A survey on Online Spam detection Methods in Social Networking sites

R.Renuga Devi, M.Hemalatha[*]

ABSTRACT

Social networks continuously grow in popularity. Most of the Internet users access the web for social networking sites. The Web sites are organized around contents in the form of text. But online social networks are organized around users. Because of the popularity most of the cyber criminals have utilized social networks as a place to perform their illegal activities. There is a need to create the techniques and methods to identify these malicious activities and the criminal in an effective and fast manner. There are many researches are going on this area. But that are insufficient. Still several methods are needed. Criminals use the social networks to send spam mails, to spread malwares, to hack the accounts, and to involve and hack some business applications, financial transactions. This paper presents a study and analysis of the online spam detection in social networks. Several methods are available to find the online spammers. The results of this analysis show that, there is a lack in social networks security. Here, given a discussion about the impact of online social networks at the hands of spammers.

Keywords: *Cyber Criminals, Malicious activities, Spammers, Spam Detection, Social Networks.*

I. INTRODUCTION

Social networks are constructed by a group of actors; actors may be individual person or the entire organizations, and the relationships between these actors. The social network gives a best way to analyse the structure of the entire social entities. Social networking sites have gained lot attractiveness in the recent years, because it provides various opportunities to the people. It allows the people to connect each other in an easy and timely manner, and allow as sharing and trading a variety of knowledge based information.

The entire social networks are represented as a graph. The graph structures are sometimes very complex to analyse. In a large network, it has different kinds of nodes and relationships. Several research processes are going o this area. The author considered several levels in a network. It is very difficult to determine the methods are used to solve the particular problem. Social network analysis is related to network theory. It is a emerging technique in a modern sociology. Social network analysis is moving from being an indicative metaphor to an analytic approach to a paradigm. It has several theoretical methods and statements. There are several social network analysis software are available for the researchers. There are two types of social network analysis are available.

i) Complete or Whole network analysis

ii) Individual or personal network Analysis

This analysis is used to study, all of the ties between the specific relations, or ties between the personal communities [1]. The difference between complete networks and personal networks is depends on the data availability. Now a day researchers used Collaboration graphs to demonstrate the good and bad relationships between people in a relation. Positive relationships are

1. Department of Computer Science, Karpagam University, Coimbatore. nicrdevi@gmail.com

2. Department of Computer Science, Karpagam University, Coimbatore. hema.bioinf@gmail.com

represented as positive edges. Like friendship, alliance, dating. And we can use a negative edge between two nodes indicate a negative relationship. Like criminals, hatred, anger. And also researchers focus on the Signed social network graphs. It can be used to guess the opportunities and future evolution of the graph. Signed social networks provide balanced and unbalanced cycles. If the product of signs is positive then it is represented as a balanced cycle. Balanced graphs characterized as a group of people who are improbable to change their opinions of all other people in the same group. In other words, unbalanced graphs denote the people who are all ready to change their opinions of all other people in the same group. For example, a

Group contains 3 people, likely A, B, and C. Here A and B, B and C have a positive relationship. C and A have a negative relationship. That is called as an unbalanced cycle. That is the group having more possibilities to morph into a balanced cycle. Here the person B has a good relationship with A, but both the person A and B have a negative relationship with the person C. The evaluation the signed social network graphs was predicted using balanced and unbalanced cycles. For better representation of a graph, Visual representations are used to recognize the network data. And it is easy method to convey the result of the analysis [2]. Most of the analytic software has the separate modules for network visualization. For easy investigation, the nodes and the relationships of the nodes are displayed in different layouts, colour, size and various properties of the nodes. Classical representation of the social network data are graphs. Graphs are not easy to understand, it does not allow sensitive analysis. Several methods have been introduced in order to represent network data in more instinctive format. Example is Sociomapping [3].

Unluckily, there are several tools of social networking area is in wrong hands. The criminals misused these tools for unwanted activities. In social network sites, spammers create duplicate accounts and access other users' accounts for personal usage. Social network spammers are different from the traditional systems spammers. In social media the spammers act as normal users. And the illegal persons attempt to change the anti spamming methods. Most of the social networks do not provider fully secured system, because of the privacy policies. Many systems do not fully concentrate on the user's content. Spammers use the networks to send unwanted messages, particularly advertising, arbitrarily. Different forms of spam are recognized. Such as social networking spam, social spam , e-mail spam, Usenet newsgroup spam, spam in blogs, Web search engine spam, instant messaging spam, online classified ads spam, wiki spam, mobile phone messaging spam, junk fax transmissions, Internet forum spam, television advertising and file sharing network spam [4].

In this paper consists of three papers, related to Spam analysis in social networks. Now a day most of the social networks are used as a well-liked platform for sharing real time information on the web, Most of the spammers target the Twitter for their malicious activities. Our analysis consists of Chao Yang, et al., methods for Cyber Criminal Ecosystem on Twitter, Kurt Thomasy, et al., methods for Suspended Accounts in Retrospect, and Hongyu Gao, et al., methods for Online Spam Filtering in Social Network for our analysis. Here given the existing methods for analyzing spammers on Twitter and the effectiveness of the existing methods and techniques.

II. Online Spam detection Methods Survey

A. *Cyber Criminal Ecosystem on Twitter*

Chao Yang, et al., analysis the cyber criminal ecosystem on Twitter, Main idea is to analyze the inner and outer relationships of the criminal account community. An Inner relationship analyze is used to find, how the criminal accounts are socially connected to form a small-world network. Criminal hubs are found in the analysis, which is in the center of the graph, are more useful to follow criminal accounts. Outer social relationships between criminal accounts and the social friends outside the criminal account community, from those three categories are exposed that have close friendships with criminal accounts. Through these analyses, a novel and effective criminal account inference algorithm by exploiting criminal accounts social relationships and semantic coordination was developed.

For experimental analysis, dataset was used from the authors previous Twitter spam account detection study [5]. The dataset contains 485,721 accounts with 14,401,157 tweets and 5,805,351 URLs. The inner relationships is represented as a directed graph, and the criminal relationship graph was named as G = (V, E). The dataset consists of 2,060 nodes and 9,868 directed edges. And further the nodes are divided, from that 8 weakly connected components are taken. Each component contains minimum of three nodes and 521 isolated nodes. The massive connected component contains 954 nodes. After visualizing sample criminal relationship graph, the graph was analysed through utilizing graph theoretical knowledge, and observed the following two main findings. First found that, Criminal accounts tend to be socially connected, forming a small-world network. To quantitatively validate these findings three graph metrics are used: graph density, reciprocity, and average shortest path length.

Criminal hubs are extracted by calculating hub scores of criminal accounts in terms of their positions in the criminal relationship graph by utilizing the HITS algorithm. Particularly, for each vertex i in the graph, Eq.(1) and (2) are used to compute its hub score Ht i and authority score Ati in the t-th iteration. When the computation converges within several iterations, the vertex's final hub score Hi was obtained.

$$H_i^t = \begin{cases} 1, & if\ t = 0 \\ \sum_{(i,j)\in E} A_j^{t-1}, & if\ t > 0 \end{cases} \tag{1}$$

$$A_i^t = \begin{cases} 1, & if\ t = 0 \\ \sum_{(j,i)\in E} H_j^{t-1}, & if\ t > 0 \end{cases} \tag{2}$$

According to this algorithm, a higher hub score of an account implies that it follows many accounts with high follower numbers. Thus, 90 criminal hubs are extracted with higher scores and 1,970 criminal leaves with lower scores by using k-means algorithm and setting k = 2.

Secondly, criminal hubs are more inclined to follow criminal accounts when compared with criminal leaves. To perform this process Criminal Following Ratio (CFR) metric was used. It indicates the ratio of the number of an account's criminal-followings to its total following number. If CFR value is high for one account, then is more inclined to follow criminal accounts. Another metric was designed, called Shared Follower Ratio (SFR), which is the percentage of an account's followers, who also follows at least one of this account's criminal-followings. This metrics is used to find that criminal hubs' SFRs are higher than criminal leaves'. Around 80% of criminal hubs' SFRs are higher than 0.4, while around 5% of criminal leaves have such values.

This observation reflects that compared with criminal leaves, criminal hubs' followers share more followers with their criminal followings. This indirectly implies that criminal hubs could obtain followers by knowing their criminal-followings' followers' information.

Chao Yang, Robert Harkreader and et al., also performed analysis on Outer Social Relationships. To extract the Criminal Supporters a Malicious Relevance Score Propagation Algorithm (Mr.SPA) Specifically was designed, Mr.SPA assigned a malicious relevance score (MR score) to each Twitter account, measuring how closely this account follows criminal accounts. A higher

MR score implies a closer "follow relationship" to criminal accounts. After extracting criminal supporters, three representative categories of supporters are observed (social butterflies, social promoters, and dummies) according to their defined thresholds.

Then Criminal account Inference Algorithm (CIA) was designed to propagates malicious scores from a seed set of known criminal accounts to their followers according to the closeness of social relationships and the strength of semantic coordination. If an account accumulates sufficient malicious score, it is more likely to be a criminal account.

The intuition of CIA is based on the following two observations:

(1) Criminal accounts tend to be socially connected;

(2)Criminal accounts usually share similar topics (or keywords or URLs) to attract victims, thus having strong semantic coordination among them.

In that, found some difficulties as well as some limitations. And the data set contain some bias. Also, the number of analysed criminal accounts is most likely only a lower bound of the actual number in the dataset, because it target on one specific type of criminal accounts due to their severity and prevalence on Twitter. However, it is extremely challenging to obtain an ideal, unbiased dataset with perfect ground truth. In addition, to reduce possible data sampling bias crawled two datasets at very different time to evaluate the performance of their CIA. An effective algorithm was designed to infer more criminal accounts by starting from a seed set of known criminal accounts and exploiting the properties of their social relationships and semantic correlations [6].

C Suspended Accounts in Retrospect: An Analysis

B. Suspended Accounts in Retrospect: An Analysis

Kurt Thomasy, et al., analyze the impact of on-line social networks at the hands of spammers through several tools, techniques. For the analysis, 1.1 million accounts are taken, which is suspended by Twitter for tumultuous activities over the course of seven months. During this seven month method, 1.8 billion tweets are collected, eighty million of that belong to spam accounts. The Twitter dataset contains of over 1.8 billion tweets collected from Twitter's streaming API [7] throughout a seven month amount from August seventeen, 2010 to March four, 2011. The primarily dataset was used to know the behaviour and lifelong of spam accounts, the campaigns executed, and also the wide-spread abuse of legitimate net services like uniform resource locator shorteners and free net hosting. The rising marketplace of illegitimate programs operated by spammers was ascertained that embody Twitter account sellers, ad-based uniform resource locator shorteners, and spam affiliate programs that facilitate modify underground market diversification. Following table 1 rundown the info, that square measure collected by the authors from varied sources.

TABLE I

SUMMARY OF DATASET

Data Source	Sample dataset size
Tweets	1,795,184,477
Accounts	32,852,752
Distinct URLs	1,073,215,755
Tweets from Suspended Accounts	80,054,991
Suspended Accounts	1,111,776
Distinct URLs from Suspended Accounts	37,652,300
Resolved URLs	15,189,365
Bit.ly URLs	10,092,013
Bit.ly Accounts	23,317

The above table is taken from the paper of Kurt Thomasy [10], et al., In order to characterize the tools and services that Twitter spammers depends on, first manually verified a sample of suspended accounts and realize the overwhelming majority were suspended for spamming, providing with an expensive supply of ground truth for menstruation spam.

The sample of suspended accounts are verified and located that, the large half were suspended for spamming, providing an expensive supply of ground truth for menstruation spam. Additionally to their Twitter dataset, resolved the primary airt of 15 million URLs to deobfuscate a layer of shortening. Finally, for 10 million URLs shortened by bit.ly, downloaded multiple statistics provided by bit.ly as well as click through and, once obtainable, the bit.ly account that shortened the computer address.An outline of their dataset is given in the Table 1. The following metric was applied for sensitivity:

Sensitivity = true positives / true positives + false negative

The above method was used to estimates the number of false positives and false negatives that result from Twitter's spam detection algorithm; approximate the algorithm's sensitivity, or the likelihood that a spam account posting URLs will be caught. Over the 31 million accounts that were not suspended, 6% are false negatives, amounting to roughly 1.9 million spam accounts that are overlooked. Another 1 million spam accounts were correctly identified by Twitter's algorithm. Found that only 37% of spam accounts posting URLs on Twitter are caught by the suspension algorithm during the period of our measurement.

In order to tag spam within the dataset, first identified the accounts, which is suspended by Twitter for violent behaviour. The suspended accounts include spam, aggressive friending, and other non-spam related offenses. Because of the slow process of the Twitter's account suspension algorithm, supposed to wait for two weeks from the last day of data collection before the accounts were suspended by Twitter. Within the combination of spam activities on Twitter, a various set of tools and strategies are recognized that build upon access to hundreds of fraudulent accounts, an array of spam URLs and domains, and automation tools for interacting with Twitter.

The analysis on Compromised and the fraudulent Accounts was carried out. From the literature review results of social networks found that 97% of accounts sending spam on Facebook were compromised [8], compared to 84% of accounts on Twitter [9]. In contrast, majority of suspended accounts in the dataset were fraudulent and created for the explicit purpose of spamming. And disparity results from how the datasets for each study were generated. From the studies, given only 8% of the URLs posted by fraudulent accounts appeared in suspected list. These lists act as a source for identifying social network spam in the existing studies, with a clear bias based on compromised accounts. The results of their analysis are viewed in conjunction with these existing studies, offers a wider viewpoint of the huge number of spamming methods in social networks.

In this work, a singular look of the behaviours of spammers on Twitter by analyzing the tweets sent by suspended users looking back. The marketplace for Twitter spam uses a various set of spamming techniques, as well as a spread of methods for making Twitter accounts, creating spam URLs, and distributing spam. The options area unit plain-woven along to make 5 of the most important spam campaigns on Twitter accounting for nearly 20% of the spam during a dataset. What is more, to found a rising spam-as-a-service market that features honoured and not so reputable associated programs, ad based shorteners, and Twitter account sellers. Primarily from the analysis, 89% of dishonourable accounts created by spammers forgo participation within the social graph, instead looking forward to unsought mentions and trending topics to draw in clicks. Amazingly, 77% of accounts happiness to spammers was suspended among someday, however despite this rate of attrition, new dishonourable accounts area unit created to require their place, sustaining Twitter spam throughout the course of their seven month measuring. By examining the accounts controlled by individual spammers as discovered by affiliate programs found one or two of actors dominant thousands of Twitter accounts, every pushing a various strategy for monetizing Twitter. Totally, the measurements expose a thriving spam scheme on Twitter that's unflustered by current defences. Their findings highlight the need of higher spam controls targeting each abusive accounts also because the services that support the spam marketplace [10].

C. Online Spam Filtering in Social Network

In the year of 2011, Hongyu Gao, et al., given a online spam filtering system. And this method was deployed as a part of the online Social Networking platform to examine messages generated by users in time period. A technique to reconstruct spam messages into campaigns for classification instead of examine them separately was planned. Tough campaign identification has been used for offline spam analysis, with the assistance of this system support the web spam detection drawback with sufficiently low overhead.

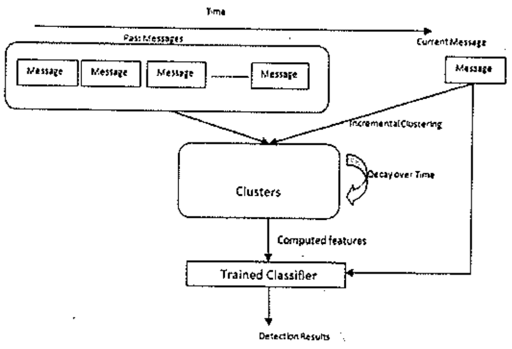

Fig. 1 Overall system Design

Fig. 1 shows the Overall design of the system [13]. System contains two major components. The components are the incremental clustering module and the supervised machine learning module.

If the system receives a replacement message from users, initially it increments the cluster and update the result. Each new message create a cluster by itself, then it combined along as one existing cluster or perhaps trigger the merge of multiple existing clusters. Within the first case, the system directly marks the message as "legitimate" while not invoking the classifier, as a result of the typical measure feature cannot be calculated. This implicit limitation can cause a false negative only if the new message is that the 1st one in an exceedingly spam cluster, that happens terribly seldom. In the other case, the values of the six options of the cluster that the new message resides in square measure (re-)calculated. The opposite clusters square measure intact. After that, the trained classifier accepts these values and decides whether or not these values represent a spam cluster. Note that if the classifier outputs "spam", it'll solely trigger a spam alarm on this message, instead of all the messages within the cluster. Since it's a web system, the choice on the previous messages within the same cluster has already been created and isn't modified. More specifically, once when the system has processed ω messages, it shrinks the six feature values related to every cluster by an element of α, wherever ω and α square measure are two necessary system parameters to see the decaying speed. If a cluster's size is contracted to a price below a threshold t, it's for good removed and every one the resources it takes up square measure freed.

The supervised machine learning module is basically a trained classifier that produces this binary call. Its input is that the feature values related to the cluster that contains the incoming message. We tend to choose call tree because the classifying module within the system.

The dataset was collected, between April and June of 2009 from Facebook crawling. For each crawled use, the complete history of received wall posts within the given time frame was recorded. The dataset was used in this study contains 187M wall posts crawled from roughly 3.5 million users in total. A labeled spam and legitimate messages to train and evaluate the system are used. For that the results of the authors previous study [11], where 199,782 wall posts are confirmed as malicious and are labeled as spam. We label the newly found wall posts as spam, too. Finally, 217,802 wall posts labelled as spam are collected. The remaining wall posts are labelled as legitimate.

i) Methodology:

The system on a server machine that has eight cores (Xeon E5520 2.2 GHz) with Hyper-Threading and 16GB memory. Arrangement of all wall post messages according to their related timestamps and then divide them into the training set and the testing set. The training data set contains first 25% of spam messages, which span from January 1, 2008 to April 21, 2008. It also contains all the real messages in the same time period. The testing set contains all the remaining messages. And "replay" the messages by feeding them strictly according to the time order in both the training and testing phases. Clusters, with the size of at least 5 were used for training.

In order to get the best accuracy, some changes over the data set were made. Based on the Zadrozny et al. suggestions [12], the authors adjusted the ratio of spam to valid message in the training set by randomly sampling to tailor the performance. In that larger ratio indicates a stronger tendency to classify a message as spam. The ratios of 10:1, 4:1, 1:1, 1:4 and 1:10. Regardless of the ratio in the training set, the full testing set is used. A 1:4 ratio in the training set results in a remarkable reduction of false positive rate by about 70%, comparing to a 10:1 ratio. On the other hand, about 13% spams slip through the detection system at this ratio. A 1:10 ratio has related effect in reducing false positive rate, but its true positive rate is fairly lower.

To measure run time performance of the system with the help of the following parameters: Latency and throughput. The time between receiving input messages from the system and the producing output results to the system is known as the latency. The time differences are recorded. The entire examination process for over 80% of messages is finished within 100ms. The average of latency is 69.7ms and the median latency is 39.6ms.

For testing the throughput, the entire dataset was provides to the system as soon as possible. Then the system computes the similarity between the before experimented messages for several new messages in comparable. It has limited hardware specialty (8 physical cores with Hyper-Threading), so executes 16 threads simultaneously. To calculate the average throughput, divide the total number of messages by the total running time, which is 1580 messages/sec.

Through the experiments we conclude that, the online spam filtering system evaluated the system with 187 million Facebook wall messages. The experimental results show that the system achieves high accuracy, low latency and high throughput. Based on their system, it adopts a collection of new features that successfully differentiate spam campaigns. It produces high accuracy and with assurance it drops messages, which is classified as "spam" before that messages reach the anticipated recipients. The system achieves average outturn of 1580 messages/sec and a median process latency of 69.7ms for every message. The high outturn and low latency guarantees that it'll not become the bottleneck of the entire OSN platform [13].

III. Discussions

The Above three papers, related to online spam system. The results of the above motioned paper are discussed here. Chao Yang, et al., analyzes the inner and outer social relationships of criminal accounts. From the analysis found two things: First one is, criminal accounts are socially connected, and it forms a small world network, and second one was criminal hubs are more inclined to follow criminal accounts when compared with criminal nodes. By analyzing the social relationships and semantic coordination of the accounts, a new algorithm was designed, called CIA to gather criminal accounts based on known dataset. CIA gathers 20 times more criminal accounts than a random selection strategy. And also proposed Mr.SPA algorithm and it extracted 5,924 criminal supporters [6].

Kurt Thomasy, et al., found that existing methods of social networks compromise the 97% of accounts sending spam on Facebook [14], when and 84% of accounts on Twitter. But in contrast, found a majority of suspended accounts in their dataset were fraudulent and created for the explicit purpose of spamming. The disparity results from how the datasets for each study were generated. One potential solution for aggregating training data and improving spam detection is to develop Twitter-specific blacklists and spam traps. Shortening services, including bit.ly and HootSuite, already employ blacklists before URLs are shortened [15, 16]. By monitoring which services underpin the spam ecosystem on Twitter, the deployed customized countermeasures for each service, reducing the support infrastructure available to spammers [10].

Hongyu Gao, et al., introduced a methods to use of spam campaigns, instead of individual spam messages, as the objects for spam classification, and solve the challenge of reconstructing campaigns in real-time. Six features are identified that can accurately distinguish spam campaigns from legitimate message clusters in OSNs. And finally the system was developed, which is easy to deploy at the OSN server side to provide online spam filtering [13].

IV. Conclusions

Social networks are gaining popularity day by day. Because of the tremendous growth of social networking sites for combating fraud is slowly getting acceptance within a particular range of sector. In this paper consists of study and analysis of the online spam detection in social networks. Several methods are available to find the online spammers. But they do not fully concentrate on secured network. Many researchers said that, social networking sites such as Facebook, Twitter, LinkedIn, MySpace, are still provide more security to the users. Each and every users of social networking sites are must know about, what they post on their social networking pages. Many sites provide encryption option when we use that sited for communication. If possible all should enable that option. By using the technologies, such as encryption methods, intrusion prevention systems and firewalls we can avoid the criminals attack. But these methods are not sufficient to the users. From the above findings we can identify that, there is a necessity of better spam controls. Most of the existing methods contain some bias. And if the number of criminal account is large, then it is difficult to analyze with the existing methods. So our goal is to focus on the fully secured social networks. Our plan is to design a full detection system by combining the existing algorithms and other detection features.

References

[1] Wellman, Barry and S.D. Berkowitz, eds, 1988. *Social Structures: A Network Approach.* Cambridge: Cambridge University Press.

[2] Bernie Hogan, Juan-Antonio Carrasco and Barry Wellman, *Visualizing Personal Networks: Working with Participant-Aided Sociograms.* Field Methods 19 (2), May 2007: 116-144.

[3] Social Network Analysis :Theory and Applications.Creative Commons Attribution-Share Alike 3.0 Unported. http:/ / creativecommons. org/ licenses/ by-sa/ 3. 0/.

[4] http://en.wikipedia.org/wiki/Spam_%28electronic%29.

[5] C. Yang, R. Harkreader, and G Gu, 2011. *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).

[6] Chao Yang, Robert Harkreader, Jialong Zhang, Seungwon Shin, and Guofei Gu ,2012, *Analyzing Spammers' Social Networks for Fun and Profit: A Case Study of Cyber Criminal Ecosystem on Twitter,* (IW3C2), April 16–20, Lyon, France. ACM.

[7] Twitter. Twitter API wiki.http://dev.twitter.com/doc, 2010.

[8] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao, 2010. *Detecting and characterizing social spam campaigns.* In Proceedings of the Internet Measurement Conference (IMC).

[9] C. Grier, K. Thomas, V. Paxson, and M. Zhang, 2010. *@spam: the underground on 140 characters or less.* In Proceedings of the ACM Conference on Computer and Communications Security (CCS).

[10] Kurt Thomasy, Chris Griery, Vern Paxsony, and Dawn Song. 2011. *Suspended Accounts in Retrospect:An Analysis of Twitter Spam.* IMC'11, November 2–4, 2011, Berlin, Germany. ACM.

[11] GAO, H., HU, J.,WILSON, C., LI, Z., CHEN, Y., AND ZHAO, B. Y, 2010. *Detecting and characterizing social spam campaigns.* In Proceedings of the 10th annual conference on Internet measurement (New York, NY, USA), IMC '10,ACM, pp. 35–47.

[12] ZADROZNY, B., LANGFORD, J., AND ABE, N, 2003. *Cost-sensitive learning by cost-proportionate example weighting.* In Proceedings of the Third IEEE International Conference on Data Mining (Washington, DC, USA), ICDM '03, IEEE Computer Society, pp. 435.

[13] Hongyu Gao, Yan Chen, Kathy Lee, Diana Palsetia and Alok Choudhary, 2011. *POSTER: Online Spam Filtering in Social Networks.* CCS'11, October 17–21, 2011, Chicago, Illinois, USA. ACM.

[14] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao, 2010. *Detecting and characterizing social spam campaigns.* In Proceedings of the Internet Measurement Conference (IMC),

[15] bit.ly. Spam and Malware Protection. 2009. http://blog.bit.ly/post/138381844/ . spam-and-malware-protection.

[16] HootSuite. Kapow! *HootSuite Fights the Evils of Phishing, Malware, and Spam.* 2010. http://blog.hootsuite. com/hootsuite-fights-malware-phishing/.

Authors Profile

R. Renuga Devi completed M.Sc., M. Phil. She is doing Ph.D (Full Time) in Karpagam University, Coimbatore. She published a paper in International Journal. And presented papers in national and International Conferences. Area of research is Social Networks, Data Mining and Neural Network.

Dr. M. Hemalatha completed M.Sc., M.C.A., M. Phil., Ph.D (Ph.D, Mother Terasa women's University, Kodaikanal). She is Professor & Head and guiding Ph.D Scholars in Department of Computer Science in Karpagam University, Coimbatore. Twelve years of experience in teaching and published more than hundred papers in International Journals and also presented more than eighty papers in various national and international conferences. Area of research is Data Mining, Software Engineering, Bioinformatics, and Neural Network. She is a Reviewer in several National and International Journals.