Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 3, Issue 4; April-June, 2016, pp. 262-265

http://www.krishisanskriti.org/Publication.html

Feature Extraction and Content Investigation of Facebook Users' using Netviz and Gephi

Ekta[1], Sanjeev Dhawan[2] and Kulvinder Singh[3]

[1]M.TECH(Software Engineering) University Institute of Engineering and Technology Kurukshetra University Kurukshetra
[2,3]Department of Computer Science & Engineering University Institute of Engineering and Technology Kurukshetra University Kurukshetra
E-mail: [1]ekta.aterja07@gmail.com, [2]rsdhawan@rediffmail.com, [3]kshandha@rediffmail.com

Abstract—*Facebook is widely used by number of user for their communication purpose. In Facebook application user can make new friends and share his or her views and also comments or likes on their user views. People are getting addicted to Facebook and they are adding more and more new friends. Sometimes these new friends can be malicious. Malicious user can access user personal information to perform malicious activity like sending spam message to personal user for its own purpose so there should be a mechanism in which these malicious users can be distinguished from normal users. In this paper, Facebook page data is analyzed by using Gephi and Facebook network data is collected using Netviz. Different groups are identified in this page dataset those having same like and interest.*

Keywords*: Facebook, Gephi, Netviz, Social Network, Malicious Users.*

1. INTRODUCTION

Social Networks are becoming a best source of communication and provides a medium to share photos, videos etc. Social Networks make personal information visible to others. Social Networks also provides privacy and security for its users. In many Social Networking sites user can choose their privacy settings and user can exchange information with the specific audiences. There are so many Social Networking sites like Facebook, Linkedin, Myspace, Instagram etc. People are using these Social Networking sites in their day to day life. Mostly people use these Social Networks to communicate with their loved ones, colleagues, neighbors etc. Facebook is one of the most powerful and most popular Social Networks. Facebook has become necessity of individual's life. People like to share their experiences, text messages, photos, videos and most importantly their emotions. Facebook is the platform where people can make new friends with ease and communicate with their friends those are far away from them. It also provides a platform for the organizations to advertise their products by creating pages of their products and services. Facebook also provides various features like create events, create groups etc. In past few years Facebook becomes most popular Social Network among the youngsters. There are also challenges regarding security and privacy. In Facebook hackers or spammers try to steal user's personal information for their personal use. Hackers create fake profiles and use these profiles to communicate with authenticated users for stealing their information. Spammers send malicious links through messages. Mostly fake profiles are used for sending the spam and stealing the information of the users. Hackers also develop the malicious applications (Apps) and convince the users to use these Apps. Once user use malicious Apps personal information is shared with the hackers and then hackers use this information for their personal use. Therefore, it is required to detect these fake profiles and prevent users from these kinds of attacks. It is needed to analyze user's Facebook personal network.

2. RELATED WORK

Much work has been done in an effort to understand the complexities of social networks and considerable information has been obtained. David *et al.* [1] provided algorithms for different Social Network Analysis and these Social Network Analysis tools were compared which execute the algorithm. Apart from this, Kurant *et al.* [2] discussed about Facebook properties and also provided how to achieve improved impartial sample by using Metropolis Hasting Random Walk(MHRW) and Reweighted Random Walk(RWRW) technique in Facebook Subgraph. Furthermore, other techniques in Social Network Analysis and Data mining provided different issues in terms of Business Application which was described by Giones *et al.* [3]. In similar way, two techniques that is Binary First Search (BFS) and Uniform (UNI) was used to detect relation between users in Social Network Analysis (Alessandro *et al*. [4]). In another research Geetanjli *et al.* [5] provided visualization of Facebook Social Network which characterizes larger number of friends that reveal unknown connection in Facebook Network. Similarly, Maharaj *et al.* [6] analyzed and visualized the connection among different communities with the help of Gephi. Authors collected dataset using Netviz application on Facebook. Furthurmore, comparative analysis of different fake profile

detection technique with their pros and cons discussed by Dr. Sanjeev Dhawan and Ekta [7]. In another research, Ekta and Dr. Sanjeev Dhawan[8] classified Diabetes dataset for predicting how many people is infected from diabetes by using NaïveBayes classifier and Ranker algorithm.

3. GEPHI TOOL

Gephi is an open source software tool which can be used for the graph and network analysis. It can be useful to process complex dataset and produce visual results. Gephi can be easily downloaded from www.gephi.org [9]. It provides many layouts for the visualization of the dataset. Gephi also processes data in CSV format.

To import dataset in Gephi tool Netviz application is used which is easily available on Facebook. With the help of Netviz user's personal data, group data, liked pages data on Facebook can be downloaded easily.

4. ANALYSIS OF FACEBOOK DATASET

Netviz application is used for the extraction of the Facebook network data like personal network, page data, group data etc. In order to analyze the Facebook dataset it can be imported to Gephi. In order to analyze the page data is collected from the Facebook which contains 501 nodes and 1296 edges. For the visualization ForceAtlas2 layout is selected. Fig. 1 shows the data which is imported in Gephi. Modularity is set for getting information of the data. Modularity in this dataset is 0.377 and number of communities are 14.

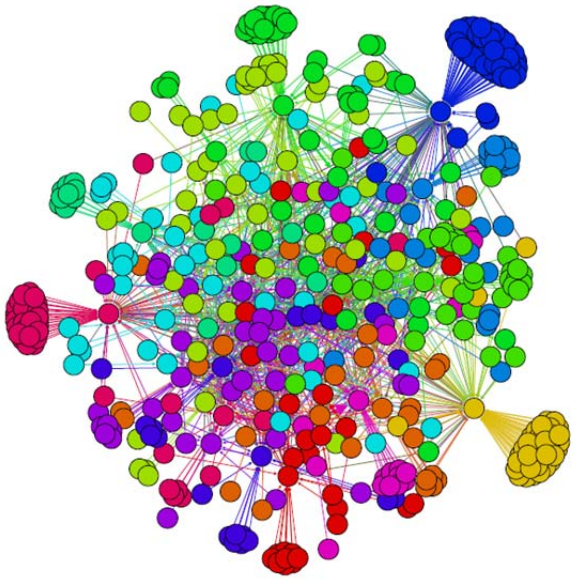

Fig. 1: Imported Facebook page Dataset in Gephi

Graph Modularity

Modularity: 0.377

Modularity with resolution: 0.377

Number of Communities: 14

Resolution: 1.0

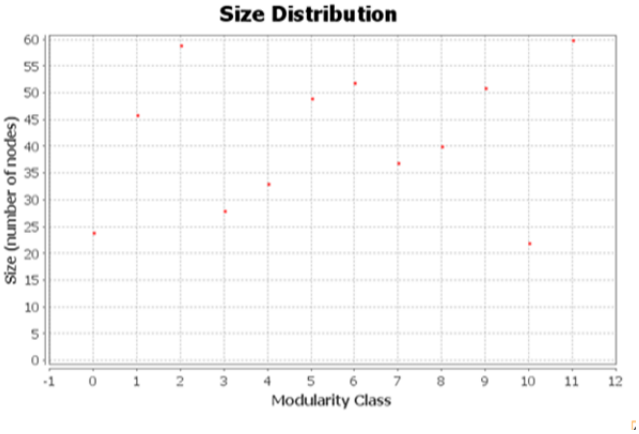

Fig. 2: Size Distribution

Graph Distance Report

Diameter: 6
Radius: 3
Average Path length: 3.427688622754491
Number of shortest paths: 250500

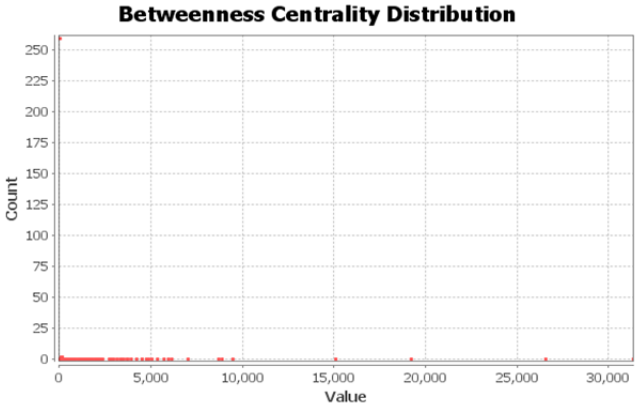

Fig. 3: Betweenness Centrality

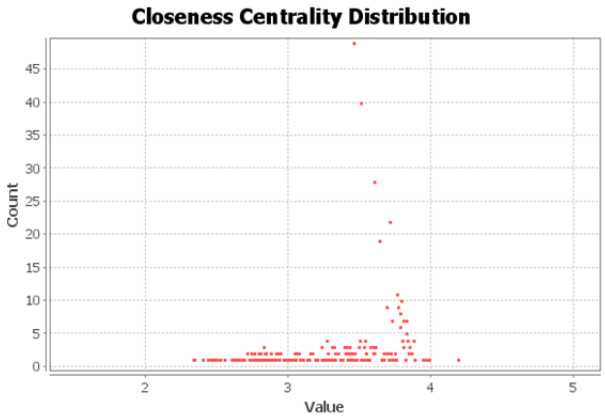

Fig. 4: Closeness Centrality

Fig. 2 shows the size distribution in the dataset and it contains modularity class and size of the number of nodes. Fig. 3 shows Betweenness centrality which shows diameter, average path length and shortest path. Betweeness centrality can be used to measure the connectivity of neighbors node. It can also measured the indirect connectivity of nodes through their direct links. It can be measured using this formula [10].

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Fig. 4 shows closeness Centrality. Closeness Centrality reflects the ability to access information through the "grapevine" of network members. Thus, closeness is the inverse of the sum of the shortest distances between each individual and every other person in the network. The shortest path may also be known as the "geodesic distance". It can be measured using this formula [11].

$$C_C(v) = \frac{1}{\sum_{t \in V \setminus v} d_G(v,t)}$$

Degree range can be set for the analysis of the Facebook page dataset and the grouping of the users. Those nodes satisfy this set degree range can be shown using graph representation as shown in Fig. 5.

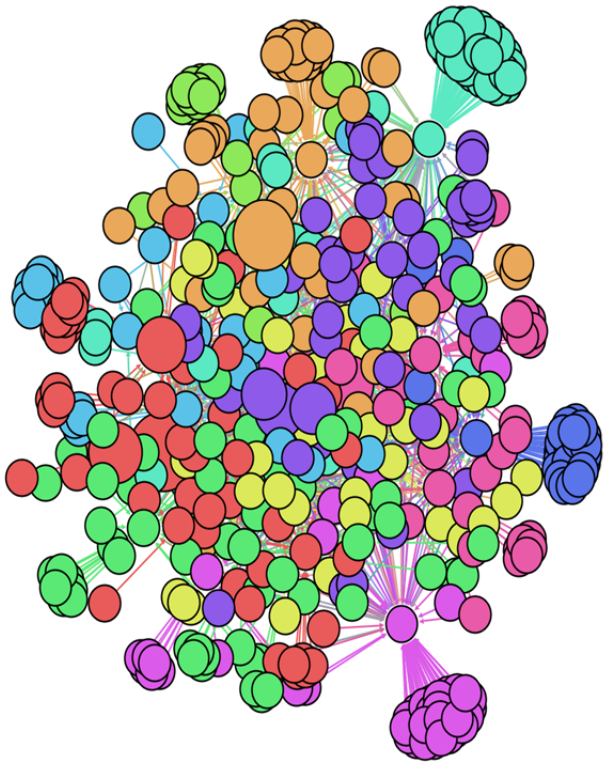

Fig. 5: Visualization of the groups in Facebook page dataset

5. PROPOSED WORK

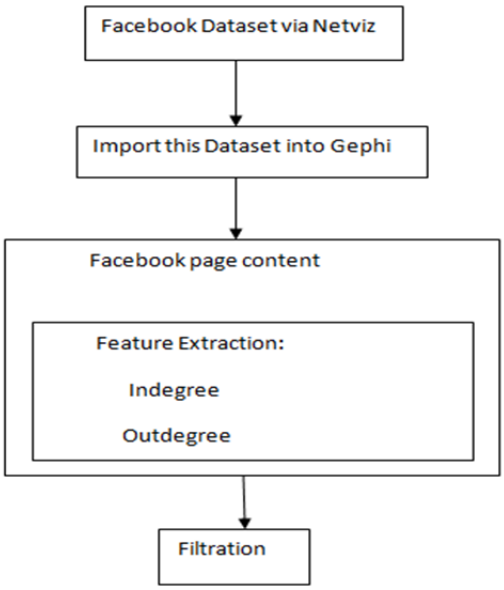

Fig. 6: Proposed Work

In this proposed method there is Netviz application is used for the extraction of data from the Facebook. When user's data is collected from the Facebook then it can be imported to the Gephi. Gephi can process the data and there different parameters will be used for the processing of the data. For the feature extraction there is indegree and outdegree will be used. Then filtration can be done on the basis of extracted features.

6. CONCLUSION

Identification of malicious or fake users is challenging task for researchers. They proposed various mechanisms to detect them. In this paper Facebook page dataset was collected by using Netviz application. This page dataset analyzed in Gephi and grouped different users. Gephi is open source tool that is specially design to analysis Facebook data. For the visualization of page dataset ForceAtlas2 layout was used. Grouping of the users was done in effective way using different parameters in Gephi.

REFERENCES

[1] David Combe, Christine Largeron, Elod Egyed-Zsigmondand Mathias Gery. "A Comparative study of social network analysis tool." *International Workshop on Web Intelligence and Virtual Enterprises 2*, 2010, pp. 1-7.

[2] Minas Gjoka and Maciej Kurant and Carter T. Butts and Athina Markopoulou, "Walking in Facebook: A Case Study of Unbiased Sampling of OSNs", *Proceedings of IEEE INFOCOM '10, San Diego*, CA, March, 2010, pp. 1-4.

[3] Francesco Bonchi, Carlos Castillo, Aristides Gionis and Alejandro Jaimes, "Social Network Analysis and Mining for Business Application", *ACM Trans. Intell. Syst. Technol*. 2, 3, Article 22 (April 2011), pp. 1-5.
[4] Salvatore A. Catanese, Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, Alessandro Provetti, "Crawling Facebook for Social Network Analysis purpose" *WIMS'11, May 25-27, 2011 Sogndal, Norway*, pp. 1-6.
[5] Nadeem Akhtar, Hira Javed, Geetanjali Sengar."Analysis of Facebook Social Network", *5th International Conference on Computational Intelligence and Communication Networks*, 2013 IEEE, pp. 451-454.
[6] Kambale Vanty Muhongya and Manoj Sewak Maharaj, "Visualising and Analysing Online Social Networks", *IEEE*,2015, pp. 1-6.
[7] Dr. Sanjeev Dhawan, Ekta, " Implications of Various Fake Profile Detection Techniques in Social Networks", *IOSR Journal of Computer Engineering (IOSR-JCE) AETM'16* ,2016, pp. 49-55.
[8] Ekta, Dr.Sanjeev Dhawan, "Classification of Data Mining and Analysis for Predicting Diabetes Subtypes using WEKA", *Vivechana: National Conference on Advances in Computer Science and Engineering (ACSE-2016)*, pp. 1-5.
[9] https://gephi.org/
[10] Betweenness centrality, https://en.wikipedia.org/wiki/Betweenness_centrality, Accessed on 11-May-2016.
[11] Closeness centrality, https://en.wikipedia.org/wiki/Centrality, Accessed on 11-May-2016.