International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 06, June 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

A Comprehensive Model for Detecting Fake Profiles in Online Social Networks

Srinivas Rao Pulluri[1], Jayadev Gyani[2], Narsimha Gugulothu[3]

[1,2]*Jayamukhi Institute of Technological Science,Narsampet,Warangal,(India)*

[3]*JNTUHCEJ, Kondagattu,Jagityal,(India)*

ABSTRACT

Online social networking sites became an important means in our daily life. Millions of users register and share personal information with others. Because of the fast expansion of social networks, public may exploit them for unprincipled and illegitimate activities. As a result of this, privacy threats and disclosing personal information have become the most important issues to the users of social networking sites. The intent of creating fake profiles have become an adversary effect and difficult to detect such identities/malicious content without appropriate research. The current research that have been developed for detecting malicious content, primarily considered the characteristics of user profile. Most of the existing techniques lack comprehensive evaluation. In this work we propose new model using machine learning and NLP (Natural Language Processing) techniques to enhance the accuracy rate in detecting the fake identities in online social networks. We would like to apply this approach to Facebook by extracting the features like Time, date of publication, language, and geoposition.

Keywords: Social Networks, Threats, Fake Profiles, Malicious, Comprehensive, Machine learning and NLP

I. INTRODUCTION

Online Social Networks are most popular through which information can be exchanged throught the world. Social Networks being the center of attraction for many applications and they incorporate a range of new information and communication tools to the user community.A Social Network is best viewed as a graphical structure with nodes and edges depicting the users and their interaction activities respectively. The nodes and edges in a Social Network graph can be labeled or unlabeled depending upon the structure of the network being used. Because of the great reputation of social intelligence, social networking sites such as Facebook, YouTube, Twitter, LinkedIn, Pinterest,Google+, Tumblr and Instagram have become the preferred means of communication and information sharing tools amongst a diverse set of users including individuals and companies. The users of the social networks will play a vital role and they are completely responsible for the contents being exchanged in the networks. Users share information by interesting websites, videos and files. People share confidential data through the set-up of great faith and others have the same faith in the data shared. The rush of online social networks' reputation and the accessibility of huge amount of data enable them simple objective to the opponents. These objectives mainly include stealing individual user's details without seeking any permission.

The attempt for the encroachment of a legitimate user profile through fake identities is considered as the mostly practiced technique [1]. As the expansion of greater security in online social networking sites it turned to be very hard to encroach into online social networks. As a result of this, antagonists create false identities to gain access to other

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

profiles. According to Cloudmark report, around 20-40% of the Facebook accounts could be fake. Five new profiles are created every second[2], there are 83 million fake profiles[3]. Because of the openness of Facebook, users are likely disclosing many personal details about themselves and their friends as presented by [4].

Newly, online social networks have come up with a radical shift in our day to day life and transformed the world wide web into social web wherein user community is the centre for online growth, business, and information sharing [5].

The main objective of any Social Networking Site is to target different user segments. The best thing about Facebook is the ability to find old friends, but, YouTube provides a platform for people to connect, inform, and inspire others across the world by video sharing. For micro blogging we use Twitter. LinkedIn is famous for maintaining professional resumes with a large number of contacts. Facebook is the most popular online social network with 1.1 billion monthly unique users while YouTube becomes the second best social network with 1 billion monthly unique users [6]. Recently, Twitter stands in the next place as the global largest micro blogging online social networking site. As per the latest statistics Twitter has about 310 million monthly active users [6]. According to latest statistics [6], LinkedIn becomes the global largest Professsional network with 255 million monthly active users. Due to greater user contribution and increased online interactions, it is very tricky to identify anomalous user behavior, pervade in Social Networks and deviate them from the regular users. In contrast, some measures should be implemented to identify suspicious profiles and indicate the profiles as malicious. Nevertheless, this could not attain the required outcome. It happens to be more tedious job, due to enhanced security and privacy policies, limitations for collecting data sets. Hence it becomes very hard to differentiate false and real accounts. Albeit, the majority of the existing work aimed at malicious content, intrusion detection, spam distribution, and detection of profile cloning[7][1], now it is the time to pay extra attention for proposing new approaches and methods to distinguish legitimate profiles from fake profiles in a comprehensive manner.

The word fake is defined as a knock-off, or someone or something that isn't what it appears to be. A fake profile is a false identity of a user in the social network to pretend as legitimate user in the network. As said by researchers [8], the behavior and attitude of fake users is dissimilar to the real users. So, the amount and the type of information that a fake user passes to the profile pose a complete deviation from the legitimate user. Today there exist a number of different approaches for generating fake profiles. Amongst these some of the important approaches will include

Phase1- Constructing the personal profile [8], in order to attract the users to the profile [9] [10].

Phase2- Profile cloning [11][7] where spammers generate a parallel profile of the normal user in an online social network by using the legitimate user profile

Phase3-Fabricating a profile with a false identity as explained in [8].

Facebook is considered as the most popular Online Social Network, which permits most of the users to make profiles, upload pictures and audio/video files, send mails and remain in contact with friends, relatives and associates. Facebook has an outstanding increase with respect to other Online Social Networks. The main goal our work is identifying, separating the real profiles and false profiles in Facebook. The majority of the techniques used to counter this problem are centered at Facebook, YouTube, Twitter and LinkedIn. Most of the social networking sites have enriched and strong APIs to get related, in time and the latest user information as per the current research needs. Application Programming Interface (API) provides the access to user account information like user, friends' behavior and other essential user information.

Profile data in online networks can be static or dynamic. The details which are provided by the user at the time of profile creation is called static data, where as the details which are acknowledged by the system in the network is called

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

dynamic data. Static data contains demographic features of a user and his/her interests and dynamic data contains user runtime behavior and locality in the network [12]. The majority of current research depends on static and dynamic data. But this is not applicable to most of the social networks, where only some of static profiles are visible and dynamic profiles are not visible to the user community. Various techniques were proposed by different researchers to detect the fake identities and malicious content in online social networks. Each technique had its own merits and demerits.

As per the literature survey, the existing research associated to Facebook has not made any effort to identify features for spam detection at profile level. Albeit such features are extremely useful for detection and classification of spam profiles in Facebook.It is treated as an essential task to identify attackers who are spreading malicious content by means of fake profiles. Finding fake profiles requires a diverse set of static and dynamic features that define its behavior. Due to privacy concerns and strong restrictions on information availability [13], none of the existing research, used for to fake profile detection are viable to implement. Thus, in this work our main aim is to find a model for resolving legal profiles and false profiles in Facebook.

1.1 Information Extraction

Information extraction is the process of identifying key words and relationships in the text. It uses pattern matching algorithms to search for predefined sequences within the text. The software used for information extraction draws new relationships among known community and locations. And also it provides significant user information over time. These tools are extremely helpful while working with large amounts of data. Conventional data mining tools presume that the data being extracted is stored in the form of tables. Unfortunately, for most of the current applications, the only means of accessing digital information is by natural language processing [14].

1.2 Natural Language Processing

NLP is a field of artificial intelligence and computational linguistics concerned with how information systems can be used to understand and process natural language documents.NLP research main objective is to gather information about how people comprehend and use natural language to accomplish the required job [15].

The origin of Natural Language Processing depend on most promising areas, such as computer and information sciences, linguistics, artificial intelligence and robotics ,mathematics, psychology, etc. NLP applications include a number of research areas like natural language text processing machine translation, language information retrieval, summarization, user interfaces, multilingual and cross speech recognition, artificial intelligence, and expert systems etc.[15]

1.3 Pre-processing methods

Preprocessing techniques play a vital role in text mining. Preprocessing is the beginning step in the text mining approach. Preprocessing is done in three steps namely, stop words removal, stemming and lemmatization and Tokenization.

A. Extraction

Extraction is the process of mining structured information from unstructured data. In many cases extraction mainly concerns with processing human language documents by means of natural language processing.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

B. Eliminating Stop Words

Stop words elimination is one of the most promising pre-processing steps in NLP. The simple idea is removing all the words that occur commonly in the documents. Naturally, articles and pronouns are considered as stop words. These words will act as the divisions of natural language.In text mining Stop words are not considered as keywords and they can be deleted from the text [16].

C. Stop word removal

There are four basic methods used to eradicate stop words from the text [16].

i. The Classic Method: This approach is focused on eliminating stop words resulted from pre-compiled lists [17].

ii. Methods based on Zipf's Law: This method uses the law for eliminating most familiar words and eliminating words that occur just the once [17][18].

iii. The Mutual Information Method

It is a supervised learning approach which operates by calculating mutual information, providing an idea of how much data the object can describe regarding the specified class. Little amount of mutual data gives that the object has a low bias power and accordingly it should be eliminated [17] [18].

iv. Term Based Random Sampling

It was introduced by Rachel Tsz-Wai Lo et. al. to physically identify the stop words from web based documents. This approach operates by repeating more separate partitions of data which are selected randomly.the.After that ranking will be given to the partitions based on their syntax values using the Kullback-Leibler divergence measure [17].

D. Stemming

Stemming is the process used to identify stem, base or root form of a word. For example, the words wait, waits, waited and waiting can be rooted to the word "wait" [19]. The main objective of stemming is to confiscate a variety of suffixes, so that inflected words can be reduced and finally we can save time and space

The important points to be considered while using stemming are

The words which do not have similar meaning should be detached.

The stem need not be identical to the morphological root of the word.

These two conventions are very useful in language processing.

Stemming and lemmatization

Due to syntactic structures, most of the text documents use numerous forms of a word. Moreover, there exist a group of words which have same meaning. In most of the cases, it appears that it would be very useful for efficient searching of words to get related documents which hold one more word in the corresponding folder. The main aim of lemmatization and stemming is to diminish conjugate forms. But these terms diverge in their essence. Stemming typically defines to be a simple heuristic procedure that cut off the ends of words with the probability of achieving the target. The technique lemmatization generally describes the proper arrangement of words with the use of morphological analysis and vocabulary of words.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

E.Tokenization

It is the process of dividing the specified document into basic building blocks or other meaningful elements called tokens. These tokens will act as inputs for further processing .Tokenization accomplishes its job by positioning word margins. Another name for Tokenization is word segmentation. The problems encountered in tokenization totally rely on the language to be used.In addition tokenizations depend on the typographical structure and writing style of vocabulary.

Principal Component Analysis

PCA is applied to reduce the dimensionality of the dataset[20] . In this proposed work PCA plays an important position by giving the great endorsement to make decisions on which profile features to be used. Principal Component Analysis (PCA) is the simplest and robust dimensionality reduction technique ever seen. In this paper we have selected a mathematical model called variance maximization for drawing PCA results. According to this model "first principal component has the highest projection variance which is the direction in feature space along.And the second component defines the direction which has highest projection variance among all the other orthogonal direction to the first component". While calculating the score on profile features both false and real accounts to be measured.

II. RELATED WORK

The availability of data in social networks has drawn many research concerns encountered by online users. A Substantial research work has been carried out for a variety of social network problems like spam filtering, information diffusion and community detection. In[21] the researchers presented the possibility of information distribution without disclosing confidential data via graph models. In the research [22], they investigated "topological characteristics of Twitter" and presented that the behavior of information distribution in online networks known by examining "retweets".In the paper[23],they elaborated a report on "click-stream data in online networking sites" and proved that click-stream data provides a means to use rich information for drawing social relations .They also proved that most of user actions in online networks include browsing. Likewise in [24] the researchers examined social communications of user community in online social networks and initiated that most of the communications in online social networks are hidden and evident actions take place infrequently.

Various research efforts have been also turned towards the identification of malicious content in social networks. In the paper [25] the researchers planned a real-time spam detection scheme for Twitter social network in which they registered browser activities while loading a page for an URL.One more significant work for identifying spam on online social networking sites is proposed in [26] .In this proposed method, they created honey-profiles based on nationality, time ,age etc.This research is aimed at different accounts representing different areas . The researchers in [27] also made an attempt to use social honeypot to entice spammers on Twitter. In [28], the researchers described a significant attempt to typify various spams on Twitter social network.

The authors in[29] elaborated that the attacks of social junction were efficient and cheaper to get private information of an individual. Further in [30] the authors presented an approach to detect the users who are involved in malicious activities on Facebook. There are two stages in this mechanism. In first stage semantic analysis was done.In second stage spatiotemporal analysis was done.After that comparison was made between the original friend graph and the spatiotemporal graph.In [31] authors presented a brief study on various social networking platforms which provide

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

different privacy settings to secure user's personal information in network[32].Various protection mechanisms offered by Facebook to protect users from spammers ,hackers,or other threats are presented in [33].Further more Facebook has its own immune system[34].Social networking sites provide greater security to protect the users by implementing encrypted route to make sure that the registered user is genuine [34] and [35] the authors classified different spammers by using graph centrality .In addition,in [36] authors detected fake profiles based on some important features.First feature was time stamp of link creation and second feature was frequency of friend request. Likewise reflective policy assessment tools presented by the authors in [37] were observed different profiles from various directions. In [38] they proposed an app called My Page Keeper.There were number of attackers whose main objective was to add some malicious content on user's timeline.

To counter these attacks My Page Keeper was used.Similar to this app in [39] authors proposed FRAppE which was used to detect malicious applications on Facebook .To thwart spam profiles in [40] they presented a technique on the basis of social interaction in Facebook.Similarly in [41] authors created honey profiles on different social networking sites.Honey profiles were used to obtain data about malicious activities. The Algorithm Random Forest was applied on collected data and identified the URL ratio of the messages.

Even the newest techniques such as Honeypots anticipated to identify the spammers fail to draw abnormal users in most of the situations.Most of the researchers have used an unsupervised detection method PCA (Principal Component Analysis)]to detect the anomalous behaviour. In[42] the authors used the profile information of a user to detect fake identities in online social networks using specific supervised machine learning techniques for feature extraction and cluster building. The proposed technique is an efficient and faster means to identify fake profiles as it only uses the attributes given by a user during profile creation. This is the first technique to detect the clusters of fake profiles created by a user on a particular online social network

III. PROPOSED WORK

In our proposed model we would like to detect and classify fake users and real users using Machine learning and Natural Language Processing Techniques. In this work we present some steps that describe the process to distinguish fake users from legitimate users through a flowchart. Figure 1 show the working paradigm of our proposed work. In first step data set will be collected through Facebook using API. After the collection of data set in next step we are applying NLP pre processing techniques like Tokenization, Stop words removal, Stemming and lemmatization to extract features of users profile. After the extraction of the features we use reduction phase to reduce the dimensionality of the dataset. We would like to filter the profiles with tokenization, stemming and stop words to optimize the outcome before reduction.

For reduction we apply Principal Component Analysis(PCA) algorithm.After the reduction phase we apply appropriate learning algorithm for classification.At the end we would like to use evaluation parameters like True positive rate (TPR) ,false positive rate (FPR) and AUC(Area Under ROC Curve) for determining the legitimate users and the malicious users.

Depending on AUC we calculate True Positive Rate and False Positive Rate.If True positive rate is higher than False positive rate then it as a legitimate us will be considered as legitimate user otherwise it will be treated as a fake user.In our work we would like to propose a frame work using Machine Learning and NLP Techniques .We would like to apply this frame work for social networking sites like Facebook and others.The main objective of our model is to increase the accuracy in identifying the spam or fake identities in online social networks.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Currently we are working on building a model using Machine Learning and Natural Language Processing Techniques. We are planning to implement a spam detector using NLP in Python.The process of building efficient, scalable approach for big data services like Facebook, requires a representative sample of data for doing research and for drawing valid conclusions.Existing techniques related to spread and mitigation of malicious content on Facebook haven't been studied completely.

Most of the existing techniques for detecting malicious content of Facebook lack inclusive evaluation.The main objective of our research work is to increase the accuracy rate in identifying the fake profiles/malicious content in online social networking sites as compared to existing research. We would like to apply the proposed approach on Facebook.

IV. WORKING PRINCIPLE OF PROPOSED WORK

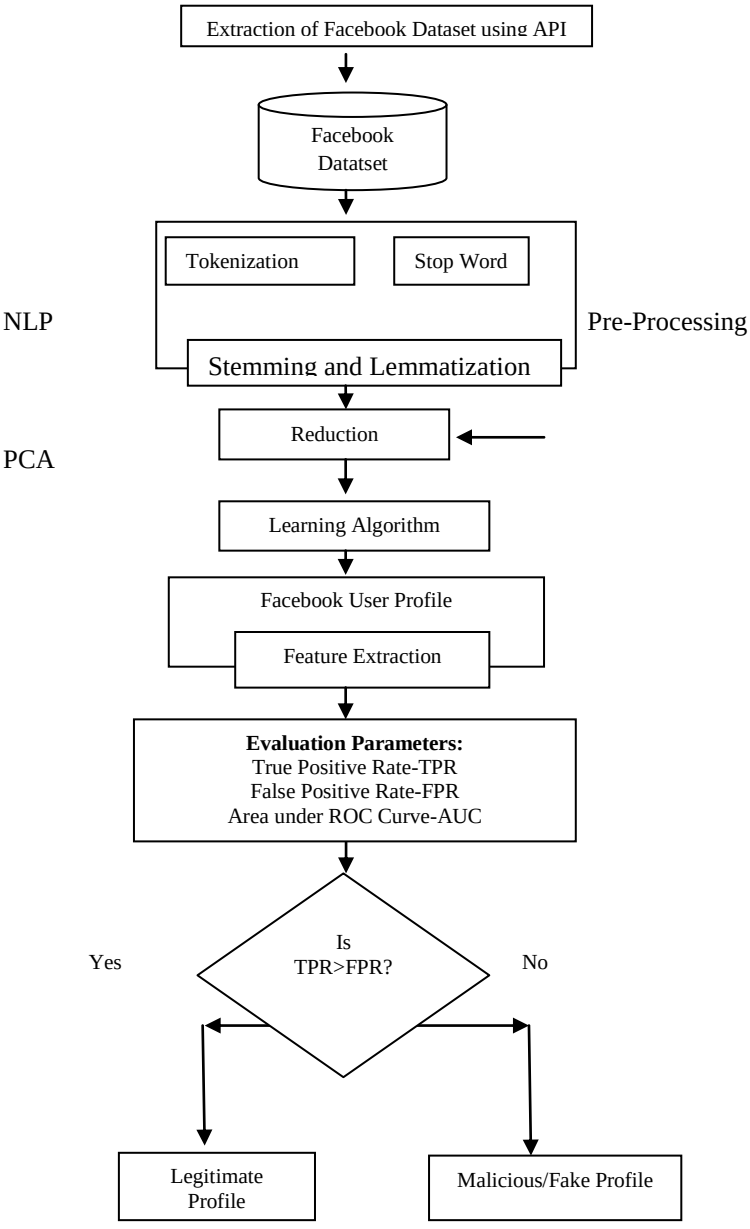

Figure-1: Working Paradigm of Proposed Work

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

V. CONCLUSIONS AND FUTURE WORK

Our proposed work presents various classification , mining and pre-processed techniques to detect malicious users , extract features, reduce the dimensionality and to prevent users from fake profiles. In this work pros and cons of each existing technique have been discussed and alternate method was proposed. After implementation of proposed technique using Machine Learning and Natural Language Processing concepts, we evaluate performance by TPR,FPR and AUC. We try to show that our techniques are most efficient as compared to existing research in terms of accuracy and performance by taking advanced features like geoposition and date of publication. In future work we concentrate on availability of social networking sites datasets and we try to propose a new enhanced model to identify fake profiles on social networks. We are planning to collect dataset using Facebook Graph search API.Important features to be considered: Time, date of Publication or posts, language and geoposition. We are also working on other attributes like typescript of user name, size, case and locality.

REFERENCES

[1.] Michael Fire et al. (2012). "Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies." Human Journal 1(1): 26-39.Günther, F. and S. Fritsch (2010). "neuralnet: Training of neural networks." The R Journal 2(1): 30-38.

[2.] Source:ALLFacebook-2012.

[3.] Source:CNN.

[4.] Saeed Abu-Nimeh, T. M. Chen, and O. Alzubi, "Malicious and Spam Posts in Online Social Networks," Computer,vol.44,no.9,IEEE2011,pp.23–28.

[5.] Howard Rheingold (2000). The Virtual Community: Homesteading on the Electronic Frontier, MIT Press.

[6.] Source:eBizMBA2017.

[7.] Kontaxis, G., et al. (2011). Detecting social network profile cloning. Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on,IEEE.

[8.] Krombholz, K., et al. (2012). "Fake identities in social media: A case study on the sustainability of the Facebook business model." Journal of Service Science Research 4(2): 175-212.

[9.] Cao, Q., et al. (2012). Aiding the detection of fake accounts in large scale social online services. Proc. of NSDI.

[10.] Bilge, L., et al. (2009). All your contacts are belong to us: automated identity theft attacks on social networks. Proceedings of the 18th international conference on World wide web, ACM.

[11.] Jin, L., et al. (2011). Towards active detection of identity clone attacks on online social networks.Proceedings of the first ACM conference on Data and application security and privacy, ACM.

[12.] Kazienko, P. and K. Musiał (2006). Social capital in online social networks. Knowledge-Based Intelligent Information and Engineering Systems, Springer.

[13.] Bradbury, D. (2011). "Data mining with Linkedin." Computer Fraud & Security 2011(10): 5-8.

[14.] Vishal Gupta and Gurpreet S. Lehal, A Survey of Text Mining Techniques and Applications, JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 1, NO. 1, AUGUST 2009.

[15.] S.Jusoh and H.M. Alfawareh, Natural language interface for online sales, in Proceedings of the International Conference on Intelligent and Advanced System (ICIAS2007). Malaysia: IEEE, November 2007,pp.224–228.

[16.] M.F. Porter, An Algorithm for Suffix Stripping, Program, vol. 14, no. 3, pp. 130-137, 1980.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

[17.] Ms. Anjali Ganesh Jivani, A Comparative Study of Stemming Algorithms, Anjali Ganesh Jivani et al, Int.J.Comp.Tech.Appl.,Vol(6),1930-1938,ISSN:2229-6093.

[18.] Deepika Sharma, Stemming Algorithms, A Comparative Study and their Analysis, International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868, Foundation of Computer Science FCS, New York,USA,Volume4–No.3,September2012–www.ijais.org.

[19.] C.Ramasubramanian and R.Ramya, Effective Pre-Processing Activities in Text Mining using Improved Porter‟s Stemming Algorithm, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 12, December 2013, ISSN (Online) : 2278-1021.

[20.] Shalinda Adikari and Kaushik Dutta, IDENTIFYING FAKE PROFILES IN LINKEDIN, PACIS 2014 Proceedings,AISeL

Sala, L. Cao, C. Wilson, R. Zablit, H. Zheng, B. Zhao, Measurementcalibrated graph models for social network experiments, in: Proceedings of the 19th International Conference on World Wide Web, ACM,2010,pp.861–870.

[21.] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media?, in: Proceedings of the 19th International Conference on World Wide Web, ACM, 2010, pp. 591–600. 26

[22.] F. Benevenuto, T. Rodrigues, M. Cha, V. Almeida, Characterizing user behavior in online social networks, in: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, ACM, 2009, pp. 49– 62.

[23.] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, B. Zhao, Understanding latent interactions in online social networks, in: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ACM, 2010, pp. 369–382.

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

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

[26.] K. Lee, B. Eoff, J. Caverlee, Seven months with the devils: A longterm study of content polluters on twitter, in: Proceedings of the AAAI Conference on Weblogs and Social Media (ICWSM), 2011.

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

[28.] Puttaswamy KPN, Sala A, and Zhao BY,” Starclique: Guaranteeing user privacy in social networks against intersection attacks”, in: Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, CoNEXT ’09. ACM 2009, New York, NY, USA, pp.157-168

[29.] Z. Halim, M. Gul, N. ul Hassan, R. Baig, S. Rehman, and F. Naz,“Malicious users’ circle detection in social network based on spatiotemporal co-occurrence,” in Computer Networks and Information Technology (ICCNIT),2011 International Conference on, July, pp. 35–390.

[30.] Liu Y, Gummadi K, Krishnamurthy B, Mislove A,” Analyzing Facebook privacy settings: User expectations vs. reality”, in: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference,ACM,pp.61–70.

[31.] Mahmood S, Desmedt Y,” Poster: preliminary analysis of google?’s privacy. In: Proceedings of the 18th ACM conference on computer and communications security”, ACM 2011, pp.809–812.

[32.] Stein T, Chen E, Mangla K,” Facebook immune system. In: Proceedings of the 4th workshop on social network systems”, ACM 2011, pp.1-8.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

[33.] Kuzma J," Account creation security of social network sites", Inter J Appl Sci Technol 1(3):2011, pp. 8–13.

[34.] Debar D, Wechsler H," Using social network analysis for spam detection", In: Proceedings of the third international conference on social computing, behavioral modeling, and prediction (SBP'10). Springer-Verlag, Berlin, Heidelberg 2010, pp. 62–69.

[35.] Cukierski WJ, Hamner B, Yang B,"Graph-based features for supervised link prediction. In: IEEE International Joint Conference on Neural Networks (IJCNN)", IEEE 2011, pp. 1237–1244.

[36.] Anwar M, Fong PW,"A visualization tool for evaluating access control policies in Facebook-style social network systems", In: Proceedings of the 27th annual ACM symposium on applied computing, ACM 2012, pp. 1443–1450.

[37.] Rahman M, Huang T, Madhyastha H, Faloutsos M," Efficient and scalable socware detection in online social networks", In: Proceedings of the 21st USENIX conference on security Symposium 2012, USENIX association, pp. 32–32.

[38.] Rahman MS, Huang TK, Madhyastha HV, Faloutsos M,"Frappe: detecting malicious Facebook applications", in: Proceedings of the 8th international conference on emerging networking experiments and technologies,ACM2012,pp.313–324.

[39.] F. Ahmed and M. Abulaish, "An MCL-Based Approach for Spam Profile Detection in Online Social Networks," IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications2012,pp.1–7.

[40.] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in ACSAC '10: Proceedings of the 26th Annual Computer Security Applications Conference. ACM Request Permissions, 2012,pp.1–9.

[41.] Cao Xiao,David Mandell Freeman and Theodore Hwa,Detecting Clusters of Fake Accounts in Online Social Networks,ACM-2015,ISBN-978-1-4503-3826-4/15/10

[42.] F. Ahmad and M. Abulaish, A Generic Statistical Approach for Spam Detection in Online Social Networks,

[43.] Computer Communications, 36(10-11), Elsevier, pp. 1120-1129, 2013

[44.] Shalinda Adikari and Kaushik Dutta, IDENTIFYING FAKE PROFILES IN LINKEDIN,AISeL,PACIS-2014