An MCL-Based Approach for Spam Profile Detection in Online Social Networks

Faraz Ahmed
Center of Excellence in Information Assurance
King Saud University, Riyadh, KSA
Email: fAhmad.c@ksu.edu.sa

Muhammad Abulaish, *SMIEEE*
Center of Excellence in Information Assurance
King Saud University, Riyadh, KSA
Email: mAbulaish@ksu.edu.sa

***Abstract*—Over the past few years, Online Social Networks (OSNs) have emerged as cheap and popular communication and information sharing media. Huge amount of information is being shared through popular OSN sites. This aspect of sharing information to a large number of individuals with ease has attracted social spammers to exploit the network of trust for spreading spam messages to promote personal blogs, advertisements, phishing, scam and so on. In this paper, we present a Markov Clustering (MCL) based approach for the detection of spam profiles on OSNs. Our study is based on a real dataset of Facebook profiles, which includes both benign and spam profiles. We model social network using a weighted graph in which profiles are represented as nodes and their interactions as edges. The weight of an edge, connecting a pair of user profiles, is calculated as a function of their real social interactions in terms of *active friends*, *page likes* and *shared URLs* within the network. MCL is applied on the weighted graph to generate different clusters containing different categories of profiles. Majority voting is applied to handle the cases in which a cluster contains both spam and normal profiles. Our experimental results show that majority voting not only reduces the number of clusters to a minimum, but also increases the performance values in terms of F_P and F_B measures from $F_P = 0.85$ and $F_B = 0.75$ to $F_P = 0.88$ and $F_B = 0.79$, respectively.**

***Index Terms*—Social network analysis, Cyber security, Social network security, Spam profile detection, Spam campaign identification.**

I. Introduction

Over the past few years, Online Social Networks (OSNs) have emerged as cheap and easily accessible social media, facilitating users around the globe for communication and information sharing. The users of these social networks are the key elements responsible for the content being shared. The social network users are the basic elements in the hierarchy of the OSNs, and the next level elements in the hierarchy are the communities formed by friends, families and acquaintances. Users share information by sharing links to interesting websites, videos and files. Moreover, the community structure of an OSN creates a network of trust and reliability. An individual shares personal information with his/her network of trust, and other users trust the information shared. A study done in [1] shows that 45% of users click on the links shared by their immediate contacts. This feature of sharing information to a large number of individuals with ease has attracted malicious parties, which also include social spammers. Social spammers exploit the network of trust for spreading spam messages promoting personal blogs, advertisements, phishing, and scams. Spammers employ different strategies for getting into a user's network of trust. Information sharing by the use of URL shortening service is an important feature of online social networking. This feature is easily exploited and is particularly harmful to users if it contains links to scam advertisements, adult contents and other solicitations, phishing attempts to capture account credentials, and pages attempting to distribute malwares. Therefore, the only security feature that protects a user from malicious parties is the network of his/her friends on the social network.

According to Symantec, globally 75.9% of email messages are spam [2]. Similarly, with social networks on the target, the current state of spam is worsening and more rigorous efforts are required to stop them. Nowadays, spammers are trying a new approach to gain access through Facebook events, in which users are invited to some events with fake titles like "*Check out who viewed your profile*!". Although, the event links direct to valid Facebook event pages, once a user selects to view more information, the malicious link is displayed [3]. Similarly, botnets, worms, and viruses have emerged on OSN sites. The study on spam conducted in [4] points out different strategies used by bots to launch successful spam campaigns. Such spam campaigns consist of a single spammer having multiple accounts on OSN sites. This campaign strategy increases the chances of a user being exposed to spam.

In this paper, we present a spam identification method to detect spam profiles on online social networks. Our study is based on a real dataset of Facebook profiles containing both normal and spam profiles. We have identified a set of features to model the social network as a weighted graph. The feature set consists of statistics related to Facebook community pages, links shared, and friends. The statistics are calculated by logging the complete wall history of profiles. The identified features are the key elements used for social network interactions, e.g., Facebook community pages are tools for publicly sharing latest updates on topics of interests. Similarly, sharing URLs with friends is the key source of information sharing. Using these features, we have proposed our graph-based social network modeling and applied Markov clustering for the identification of spam profiles. In the graph model, a node represents a profile and an edge represents a connection between a pair of profiles. The connection between any two profiles is defined by correlating information extracted from actual profiles. The information consists of list of URLs shared, list of friends and list of Facebook fanpage-likes of each profile. Using these lists, we generate an n x n adjacency matrix M where n is the number of nodes in the weighted graph. The matrix M is then fed into the well-known graph clustering method (Markov clustering) to group similar profiles together. As a result of applying MCL, the graph nodes are grouped into three different types of clusters – (i) clusters containing all spam profiles, (ii) clusters containing all normal profiles, and (iii) clusters containing both spam and benign profiles. In third case, we have applied majority voting concept to resolve the class of a test profile.

The rest of the paper is organized as follows. Starting with a brief review of the related works on online social network security in section II, we have described our dataset in section III. Section IV presents the social network modeling approach using a weighted graph. Markov clustering and Experimental results are presented in section V. Finally, section VI concludes the paper and provides some insights of our future works.

II. Related Work

The huge amount of information available through the online social networking sites has attracted researchers to mine this information and study issues faced by the social network community. Considerable work has been done for collecting and mining the information for various problems such as community detection, information diffusion and spam filtering. In [5], the authors investigated the feasibility of using measurement calibrated graph models for sharing information among researchers without revealing private data. In [6], the authors presented a study of topological characteristics of Twitter OSN. The authors investigated the behavior of information diffusion over the Twitter network by analyzing *retweets* and found that an information retweeted once reaches on average 1000 users. The authors in [7] presented a study of clickstream data of social networks. Their analysis shows that the use of clickstream data provides rich information about social interactions, and that a majority of user activities on social networks consists of "browsing". Similarly in [8], the authors investigated social interactions of users on OSNs and proposed that a majority of interactions on OSN sites are latent in nature, whereas visible events occur less frequently.

There has been some research for the detection and prevention of spam on OSNs. In [9], the authors proposed a real-time URL-spam detection scheme for Twitter. They logged browser activity as a URL loads in the browser and monitored a multitude of details including redirects, domains contacted while constructing a page, HTML content, pop-up windows, HTTP headers, and Java script and plugin execution to detect spam links. Another substantial work on detection of spam on OSNs is presented in [4]. In this work, the authors created honey-profiles representing different age, nationality, etc. Their study is based on data collected from profiles of several regions including USA, Middle East, and Europe. They logged all types of requests on Facebook, as well as wall posts, status updates, and private messages. On MySpace, they recorded mood updates, wall posts, and messages. On Twitter, they logged tweets and direct messages. Based on these activities they distinguished spam profiles from normal profiles. The authors in [10] also utilized the social honeypot concept to lure content polluters on Twitter. The Twitter users harvested are analyzed and a set of features is proposed for the classification purpose. The technique is evaluated on a new dataset of Twitter spammers collected from "@spam mention" provided by Twitter to flag spammers. In [11], the authors analyzed a large dataset of wall posts on Facebook user profiles for the detection of spam accounts. They built wall post similarity graph for the detection of malicious wall posts. Similarly in [12], the authors presented a thorough analysis of profile-based and content-based evasion tactics employed by Twitter spammers. In this work, the authors proposed a set of 24 features consisting of graph-, neighbor-, automation-, and timing-based features. The features are evaluated by the application of machine learning techniques on the datasets. The authors also formalized the robustness of the proposed feature set. In [13], the authors presented a large-scale effort to characterize spam on Twitter. Using click-through data generated from spam URLs, the authors analyzed the success of Twitter spam at luring over 1.6 million users into visiting spam webpages. The authors clustered spam URLs present in Tweets to identify trends which can distinguish spam, malware and phishing. In [14], the authors proposed a combination of content-based and user-based features for the detection of spam profiles on Twitter. In order to evaluate the importance of these features, the collected dataset is fed into traditional classifiers. A study of monetary relationships of spammers is given in [15]. The paper gives an analysis on the behavior of Twitter spammers. Based on a large Twitter dataset, the authors identified monetary relationships of spammers with vendors seeking to distribute their URLs. The authors also analyzed major spam campaigns and their life spans.

Compared to the existing work, our study presents a different methodology for the detection of spam profiles. Firstly, we employ a new graph clustering technique (Markov Clustering), which does not require the number of clusters to be supplied by the users as in the case of k-means algorithms. Secondly, we apply majority voting to improve the clustering results. Moreover, most of the previous works target Twitter spam, whereas we have worked in this paper on the identification of Facebook spam. The most prominent work on the detection of Facebook spam in presented in [11], however the authors have worked on the detection of individual malicious wall posts rather than complete profiles and the spam campaign analysis identifies campaigns on the basis of the bait (for example free gifts, jobs, etc.) used to lure victims. Although, identification of spam posts is necessary for securing a users' profile, but in order to minimize the number of spam on Facebook, it is desirable to identify profiles that are under the control of a single spammer. This spam-profile detection strategy will have a significant impact in making spam-free online social

TABLE I
STATISTICS OF FACEBOOK PROFILES DATA

	Links	Likes/Hashtags	Friends/Mentioned
Facebook Normal	20175	21975	42124
Facebook Spam	53836	67536	107953

networking sites.

III. DATA COLLECTION

Facebook is the most popular online social network claiming 800 million active users [16]. The popularity of Facebook can be associated to its platform features that make social interactions and information sharing more interactive. To develop a proof-of-concept model of the social network graph, we crawled publicly accessible Facebook data containing both normal and spam profiles. Since in our case, the graph is a representation of true social interactions, we logged information that are only related to the social interactions of profiles. Some of the popular features considered by our data collection module are wall posts, fan pages and tags. A brief description of these features are presented in the following paragraphs.

- ***Wall Posts***: A users' Facebook wall is a place where her friends (or everyone depending on the privacy settings) can interact by posting messages and useful links. Users can also like and comment on the wall posts. According to Facebook statistics (November 2011), in a single day about 2 billion wall posts are liked or commented.
- ***Pages***: Facebook pages are designed for celebrities, business organizations, etc., that intend to share information to people outside their real social circle. Users can `like` certain pages to get latest updates about their interests. According to Facebook statistics (November 2011), a single user has indirect connection to larger groups of users via 80 (on average) community pages, groups and events.
- ***Tags***: Facebook tagging feature allows users to tag friends and pages in posts (analogous to twitter mention). Once tagged in a post the content being shared becomes visible on the subjects' wall and hence affects information diffusion.

For Facebook profiles users' activity on his/her Facebook wall were logged. Only information available for public view was collected and users with restricted view of their profiles were not considered. We logged activities related to friendship requests, wall posts, fan page likes and links shared. We logged only the visible interactions of a profile. We logged this information from a total of 320 Facebook profiles, including 165 spam profiles and 155 normal profiles. A profile is categorized as spam on the basis of its visible activities in the network. Spammers exhibit major wall post activity, consisting of links directing to mostly fake pornographic websites, personal blogs, advertisements, and so on. Our dataset consists of 104 spammers who solely exploit the posting feature. Another category of spam profiles consists of compromised accounts, infected or hacked by malicious Facebook applications. Such accounts exhibit a plethora of posts sharing the same link directing to some advertisement campaign. We identified 16 compromised accounts in the spam dataset. A majority of spammers use the Facebook tagging feature; in each post Facebook friends and fan pages are tagged which makes the link visible to more people than originally tagged. Our dataset consists of 34 such spammers. Profiles of the users whose behavior were contradictory to the spam users were categorized as normal profiles. Some profiles showing no other activity, except sharing a large number of links to Zynga games or Youtube videos were not considered. Table I gives a detail statistics of our collected Facebook dataset.

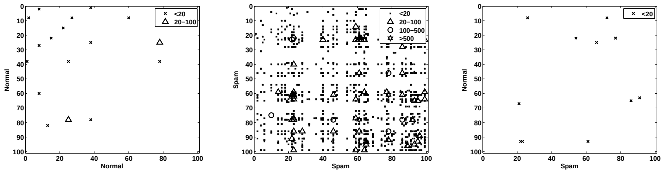
IV. FEATURE IDENTIFICATION AND SOCIAL NETWORK MODELING

In this section, we discuss the modeling process of social networks data using a weighted graph, in which user profiles are represented as nodes and their interactions as edges. It should be noted that the linkages between profiles are used to model user interactions, rather than their friendship relations. Formally, the social network is defined as a weighted graph $G = (V, E, W)$, where V is the set of profiles, $E \subseteq V \times V$ is the set of edges, and $W \subseteq \Re$ is a set of weights assigned to edges. For each node $v \in V$, a 3-dimensional feature vector comprising of number of *active friends*, *page likes* and *URLs shared* is identified. Using the feature vectors of nodes, the weight of an edge $e_{ij} = (v_i, v_j)$ is calculated as an aggregation of the common *active friends*, *page likes* and *URLs shared* of nodes v_i and v_j. Further details about the features and the weight calculation process is presented in the following paragraphs.

Active Friends: This feature captures the interaction frequency of a user with its friends in the network. For a user v_i with F_i as the set of friends, the set of active friends F_i^a can be calculated as an intersection of the set F_i and the set of the friends of v_i who were either contacted by v_i or those who interacted with v_i through *wall posts*, *comments* or *tags*. Mathematically, this can be defined using equation 1 in which I_i is the set of users with whom v_i has interactions in the network. For a node v_i, the value of the "active friends" feature is taken as the cardinality of the set of its active friends F_i^a. Similarly, the set of common active friends in the network with whom a pair of users v_i and v_j have interacted is calculated as the intersection of their active friend sets F_i^a and F_j^a, respectively, as given in equation 2. For an edge $e_{ij} = (v_i, v_j)$, the value of the "active friends" feature is taken as the cardinality of the set of common active friends F_{ij}^a.

$$F_i^a = F_i \cap I_i \tag{1}$$

$$F_{ij}^a = F_i^a \cap F_j^a \tag{2}$$

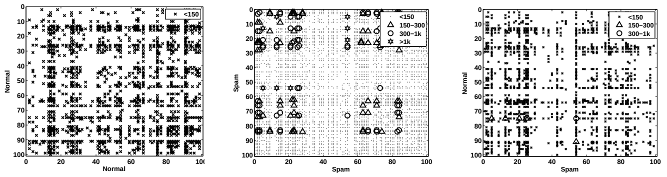


(a) Normal vs Normal

(b) Spam vs Spam

(c) Normal vs Spam

Fig. 1. Sparsity pattern plot for *active friends*

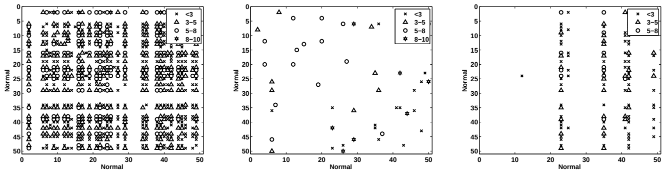


(a) Normal vs Normal

(b) Spam vs Spam

(c) Normal vs Spam

Fig. 2. Sparsity pattern plot for *page-likes*



(a) Normal vs Normal

(b) Spam vs Spam

(c) Normal vs Spam

Fig. 3. Sparsity pattern plot for *URLs*

On analysis, we found that normal users generally interact with a small portion of their friends, usually with the ones who are more active or have similar interests. In contrast, spammers have a large number of friends and their activity consists of mainly one way interactions with majority of their friends. This feature is important for the identification of spam campaigns. Most spammers use multiple profiles to launch an effective spam campaign. These multiple spam profiles are interlinked either by direct friendships or through mutual friends. Such mutual friends are also mutual targets of the spam profiles. Figure 1 shows zoomed sparsity pattern plots of active mutual friends of 100 randomly selected nodes. Figure1(a) shows that some normal profiles are inter-connected through a maximum of 20-100 mutual friends. Spam profiles are clearly more inter-linked through active mutual friends, numbering upto 600. This shows the penetration of coordinated spam profiles. Moreover, as shown in figure 1(c), some normal profiles are connected to spam profiles as they have some common friends with spam profiles. The small number of connections shows that once a friend is spammed, the identity of the fake profile is exposed and it becomes difficult for the spammer to expand its network.

Page-Likes: This feature captures the *page-likes* frequency of the users in a social network. For an edge $e_{ij} = (v_i, v_j)$, the common *page-likes* of v_i and v_j, P_{ij}, is calculated as the intersection of the sets of *page-likes* of v_i and v_j, as given in equation 3, and the *page likes* attribute value is calculated as the cardinality of the set P_{ij}.

$$P_{ij} = P_i \cap P_j \tag{3}$$

On analysis, we found that spammers greatly exploit the page-likes feature for spreading spam. Tagging a community page in a spam post or directly posting spam on the pages' wall makes the spam link visible to all members of the community. This greedy behavior of spammers is depicted in Figure 2, which shows that most of the spammers target the same community pages, usually the popular ones. We also found that some of the spammers shared same target, i.e., more than 1000 common community pages, whereas normal users had generally less than 150 common pages, as shown in figure 2(a). Figure 2(c) also shows that most of the spam and normal profiles have common page-likes numbering upto 150, depicting the level of infiltration of spam in the normal user community.

URLs: This feature captures the URL sharing patterns of the social network users. For an edge $e_{ij} = (v_i, v_j)$, the common URLs of v_i and v_j, U_{ij}, is calculated as the intersection of the sets of URLs shared by v_i and v_j. For each node $v_i \in V$, we generate frequency histogram for URLs shared by the corresponding users, and for an edge connecting a pair of nodes, the URL attribute value is calculated as a fraction of the URLs commonly shared by them using equation 4.

$$U_{ij} = \frac{U_i \cap U_j}{U_i \cup U_j} \tag{4}$$

Our analysis reveals that spammers promoting personal blogs and websites generally share links to different pages of the same website. We map every unique URL on a unique website identifier and all webpages of a single site on the same website identifier. Figure 3(a) shows that normal users have larger number of common URLs (more than 8 websites) including links to Youtube videos, Zynga games, popular news articles, etc. It can be observed in figure 3(b) that some spam profiles have significant number of common URLs (more than 8 common website identifiers). This behavior reveals the spam campaigns carried out for promoting blogs, advertisements, and so on. In Figure 3(c), some spam profiles have shared URLs posted by most of the normal profiles, which depicts the evasion tactics employed by the spammers.

On the basis of the above-mentioned features, each edge $e_{ij} = (v_i, v_j)$ is assigned a weight $\omega(e_{ij})$ that is calculated as an aggregation of the individual feature values as given in equation 5. In equation 5, $|.|$ represents the cardinality of a set.

$$\omega(e_{ij}) = |F^a_{ij}| + |P_{ij}| + |U_{ij}| \tag{5}$$

V. Profile Classification and Experimental Evaluation

Given a set of nodes $V = \{v_i : i = 1, ..., n\}$ and weighted edges $E = \{(v_i, v_j) : 1 \leq i, j \leq n\}$, an adjacency matrix $A_{n \times n}$ is created to represent the graph in a machine-readable format. The value of a cell $A(i, j) = a_{ij} \geq 0$ represents the weight of the edge connecting the nodes v_i and v_j, which is calculated using equation 5. On the basis of the way weight values are calculated (see equation 5), the weight value of an edge $e_{ij} = (v_i, v_j)$ can be considered as a similarity value between the nodes v_i and v_j. On matrix A, we apply Markov clustering which uses a random walk on the weighted graph. It calculates the probability of transition, i.e., probability of moving from one node to another in the graph. So, for a similarity matrix, A, the normalized adjacency matrix M is the transition matrix for a Markov random walk and $M(i, j) = m_{ij}$ is the transition probability. Considering the transition probability from one node to another in t steps as $M.M^{t-1}$, the transition probability is inflated, i.e., higher transition probabilities are increased and lower transition probabilities are decreased. This is done by taking m_{ij} to the power $r \geq 1$, as given in equation 6, where r is an inflation parameter.

$$\Upsilon(M, r) = \left\{ \frac{(m^r_{ij})}{\sum_{a=1}^{n}(m^r_{ia})} \right\}^{n}_{i,j=1} \tag{6}$$

The markov clustering method performs matrix expansion and inflation iteratively, i.e., it takes successive powers of M and then performs the inflation process. The iteration terminates when the matrix difference in terms of *Frobenius norm* as given in equation 7 falls below a threshold $\epsilon \geq 0$. In our experiment, the value of ϵ is 0.001.

$$||M_t - M_{t-1}||_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(m_{ij_t} - m_{ij_{t-1}})^2} \tag{7}$$

To evaluate the performance of our approach, we used quality metrics `B-cubed` (F_B) and F_P`-measure` (F_P) [17]. These measures are used to quantify the quality of disambiguation. A brief description of these measures is given in the following paragraphs.

B-cubed (F_B): For each profile p, let S_p is the set of profiles whose class labels match with the label of p in test dataset, and A_p is the set of profiles whose class labels match with the class of p in the resultant clusters obtained after applying the clustering algorithm. Using these sets, the local precision ($Precision_p$) and local recall ($Recall_p$) are calculated using equations 8 and 9, respectively. Thereafter, *Precision* (*Recall*) values are calculated as an average of $Precision_p$ ($Recall_p$) over all objects p. Finally, the value of *B-cubed* is calculated as a harmonic mean of the *Precision* and *Recall*.

F_P**-measure** (F_P): Considering a set S of clusters present in the test dataset and a set A of clusters obtained by applying the algorithm, the F_P-$measure$ is defined as the harmonic mean of $Purity$ and $InversePurity$ defined using equations 10 and 11, respectively. In these equations, P is the set of all profiles in the test dataset.

TABLE II
Performance evaluation results in terms of F_P and F_B measures

r	1.2	1.5	1.7	2.0	2.1	2.3
F_P	0.6791	0.8529	0.7891	0.7885	0.7810	0.7772
F_B	0.6549	0.7528	0.6770	0.6775	0.6670	0.6609

TABLE III
Cluster details at $r = 1.5$

Clusters	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Size	9	4	138	169
Spam	4	4	128	30
Normal	5	0	10	139

$$Precision_p = \frac{|A_p \cap S_p|}{|A_p|} \tag{8}$$

TABLE IV
PERFORMANCE EVALUATION RESULTS AFTER APPLYING MAJORITY VOTING

r	1.2	1.7	2.3	2.7	3.5	7	10
F_P	0.6791	0.8656	0.8688	0.8719	.8750	.8719	.8750
F_B	0.6549	0.7752	0.7792	0.7832	.7886	.7845	.7873

$$Recall_p = \frac{|A_p \cap S_p|}{|S_p|} \tag{9}$$

$$Purity = \sum_{A_i \in A} \frac{|A_i|}{|R|} max_{S_j \in S} \frac{|A_i \cap S_j|}{|A_i|} \tag{10}$$

$$InversePurity = \sum_{S_i \in S} \frac{|S_j|}{|R|} max_{A_i \in A} \frac{|A_i \cap S_j|}{|S_j|} \tag{11}$$

Table II shows the results obtained for different values of r. The best results obtained are at $r = 1.5$ for both F_B and F_P with $F_B = 0.7528$ and $F_P = 0.8529$. Table III shows further details about the 4 clusters formed at $r = 1.5$. Clusters 3 and 4 are the two main clusters constituting majority of nodes. Cluster 3 has majority of spam nodes and 4 has majority of normal nodes. In order to further improve the results, we propose majority voting to include outlier clusters in the majority clusters.

The clusters obtained after performing Markov clustering are divided into three categories. The first category represents clusters containing purely spam profiles, the second category includes clusters consisting purely normal profiles, and the third category of clusters have mixup of both spam and normal profiles. In third case, the cluster is termed as *outlier cluster* and we have applied the concept of majority voting to assign a class (spam or normal) to the corresponding cluster and accordingly merge with the similar clusters. According to the results at $r = 1.5$, there are two small clusters. In order to converge the number of clusters to 2, we use simply majority voting to include the outlier nodes or clusters in the major clusters. For each outlier cluster O_i the number of spam and normal nodes defines the association of O_i with either the spam-majority cluster or the normal-majority cluster. Table IV shows the performance values in terms F_P and F_B measures that are obtained after applying majority voting. Majority voting reduces the number of clusters to 2, however the voting scheme allows selection of outlier clusters, which were part of a wrong majority cluster at lower values of r. As the value of r is increased, more outlier clusters are formed and are voted into their respective majority clusters. The maximum number of clusters obtained are 18 for $r \geq 10$. The best results are achieved at $r = 3.5$ with $F_P = 0.88$ and $F_B = 0.79$.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an MCL-based approach to detect spam profiles in Online Social Networks (OSNs) like Facebook. To this end, we have identified a set of three features and used them to model social interactions of OSN users using a weighted graph. MCL is applied on the social graph to exploit the behavior similarity of profiles and unearth the clusters present in profile dataset. The concept of majority voting is used to determine the class of outlier clusters and merge them with appropriate clusters. Presently, we are enhancing our dataset and working towards identifying more discriminative features and evaluating the efficacy and scalability of the proposed method on a larger real dataset collected from different social networks.

It should be noted that, supervised learning techniques like naive Bayes, decision tree, etc. can be a choice if the target is to classify the profiles either as *normal* or as *spam*, but generally these techniques are inappropriate if the target is to identify different types of spam campaigns – as in such situation it would not be possible to determine the exact number of classes present in the dataset. Therefore, MCL and similar clustering technique is an obvious choice if the target is to classify the dataset into n classes, where the value of n is not known in advance. Although, the spam campaign identification is not the focus of this paper, we have applied MCL algorithm instead of naive Bayes or decision tree to enhance the proposed approach in future to identify more refined classes of profiles, i.e., different sets of spam profiles constituting different types of spam campaigns.

ACKNOWLEDGMENT

The authors would like to thank King Abdulaziz City for Science and Technology (KACST) and King Saud University for their support. This work has been funded by KACST under the NPST project number `11-INF1594-02`.

REFERENCES

[1] 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 (WWW'09)*. ACM, 2009, pp. 551–560.

[2] Symantec, "Symantec Intelligence Report: August 2011."

[3] http://www.geeksugar.com/Spammers-Target-Facebook-Events-15447506.

[4] 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.

[5] A. Sala, L. Cao, C. Wilson, R. Zablit, H. Zheng, and B. Zhao, "Measurement-calibrated graph models for social network experiments," in *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*. ACM, 2010, pp. 861–870.

[6] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*. ACM, 2010, pp. 591–600.

[7] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (IMC'09)*. ACM, 2009, pp. 49–62.

[8] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, and B. Zhao, "Understanding latent interactions in online social networks," in *Proceedings of the 10th Annual Conference on Internet Measurement Conference (IMC'10)*. ACM, 2010, pp. 369–382.

[9] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, “Design and evaluation of a real-time url spam filtering service,” in *IEEE Symposium on Security and Privacy*, 2011.

[10] K. Lee, B. Eoff, and J. Caverlee, “Seven months with the devils: A long-term study of content polluters on twitter,” in *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM’11)*, 2011.

[11] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao, “Detecting and characterizing social spam campaigns,” in *Proceedings of the 10th Annual Conference on Internet Measurement (IMC’10)*. ACM, 2010, pp. 35–47.

[12] C. Yang, R. Harkreader, and G. Gu, “Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers,” in *Proceedings of the 14th International Symposium on Recent Advances in Intrusion Detection (RAID’11)*, 2011.

[13] 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 (CCS’10)*. ACM, 2010, pp. 27–37.

[14] M. McCord and M. Chuah, “Spam detection on twitter using traditional classifiers,” *Autonomic and Trusted Computing*, pp. 175–186, 2011.

[15] K. Thomas, C. Grier, D. Song, and V. Paxson, “Suspended accounts in retrospect: An analysis of twitter spam.”

[16] www.facebook.com/press/info.php?statistics.

[17] D. V. Kalashnikov, Z. Chen, S. Mehrotra, and R. Nuray-Turan, “Web people search via connection analysis,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, pp. 1550–1565, 2008.