Early malicious activity discovery in microblogs by social bridges detection

Antonia Gogoglou[1], Zenonas Theodosiou[2], Tasos Kounoudes[2], Athena Vakali[1] and Yannis Manolopoulos[1]
[1]Department of Informatics
Aristotle University of Thessaloniki, Thessaloniki, Greece
Email: {agogoglou, avakali, manolopo}@csd.auth.gr
[2]SignalGeneriX, Cyprus
Email: {z.theodosiou, tasos}@signalgenerix.com

***Abstract*—With the emerging and intense use of Online Social Networks (OSNs) amongst young children and teenagers (youngsters), safe networking and socializing on the Web has faced extensive scrutiny. Content and interactions which are considered safe for adult OSN users, might embed potentially threatening and malicious information when it comes to underage users. This work is motivated by the strong need to safeguard youngsters OSNs experience such that they can be empowered and aware. The topology of a graph is studied towards detecting the so called social bridges, i.e. the group(s) of malicious users and their supporters, who have links and ties to both honest and malicious user communities. A graph-topology based classification scheme is proposed to detect such bridge linkages which are suspicious for threatening youngsters networking vulnerability. The proposed scheme is validated by a Twitter network, at which potentially dangerous users are identified based on their Twitter connections. The achieved performance is higher compared to previous efforts, despite the increased complexity due to the variety of groups identified as malicious.**

I. Introduction

Malicious behavior on the Web has emerged in various internet applications including, but not limited to, email services, shopping and recommendation platforms, crowdsourcing websites, mashups and OSNs. Such behavior has heavily impacted popular and widely used OSN platforms and applications, since they are open and easily accessible large crowds forums, forming structures such as the social graph [1]. Therefore, social networks constitute a breeding ground for the spread of malicious behavioral patterns, such as spamming, link farming, Sybil attacks (forged profile identities), phishing and the even more dangerous pedophile attacks, online grooming, etc [2]. In this direction, the social network providers (Twitter, Instagram, Facebook, Flickr, YouTube, etc.), the authorities, as well as the scientific community, are invested in analyzing social media data and identifying or even predicting the aforementioned behavioral patterns. In order to perform this analysis, data from web-based communities and user generated content needs to be utilized, such as connections from social-networking sites, video sharing sites, blogs, folksonomies, etc.

In the context of the present article, we conducted an empirical analysis of the social dynamics of spam accounts in OSNs and the ways they form connections with the rest of the network in order to reach the *honest* users. The concept of spamming in OSNs and the ways to identify it have been extensively studied with approaches including automatic dissemination of spam like [3], [4], tools used by spammers to deceive search engines [5] or faking honest behaviors [6]. Although these approaches are efficient and their prediction results seem promising, they do not attempt to identify all potentially dangerous users in real world networks. As it is often the case, spammers manage to mimic honest users' behavior and, by connecting with them, they penetrate the strongly connected component of a network making it challenging to identify them.

More specifically, we contemplated the social behavior these spam users display, in order to increase their impact. We proceeded to expand the concept of *dangerous or malicious* users in OSNs, beyond the obvious spam accounts, to facilitate the needs of more sensitive OSN users, such as young adolescents and children. *A motivating scenario* would be a young child that makes a new connection in an OSN with a user that appears to be connected with other children from the same school or neighbourhood. If this new user has not explicitly shared malicious content online, conventional detection systems would not provide an alert for this new connection and that might be justifiable for adult users. But for a kid this might not suffice to protect them from exposure to inappropriate groups of users. Should this new connection have links to spammers or generally malicious users, the child could be exposed to other far more dangerous new connections by entering a part of the network with criminal communities. A young user could also be faced with inappropriate shared content that is being spread in this part of the network. In this article we refer to the users that help link spammers to the core of the network as *social bridges*.

Various groups of users tend to follow criminal accounts (e.g. spam accounts), and they display certain identifiable behavioral patterns. In [7] criminal hubs and criminal leaves were identified as users that follow a large number of criminal accounts and the ones that have limited connections to the criminal communities respectively. A further categorization of the extracted criminal hubs was conducted by dividing them into social butterflies, social promoters and dummies. Each of these categories has different motives for following criminal accounts, either knowingly or not, and could prove being dangerous themselves. The spam-neighbourhood has

been contemplated in [8] and the spam followers have been found to be increasingly influential nodes in a network.

In this study, we manage to address the aforementioned issues through:

- The design of an expanded detection framework for malicious users that identifies both spammers and their *social bridges* to the rest of the network. The implementation of this classification framework relying on the topology of the social graph. This information is already available in most OSNs, as opposed to private information or shared content that might not be accessible to a user, until the new connection is added. In this way alerts can occur in advance, when a new contact is being added by an underage user.
- The addition of the *k-shell decomposition* concept to better analyze the topological behavior of spammers in a network and identify their *social bridges*. Moreover, we utilized the *k-core* numbers as a distinguishing feature to yield increased accuracy for our detection framework.
- The combination of sampling and *cost-aware methodologies* to facilitate the classification of malicious and innocent users. This leads to enhanced performance of our framework, as compared to using them independently.

The rest of the article is organized as follows: Section II summarizes the literature overview; Section III introduces the dataset utilized in this work. Section IV presents the analysis of our dataset and the observations made on the topological features of various user categories, while Section V describes the experiments conducted and their results. Section VI concludes the article.

II. Related Work

There are three basic approaches to the study of malicious behavior in OSNs: i) focusing on link analysis (URLs, clickstreams, etc.) [9] ii) focusing on content mining (hash tag mining, comments or status semantics analysis, image processing etc.) [10] iii) focusing on networks features (centrality, connectivity, degrees, community detection, shortest path, small world properties, etc.) [11]. Each of these approaches mines different categories of data crawled from online platforms in an attempt to extract valuable insights on the interrelations of the social graph. Link based methodologies often produce misleading results, as it is fairly common for honest users to be redirected to malicious links, which hinders the distinguishing of actual spammers. In [12], [13] systems that rely on content information present promising results in detecting spam, cyberbullying and aggressive behaviors by utilizing either text from posts and comments. Such systems are often valuable from the OSN providers' side to distinguish users distributing malicious content, but on the user's side content information about other users is often unavailable before connecting with them.

An attempt to combine content and network analytics to assign to each user a probability of engaging in cyberbullying was conducted in [14], [15]. The combined features proved to be better performing than using social or textual features alone. However, the achieved accuracy was only slightly higher than the one achieved with the social features alone, meaning that the addition of text mining increased the overhead and computational cost rather than considerably improving the resulting accuracy. An analogous approach was adopted in [16] where user demographics, network features and sentiment analysis of posts were combined to identify connections sharing bullying content in MySpace threads.

Network oriented methodologies present the advantage of requiring limited information about the user including only their connections in an OSN. This information is often publicly available, and the status of a user (dangerous or not) can be assessed based on their network position. A number of studies [17], [18], [19] utilize the community structure and the topology of the OSN graphs to create user profiles, according to their network metrics, such as in-degree, out-degree, centralities, community memberships, etc. and common features they share with other users (personal information, activities, shared groups and posts, etc.).

In [20] user profiles were extracted using the degrees of each user in a Facebook network, weighted by the common attributes shared between connected nodes. It was inferred in this study that a user can to a great extent be characterized by his/her connections in an OSN and even profile features, such as educational background, profession, etc., may be predicted from their neighbours' features. In an analogous attempt, Fire et al [21] developed the Social Protector, which works from a user's side to evaluate a Facebook user's connections utilizing the users' shared attributes. However, this system assesses only the existing friend's network of a user and does not provide any indication whether a new connection could be dangerous or not. Bhat et al. [11] focused on detecting spammers by grouping users into communities and studying the differences in their community-based network features as compared to the ones of honest users. The spam behavior was simulated using Random Link Attack (RLA) model was conducted in a Facebook dataset.

Studies that utilize simulated spam behavior may fail to caption the real world dynamic process according to which malicious users attach to communities of honest (or innocent) users. The notion of *maliciousness* of a user is rather vague, thus it is difficult to simulate effectively. Not only a user sharing malicious content should be considered as a dangerous connection. Various attempts [22], [23] have been made to define what may constitute a malicious user; Athanassopoulos et al. [22] extensively contemplated different scenarios of attacks that can be performed using Facebook, including but not limited to fake and compromised real user accounts, attack campaigns such as social spam, malware distribution, and online rating distortion. More serious offenses in OSNs that include spreading of pedophilic content, online grooming, sexual harassment, cyberbulllying, etc are analyzed in [24]. As a result, unified and broad approaches need to be adopted to identify any potential danger to the various groups of OSN users.

III. Dataset Formulation

For the purposes of our research a subset of an OSN graph needed to be formulated built around the connections of a groups of identified malicious users. We opted to experiment with the Twitter network in particular, due to its openness, accessibility and popularity amongst youngsters; in the USA 42% of the teenagers between 15-17 actively use Twitter, whilst the popularity in younger kids has risen also reaching 21% of 13-14 year old kids using Twitter in 2015 [1]. a subset of an OSN graph needed to be formulated containing the groups in Figure 1. The Twitter dataset referred in [25] [8] was utilized to formulate our dataset. The original dataset contains over 50 million users and 2 billion links between those users as well as a set of more than 40,000 identified spammer accounts, which have been officially suspended by Twitter. Details of the dataset can be found here [2]. From this huge part of the Twitter network, we sampled a graph starting from 500 spammer users and extracting their connections according to the relationships depicted in Figure 1. This process led to a graph containing 303,999 unique users and a total of 1,002,316 links between them. The graph constructed is directed, based on the ”following” relationship of Twitter, and unweighted, as there is no natural measure of relationship strength in Twitter followers.

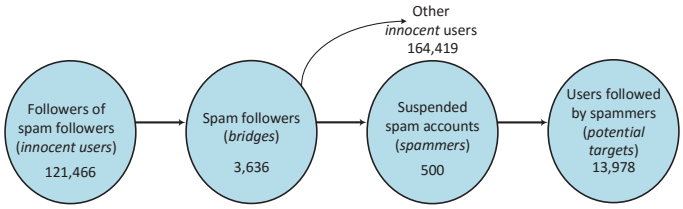

Fig. 1. Groups of users and their respective sizes in the Twitter graph built around a group of spammers.

IV. Social Bridges Detection by Social Graph Analysis

Firstly, in order to identify the social dynamics of spammers and their followers we have isolated these groups and their followers. We have extracted the connected components present in this part of the network using the Tarjan algorithm [26]. Then, we calculated the percentages of spammers and their associated connections that belong to each component. As can be seen in Figure 2, there are three identified components presented with *black* , *red* and *green* nodes. The majority (92%) of the least populated component (*black*) is comprised of spam users, while the *red* component contains 87% spam followers and the biggest of the components is mainly (96%) populated with honest users whom the other two components follow. The center of the graph represents the largest connected component in this subset of the Twitter graph and, as can be seen in Figure 2, the seemingly unpopular (disconnected) spam users (depicted on the bottom right of the figure with *black*) use their connection to the second connected component (*red*) to penetrate the dense center of the network (*green*). Therefore, these spam followers, constituting the majority of the second component, can be considered as a dangerous influence especially for the most vulnerable and impressionable users of the Twitter network (i.e. children) and constitute the *social bridges* of spammers. The group of spammers and social bridges will be referred to as *malicious* users.

Figure 3 depicts a subsample of spammers and the users they follow but without the social bridges. Consequently, the spam nodes become severly disconnected from the major connected component. It can be detected that, when the bridge users are removed from the spammers' connections they lose their access to the core of the network, meaning that the users they manage to connect to are not central influential nodes. As a result, ir becomes challenging for them to increase their connections and their impact. This observation indicates the dangers innocent users face when connecting with the social bridges; they become part of the expansive network of malicious users increasing the probability of connecting with the criminals (spammers) themselves.

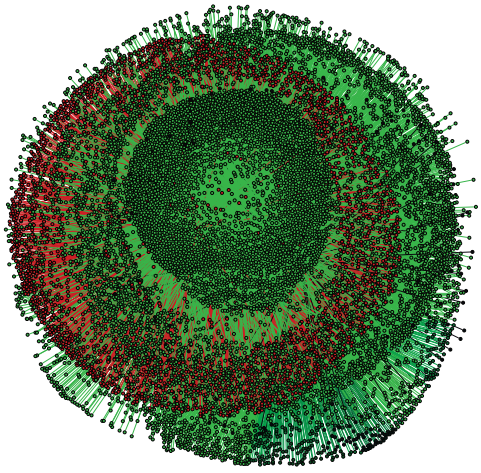

Fig. 2. A subsample of our Twitter graph including the followers of spammers (SF) and the SF's followers. Three different components were identified depicted in black, red and green.

To further analyze the different topological positioning of the malicious group in the Twitter graph as compared to the honest users, we have utilized a set of widely used network features [27]:

In-Degree defined as the number of incoming connections (followers) a user has.

Out-Degree defined as the number of users a node follows (outgoing connections).

Betweeness centrality which is equal to the number of shortest paths from all users to all others that pass through that specific node (i.e. user). It is a metric indicative of a user's influence in a network. As previous studies have indicated, we have confirmed that the bridge users are highly influential nodes in a graph ranking high in betweeness centrality values.

[1] http://www.statista.com/statistics/184307/usage-of-twitter-among-us-teenagers-by-age-group/

[2] http://socialnetworks.mpi-sws.org/datasets.html

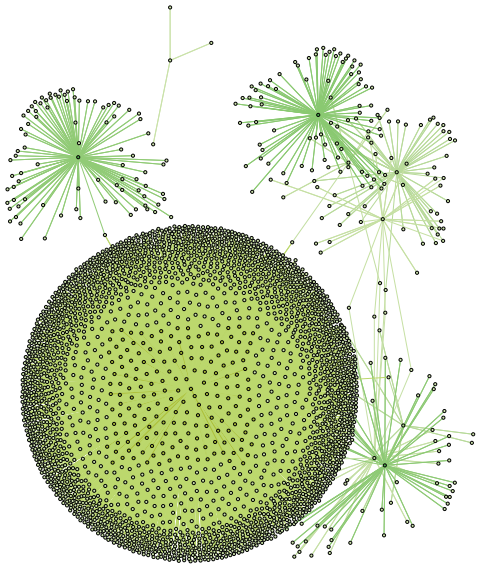

Fig. 3. A subset of our Twitter graph including a set of spammer nodes (the isolated nodes) and the users they follow, excluding the social bridges. The major connected component is depicted in the center.

Closeness centrality which is the mean distance from a vertex to other vertices. For our Twitter graph, as it contained a number of disconnected nodes, we utilized the harmonic mean to calculate representative values for the closeness centrality. Nodes with a low value in this metric might have better access to information than other nodes or more direct influence on other users of the network.

Eigenvector centrality is an extension of in-degree centrality that awards higher importance to links coming from more relevant nodes. In other words, a node is important (high eigenvector value) if it is linked to other important nodes.

k-core number [28] defined as the largest integer *k* for a node such that this node exists in a graph where all vertices have degree $>= k$. As it is often the case, the nodes belonging to the highest k-core (k_{max}) comprise a well connected globally distributed subset of the network, identified as the *nucleus* in an analogous study [29] on linkage between webpages. In the case of Twitter users, the k_{max} core is comprised of 72% malicious users and 28% of honest users. This is a surprising finding indicating that particularly the social bridges are often well connected users that can influence a large part of the network. This justifies the spammers tendency to attach to them in order to approach the majority of honest users. The largest connected component of the $k_{max} - 1$ shell (the second largest k-core) constitutes the *peer-component* (as named in [29]) that is the most well-connected component of the majority of users that remains connected even when we remove the k_{max} group. The 88% of the peer-component is comprised of honest users and the remaining percentage represents the malicious users. The rest of the graph contains low-connected users that would become entirely disconnected, if the peer-component and k_{max} were removed.

In Figure 4 we have summarized the percentage differences of the network features discussed above for the three identified user categories; the differences are calculated as:

$$DF(\%) = \frac{mean(group1) - mean(group2)}{mean(group2)} \quad (1)$$

We observe that the values for the spammer users present lower values of in-degree and closeness centrality compared to honest users. The social bridges display all the characteristics of influential users with high centralities, as discussed above. As it is often the case, the social bridges display similarities with both honest users and spammers in their topological features.

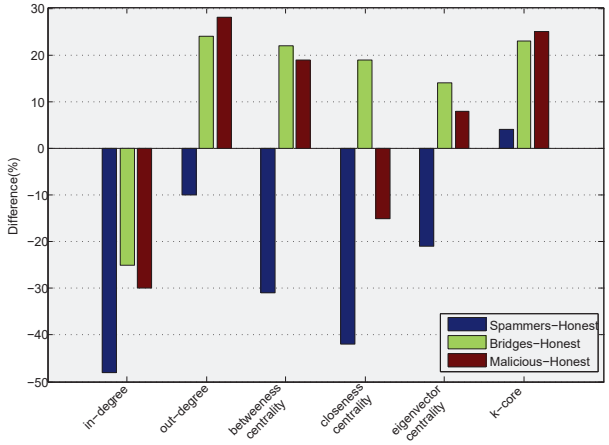

Fig. 4. Bar chart of the percentage differences for the 6 network features amongst the 3 identified user groups.

V. Experiments and Results

After investigation of the potential dangerous users in a Twitter network, we designed a framework for early identification of malicious users based on publicly available information, suitable for the protection of underage users in OSNs. As discussed in sections I and II, the predictive models relying on the social topology utilizing a user's connections are the most appropriate ones for alerting users beforehand about the dangers of making a new connection with another user. Consequently, we developed a classifier using the six network features discussed in the section IV for identifying two groups of users: malicious (spammers and social bridges) and non-malicious (honest) ones. The nature of the problem implies dealing with highly imbalanced classes, which can severely affect the performance of the classifier. To tackle this issue of skewed classes, we have chosen four different approaches and compared the performance of the resulting classifiers.

The first approach is based on 'SMOTE' (Synthetic Minority Oversampling TEchnique), which is widely used for skewed classification problems and has been applied specifically to spam detection [14]. SMOTE works by over sampling the minority class while under sampling the majority class to create balanced classes for the training dataset. That, however, could lead to potential overfitting due to the replication of the data points, thus this approach is completed with added synthetic data points following an analogous behavior in the feature space of the original training dataset. Another approach for skewed classification problems, is cost sensitive learning,

according to which the misplaced points belonging to the minority class are assigned a higher penalty than the ones of the majority class. One more approach was utilized as a comparison, which is based on rejection sampling and majority voting [30]. More specifically, cost proportionate rejection sampling from class c of class set C is applied instead of standard sampling in order to generate the appropriate training set. According to this method, each data point is independently included or not given the probability P, which is determined by the misclassification cost of the class and the maximum misclassification cost based on this formula:

$$P(c) = \frac{Cost(c)}{max[Cost(c) \forall c \in C]} \quad (2)$$

The resampling will be repeated a number of times and the results will be combined to increase the consistency and average performance of the classifier.

In our case, we have chosen 70% of our majority class (non malicious users) as our training set and that lead to 72,709 training data from the majority class. Using either the SMOTE approach or rejection sampling we have generated an analogous set of points from the 70% of the minority class (malicious users). The remaining 30% is used as test set, where the natural imbalance of the classes is preserved. For the cost sensitive learning the minority class was assigned a 10 times higher cost than the majority class, after performing a grid search for choosing the optimal cost parameter and weight. The 'SMOTE' approach and the cost sensitive learning were applied individually and also as a combination to identify the better performing model. In the case of applying the cost-sensitive learning approach individually, the sampling to form the training data was conducted randomly maintaining the natural imbalance of the classes.

The above mentioned approaches were combined with an SVM classifier and in particular the C-SVM classifier from libSVM [3], which is a widely used library for SVMs. We have opted for an SVM classifier, because it is well combined with sampling methodologies and the C-SVM in particular allows for efficient cost sensitive learning. The most popular metric for evaluating a classifier's performance is *accuracy*, but in our case of heavily skewed classes the misclassified items belonging to the two classes cannot be weighted equally. A basic majority voting classifier could yield an accuracy score of more than 95%; however, that score would not be representative of its distinguishing power in successfully identifying the malicious users, who constitute a small percentage of the total graph. Hence, we have opted for two different performance metrics, the *F-measure*, which is a combination of precision and recall, and the *sensitivity* or *true positive rate* for the minority class.

In addition, we perform a sensitivity analysis for our chosen features and we compare performances when excluding one feature at a time. The combinations of features and the methodologies we have applied are summarized on Table I. Columns correspond to the approaches for dealing with the imbalanced classes and rows to the combinations of features, that are excluded one at a time to explore their influence on the final result. The classification results for the various approaches are depicted in Table II according to the two aforementioned performance metrics. The best performing combinations for each approach appear in bold. For each method and combination of features we have run the classifier 10 times to avoid bias in the selected training and data set and we present the average for each of the two metrics.

[3]https://www.csie.ntu.edu.tw/ cjlin/libsvm/

TABLE I
COMBINATIONS OF FEATURES AND METHODS APPLIED IN THE EXPERIMENTS.

Features:	**1.**In-Degree	**2.**Out-Degree	**3.**Betweeness Centr.
	4.Closeness Centr.	**5.**Eigenvector Centr.	**6.**k-core numbers
SMOTE:	SMOTE sampling		
Cost:	Cost sensitive learning		
SMOTE&Cost:	SMOTE sampling with Cost-sensitive learning		
Probabilistic:	Probability based cost proportionate rejection sampling		

TABLE II
RESULTS IN TWO PERFORMANCE METRICS (F-MEASURES AND SENSITIVITY) FOR THE 4 APPROACHES EMPLOYED AND MULTIPLE COMBINATIONS OF FEATURES (EXCLUDING ONE FEATURE AT A TIME)

	SMOTE		Cost		SMOTE&Cost		Probabilistic	
	F	S	F	S	F	S	F	S
All features	**0.623**	**0.650**	**0.512**	**0.529**	**0.802**	**0.830**	**0.732**	**0.758**
All minus 1	0.478	0.494	0.304	0.333	0.651	0.682	0.503	0.538
All minus 2	0.380	0.412	0.392	0.403	0.690	0.703	0.545	0.581
All minus 3	0.567	0.591	0.497	0.514	0.790	0.808	0.688	0.727
All minus 4	0.377	0.394	0.279	0.264	0.619	0.633	0.511	0.542
All minus 5	0.542	0.578	0.476	0.470	0.711	0.730	0.612	0.644
All minus 6	0.512	0.523	0.420	0.456	0.700	0.724	0.598	0.632

As indicated by the results in Table II the optimal performance is achieved for the SMOTE&Cost approach. It appears that the appropriate sampling (i.e. SMOTE) is allowing for the biggest increase in performance, since in the Cost approach, where no sampling was applied and only different costs were assigned to each class, the lowest performance was achieved. In addition, the combination of cost sensitive learning and SMOTE provides better results than the Probabilistic approach, which employs a different sampling methodology and incorporates the cost-weights of the different classes in the sampling itself.

As far as the features are concerned, we detect that combining all of the network features, thereby fully leveraging the topological position of each user, yields the highest performing classifier in all four approaches. That is to be expected, as more detailed representations allow for more accurate classification results. Moreover, the total number of network features is only 6, thus not causing a very high-dimensional feature space that could lead to overfitting. The most important of the social features is the closeness centrality (feature 4), as removing this feature results in the lowest achieved performance. That is to be expected as for graphs containing disconnected nodes (isolated spammers or unpopular users followed by spammers) the adjusted closeness centrality is the most representative metric for the connectivity and topological position of a node. The addition of the *k-core* numbers as a feature also helps improve performance, as the core numbers are 27%

different in malicious users compared with honest ones (see section IV). Approaches that report analogous performance with ours, like [11], [14], use either simulated spam behavior or calculate performance metrics by averaging the results of both classes, which can yield favorable results. Consequently, direct comparisons cannot be performed. Our proposed combination of these 6 network features with SMOTE and cost sensitive learning yields better scores in performance metrics compared with existing approaches evaluated with analogous performance metrics on real-world datasets [15], which reach a maximum of 0.760 in sensitivity and F-measure.

VI. Conclusion

We have designed a classification based framework using social network features to identify spam users and their *social bridges*, whom they utilize to access the connected components of the Twitter graph. The different behavioral patterns of these two categories of users that pose dangers to particularly vulnerable groups of users (such as children) are explored and automatically predicted to allow for alerts to occur when a new connection is added. In the future, we plan to expand the features used to include semantic or textual information and apply our best performing classification scheme to other OSNs, as well.

Acknowledgment

The work presented in this paper is a result of ENCASE project. This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 691025

References

[1] D. Yin, Z. Xue, L. Hong, B. Davison, A. Kontostathis, and L. Edwards, "Detection of harassment on web 2.0," in *Proceedings of the Content Analysis in the WEB 2.0 (CAW2.0) Workshop at WWW2009*, 2009.

[2] M. Yip, N. Shadbolt, and C. Webber, "Structural analysis of online criminal social networks," in *IEEE Conference on Intelligence and Security Informatics (ISI)*, 2012, pp. 60–65.

[3] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: The underground on 140 characters or less," in *Proceedings of the 17th ACM Conference on Computer and Communications Security*, 2010, pp. 27–37.

[4] Z. Xianghan, Z. Zhipeng, C. Zheyi, Y. Yuanlong, and R. Chunming, "Detecting spammers on social networks," *Neurocomputing*, vol. 159, pp. 27 – 34, 2015.

[5] R. Heartfield and G. Loukas, "A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks," *ACM Comput. Surv.*, vol. 48, no. 3, pp. 37:1–37:39, 2015.

[6] 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*, ser. IMC '11, 2011, pp. 243–258.

[7] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, "Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on twitter," in *Proceedings of the 21st International Conference on World Wide Web*, ser. WWW '12, 2012, pp. 71–80.

[8] S. Ghosh, B. Viswanath, F. Kooti, N. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. Gummadi, "Understanding and combating link farming in the twitter social network," in *Proceedings of the 21st International Conference on World Wide Web*, ser. WWW '12, 2012, pp. 61–70.

[9] J. Wei, Z. Shen, N. Sundaresan, and K. Ma, "Visual cluster exploration of web clickstream data," in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2012, pp. 3–12.

[10] M. Kandias, V. Stavrou, N. Bozovic, L. Mitrou, and D. Gritzalis, "Can we trust this user? predicting insider's attitude via youtube usage profiling," in *Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC)*, 2013, pp. 3–12.

[11] S. Bhat and M. Abulaish, "Community-based features for identifying spammers in online social networks," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2013, pp. 100–107.

[12] H. Hosseinmardi, S. Li, Z. Yang, Q. Lv, R. Rahin, R. Han, and S. Mishra, "A comparison of common users across instagram and ask.fm to better understand cyberbullying," in *BDCloud*. IEEE, 2014, pp. 355–362.

[13] M. Giatsoglou, D. Chatzakou, N. Shah, C. Faloutsos, and A. Vakali, "Retweeting activity on twitter: Signs of deception," in *Advances in Knowledge Discovery and Data Mining*. Springer International Publishing, 2015, pp. 122–134.

[14] M. A., K. D., and S. R., "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the twitter network," *Computers in Human Behavior*, vol. 63, pp. 433 – 443, 2016.

[15] Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber bullying detection using social and textual analysis," in *Proceedings of the 3rd International Workshop on Socially-Aware Multimedia*, ser. SAM '14. ACM, 2014, pp. 3–6.

[16] A. Squicciarini, S. Rajtmajer, Y. Liu, and C. Griffin, "Identification and characterization of cyberbullying dynamics in an online social network," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ser. ASONAM '15, 2015, pp. 280–285.

[17] M. Fire, G. Katz, and Y. Elovici, "Strangers intrusion detection-detecting spammers and fake proles in social networks based on topology anomalies," *Human Journal*, vol. 1, pp. 26–39, 2012.

[18] J. L. C. X. Jin, X.and Lin and J. Han, "Socialspamguard: A data mining-based spam detection system for social media networks," in *PVLDB*, vol. 12, 2011, pp. 1458–1461.

[19] D. DeBarr and H. Wechsler, "Using social network analysis for spam detection," in *Proceedings of the Third International Conference on Social Computing, Behavioral Modeling, and Prediction*, 2010, pp. 62–69.

[20] A. Mislove, B. Viswanath, K. P. 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*, ser. WSDM '10, 2010, pp. 251–260.

[21] M. Fire, D. Kagan, A. Elyashar, and Y. Elovici, "Friend or foe? fake profile identification in online social networks," *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 1–23, 2014.

[22] E. Athanasopoulos, A. Makridakis, S. Antonatos, D. Antoniades, S. Ioannidis, K. G. Anagnostakis, and E. Markatos, *Antisocial Networks: Turning a Social Network into a Botnet*, 2008, pp. 146–160.

[23] Q. Cao, X. Yang, J. Yu, and C. Palow, "Uncovering large groups of active malicious accounts in online social networks," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 477–488.

[24] S. Wachs, G. K. Jiskrova, A. T. Vazsonyi, K. D. Wolf, and M. Junger, "A cross-national study of direct and indirect effects of cyberbullying on cybergrooming victimization via self-esteem," *Psicologa Educativa*, vol. 22, no. 1, pp. 61 – 70, 2016.

[25] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, Washington DC, USA, May 2010.

[26] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM Journal on Computing*, vol. 1, pp. 146–160, 1972.

[27] E. Estrada, *The structure of complex networks: theory and applications*. Oxford University Press, 2012.

[28] S. B. Seidman, "Network structure and minimum degree," *Social Networks*, vol. 5, no. 3, pp. 269 – 287, 1983.

[29] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, "A model of internet topology using k-shell decomposition," *Proceedings of the National Academy of Sciences*, vol. 104, no. 27, pp. 11 150–11 154, 2007.

[30] A. Cohen, "An effective general purpose approach for automated biomedical document classification," in *AMIA Annual Symposium Proceedings*, 2006, pp. 161–165.