IEEE, 2010, pp. 1–10.

[32] M. Bosma, E. Meij, and W. Weerkamp, "A framework for unsupervised spam detection in social networking sites."

[33] B. Levine, C. Shields, and N. Margolin, "A survey of solutions to the sybil attack," *University of Massachusetts Amherst, Amherst, MA*, 2006.

[34] D. Quercia and S. Hailes, "Sybil attacks against mobile users: friends and foes to the rescue," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–5.

[35] H. Yu, P. Gibbons, M. Kaminsky, and F. Xiao, "Sybillimit: A near-optimal social network defense against sybil attacks," *IEEE/ACM Transactions on Networking (ToN)*, vol. 18, no. 3, pp. 885–898, 2010.

[36] G. Danezis and P. Mittal, "Sybilinfer: Detecting sybil nodes using social networks." NDSS, 2009.

[37] H. Yu, "Sybil defenses via social networks: a tutorial and survey," *ACM SIGACT News*, vol. 42, no. 3, pp. 80–101, 2011.

[38] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, 2008.

[39] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American society for information science and technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[40] M. Fire, L. Tenenboim, O. Lesser, R. Puzis, L. Rokach, and Y. Elovici, "Link prediction in social networks using computationally efficient topological features," in *Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third*

International Confernece on Social Computing (SocialCom). IEEE, 2011, pp. 73–80.
[41] AnyBeat, "Anybeat about," http://www.anybeat.com/about.
[42] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
[43] Google, "Google search by image," http://www.google.com/insidesearch/searchbyimage.html.
[44] A. Pinar Saygin, I. Cicekli, and V. Akman, "Turing test: 50 years later," *Minds and Machines*, vol. 10, no. 4, pp. 463–518, 2000.