IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-10, May 2019 E-ISSN: 2347-2693

Reaction Based Approach to find Malicious Posts in Online Social Networks

Rubina khan[1*], Anshul khurana[2]

[1,2]Dept. of Computer Science, Shri Ram college of Institute & Technology, RGPV University, Jabalpur, India

Corresponding Author: rubina273khan@gmail.com, Tel.: 8989522110

DOI: https://doi.org/10.26438/ijcse/v7si10.144148 | Available online at: www.ijcseonline.org

***Abstract*—** Social network is a platform of connected peoples where peoples share relation, emotions, activities etc. Second generation social network come in existence with lots of emerging applications, which also support service oriented environment, during all kind of activities massive information is generated. These all information are in the form of post. Hence it is necessary to find category of post. A post may be legitimate or malicious. In this dissertation we are trying to find malicious post on the basis of reaction and share on particular post. All the post collected by facebook through app known as Netvizz. Entire concept implemented in R studio which is an IDE of R programming. Dissertation also contain statistical analysis of page network with the help of well known tool gephi, It is based on predefine parameter such as Eigen vector centrality,closeness ,betweenness etc.

***Keywords*—**Facebook,Netvizz,Gephi,Visualization.

I. INTRODUCTION

A social networking site is a website where every user has its own profile and has contacts with its friends, family members, employees, share their updates, and join new communities and groups having same interests. Most of these sites allow users to post blog entries, search for others with similar interests and share lists of contacts. User profiles often have a section dedicated to comments from friends and other users [1]. To protect user privacy A social, social networks typically have controls, that allow users to choose who can view their profile, contact them, add them to their list of contacts, and so on.

1.1 Social Graphs

Graph for OSN (ONLINE SOCIAL NETWORKING) is a representation of the interconnection of relationships in an online social network[4].

The social graph has been referred to as "The global mapping of everybody and how they are related"

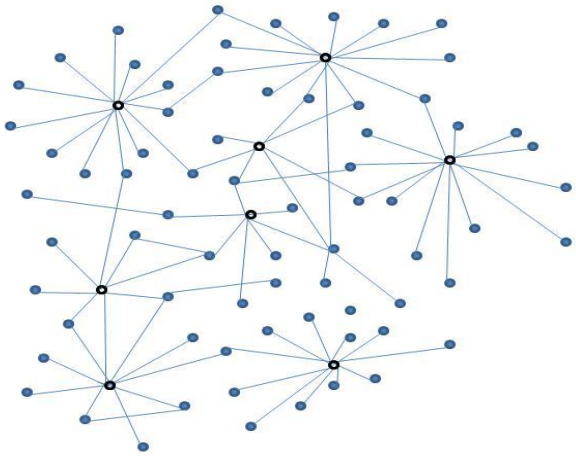

Figure:1 Online Social Graph

Second generation social networks comes in existence with lots of emerging applications, which create service oriented environment[2]. In a simple service oriented environment there are many recommender between service requester and service provider [7]. Hence trust is big factor of service Consumer and recommender plays big role for trustworthy service. Service requesters always require the evaluation of the trustworthiness and satisfaction of service provider.

1.2 Trust in Online Social Network-.
Second generation social networks comes in existence with lots of emerging applications, which create service oriented environment[1]. In a simple service oriented environment there are many recommender between service requester and service provider[12] .Hence trust is big factor of service Consumer and recommender plays big role for trustworthy service[3] Service requesters always require the evaluation of the trustworthiness and satisfaction of service provider.

II. MOTIVATION

Number of users are interconnected in OSN with each other by various communities groups, pages and posts, activities etc. People are connected to each other with same kind of interest.[5] In Online social networks like facebook groups there are different members are communicating with each other. Let us take an example, Facebook group in which admin post related to group and user also post to the group but this post posted by user may be irrelevant to that Face book group .A post may become irrelevant due to less number of Likes, less number of Share etc. To identify these malicious posts is a difficult task.Face book is easy way to connect communication among many users[4]. User engages to each other with lots of activities for Effective communication. Create of page, group or event is effective way to better communication, but some time user can make unnecessary content on page and group. This type of content in the form of mostly in malicious post[9]. To find malicious post in a community or page is challenging task. In this dissertation. We apply enhance reputation base approach for detection of malicious post, decision of malicious is base on reputation index. data is collected by face book with the help of netvizz app, for the purpose of implementation we use R programming language with R studio and visualization of data we use gephi**.**

The motivation of finding malicious post based on share and like, reputation index. Decision policy for reputation index is based on manually reputation selection, this is a key point for motivation .Hence this concept may apply to calculate malicious post in online social network.

III. PROPOSED METHODOLOGY

In this dissertation we are trying to find malicious post on the basis of reaction and share on particular post. All are post collected by face book through app known as Netvizz. Entire concept implemented in R studio which is a IDE of R programming. Dissertation also contain statistical analysis of page network with the help of well known tool gephi,It is based on predefine parameter such as Eigen vector centrality, closeness, betweens etc.

Terms in use-

Table: 1 Activities in Different Social Networks

OSN	Activity (direct)	Activity(indirect)
facebook	like, comment, tag, recommendation	share, recommendation
Orkut	like, comment, recommendation	share, recommendation
Twitter	Comment, tag	Re tweet

a.Malicious Post-There are many possibility related to malicious post but in our context a post with less number of activates
b.Legitimate Post- There are many possibilities related to malicious post but in our context a post with more number of activities.
c.Reputation Index- It is a threshold value for the selection of post category.
d.Likeshare- It is the sum of Total number of activities on the post.

Figure:2 Facebook Reactions

Table:2 Different Activities

Activity category	Sub Category
Sharing	Wallpaper(W_l), Photo (W_p), Video share(W_v)
Chat	Chat (Wc) Video-Chat(W_{vc})
Game playing	Playing-online- games(W_g)
Personal information	Messaging(W_m) Recommendation(W_r)

// Algorithm for malicious post**//**

1 Data collection from online social networking [Netvizz is used]

2 Import data set with selected platform

3 Data Cleaning and Preprocessing
(a)Remove unnecessary information
(b) Remove NA values
(c) Apply factoring or Data scaling

4 Calculate Reputation
(a) Likeshare = reactions on posts + share on posts
(b) Reputaion = likeshare / comments on posts.

5 Selection of Threshold Value (Reputation index) [Select manually or select on the basis of standard deviation or mean value]

6 for (1 to number of post if reputation<=reputation index)
Then post is malicious
Else post is legitimate

The Algorithm shows that for posts to be legitimate the reputation of the posts should be greater then reputation index. In our context the reputation index is in between 1 to 2.
Figure of flow chart based on used technique is shown below which represent different steps for this research and analysis.

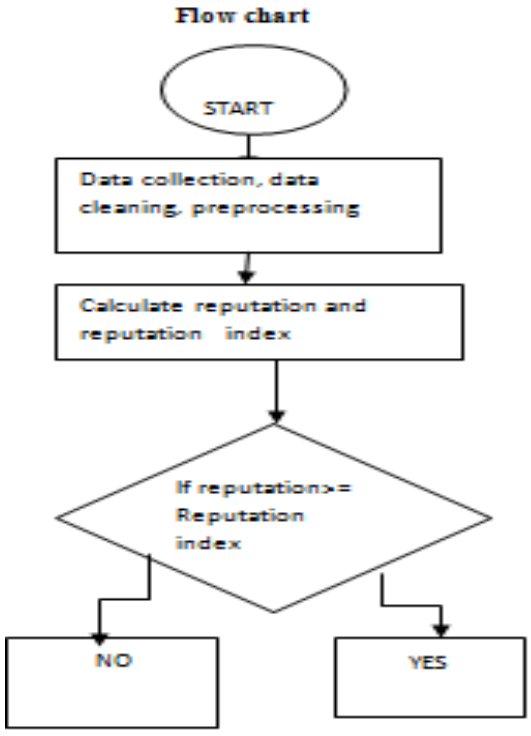

Fig. 3

Gephi statistics – Gephi is open source software which is mostly use in network visualization. It help to visualized many data sets specially online social networks. Complex set of networks can easily visualized and analysis with the help of gephi.
Following statistics can easily calculate by gephi:-
Global and Local measure-. Global centrality measures require complete topological information for their computation, whereas local centrality measures only require local topological information from neighboring nodes.
Betweenness: It measures the fraction of shortest paths passing through a single node relative to the total number of shortest paths in the network.
Closeness centrality measures the mean distance from a vertex to other vertices.
Closeness - Closeness is a global measure of geodesic distance of a node to all other nodes.
Betweenness centrality It is measures the extent to which a vertex lies on paths between other vertices.

IV. IMPLEMENTATION & ANALYSIS

- **Data Collection-**

Data is collected with the help of social networking API. There are many API for collection of data from online social networks.
Netvizz-
Netvizz is one of application which helps us to collect data from faceook. Netvizz is a tool data collection and extraction in user's face book account. It allows to researches to export data in stranded file format.

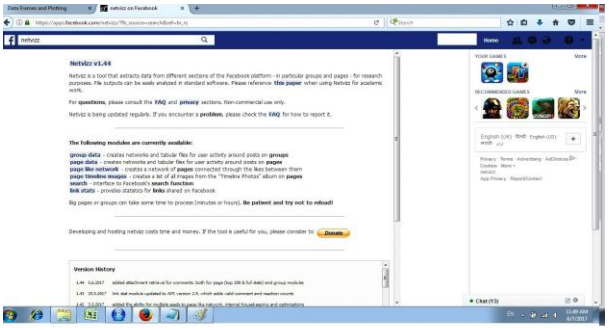

Fig. 4

- **Data Cleaning and Preprocessing**
  Collected data is form of raw data, it is necessary to convert data in the form of tabular format and remove unnecessary redundancy of data. In our data unnecessary information such as post url, comments, etc are removed before importing data in r studio.

(a)R and R Studio-
R is powerful programming tool in data implementation ,R studio is an integrated environment which provides all packages in single platform. Feature of R describe in following ways.

Post=949	Legitimate=805	Malicious=144	15.17 % Malicious

(i)R is known comprehensive statistical platform, which is used fordata visualization, data analysis.
(ii)R contain advanced statistical package.

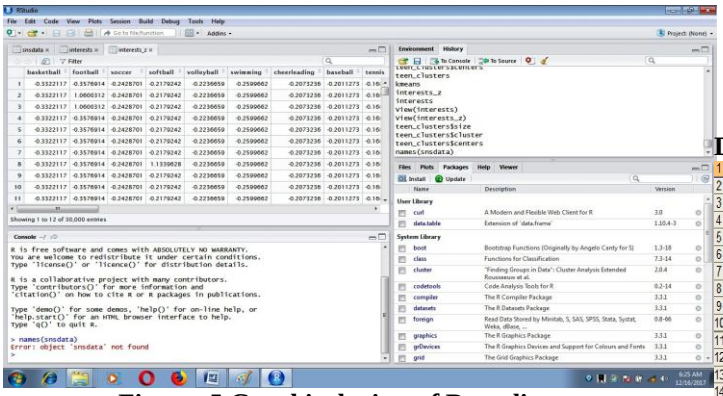

Figure: 5 Graphical view of R studio

Gephi- Gephi is open source software which is mostly use in network visualization. It help to visualized many data sets specially online social networks. Complex set of networks can easily visualized and analysis with the help of gephi.

- **Reputation Selection-** In our work Histogram is plotted which shows how reputation values is spared.According to diagram most of values between1 to 2.

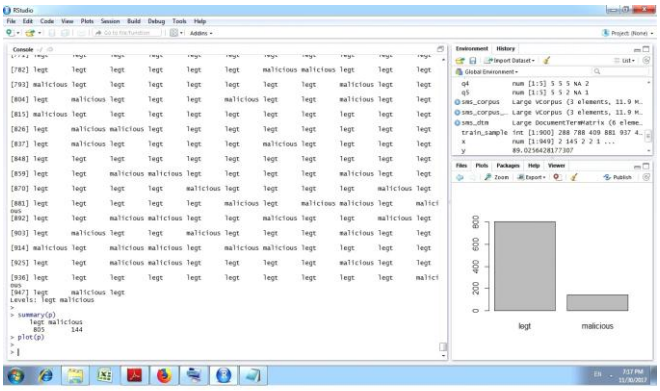

Figure: 6

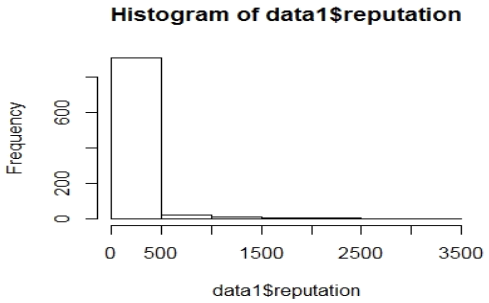

Figure:7 Histogram of reputation

Data Result-

1	3	3	3	3	3	3	3	3	3	3	3	3
2	2	56	874	1987	16	2315	8	2339	2323	145.1875	145.1875	legt
3	3	59	873	0	1	1	1	1	2	2	2	legt
4	3	59	872	0	1	1	1	1	2	2	2	legt
5	4	59	871	0	0	0	1	1	1	1	1	malicious
6	2	59	870	0	0	2	0	2	2	1	1	malicious
7	3	59	869	0	1	1	1	1	2	2	2	legt
8	3	59	868	0	1	1	1	1	2	2	2	legt
9	3	59	867	0	1	1	1	1	2	2	2	legt
10	4	56	866	203	17	242	0	259	242	14.23529	14.23529	legt
11	4	56	865	6397	274	10147	519	10940	10666	38.92701	38.92701	legt
12	3	59	864	0	1	1	1	1	2	2	2	legt
13	2	56	863	2444	44	3176	71	3291	3247	73.79545	73.79545	legt
14	3	59	862	0	1	1	1	1	2	2	2	legt
15	3	59	861	0	1	1	1	1	2	2	2	legt
16	3	59	860	1	1	1	1	1	2	2	2	legt
17	3	59	859	2	1	1	1	1	2	2	2	legt
18	3	59	858	0	1	1	1	1	2	2	2	legt
19	3	59	857	0	1	1	1	1	2	2	2	legt
20	4	56	856	8358	168	11323	495	11986	11818	70.34524	70.34524	legt
21	2	56	855	3859	53	4796	86	4935	4882	92.11321	92.11321	legt
22	1	59	854	0	0	0	0	0	0	1	1	malicious
23	3	59	853	0	1	1	1	1	2	2	2	legt
24	3	59	852	0	1	1	1	1	2	2	2	legt
25	3	59	851	1	1	1	1	1	2	2	2	legt

Figure: 8 Result Data

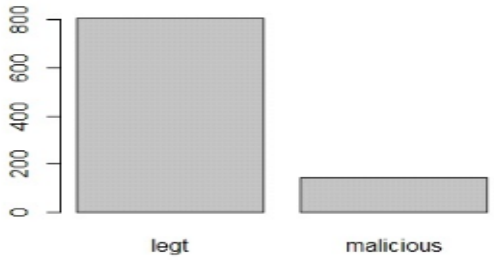

Figure: 9 Bar graph

These two diagrams are description about result, bar graph present legitimate and malicious post overall **15.17% are malicious** and **84.9% are legitimate.** Next is final result data sheet Last column in sheet represent category of post (legitimacy, malicious).

GEPHI ANALYSIS

Gephi analysis makes it eay to analyze social network easy creation of social data connectors to map community organizations and small world networks.

A. Betweenness

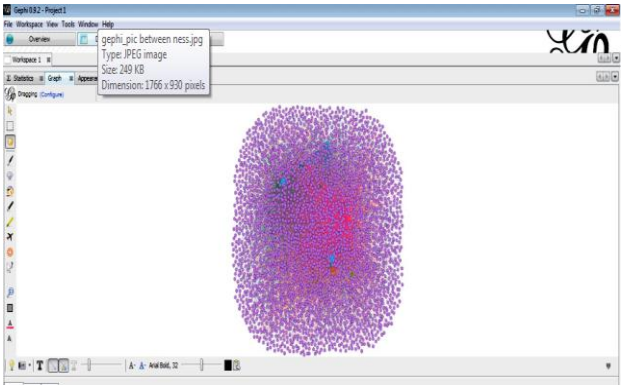

Figure:10 Betweenness in network

B. Cluster Cofficient

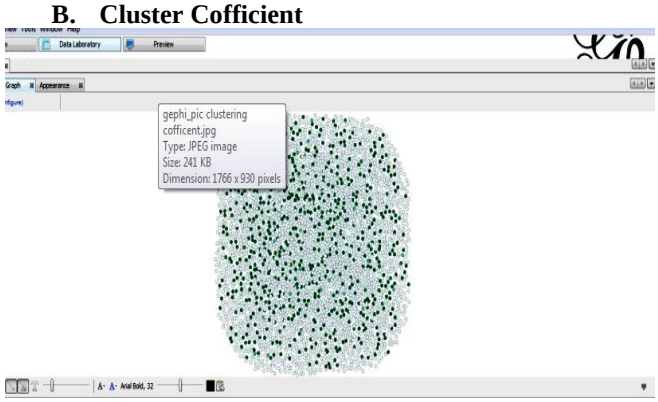

Figure:11 Cluster Coefficient in Posts network

C. Degree

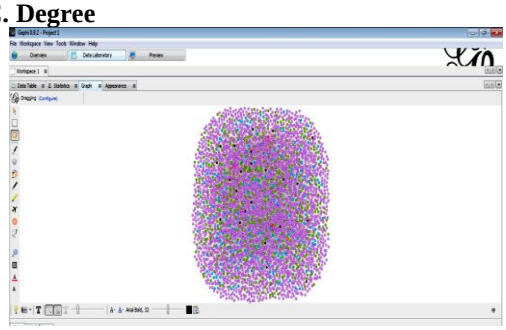

Figure:12 Degree Distribution in Post Network

Summary-This chapter is all about result and discussion, chapter is divided in two part first part related to 949 posts which is collected through netvizz data collection tool in face book, and R programming in use as implementation platform. Result comes in terms of legitimate and malicious post. Next result related to gephi tool. Predefine statistics are applied on data such as cluster coefficient, betweens, closeness, Eigen vector centrality etc Result comes in the terms of Network visualization.

V. CONCLUSION & FUTURE SCOPE

CONCLUSION

This dissertation is a novel approach to find malicious post in online social network.There are huge number of post generated in facebook. It is mandatory to classify posts. There are two major possibilities either post may be legitimate or malicious. This work distinguished posts in the terms of reaction or share .It seems malicious post has less number of like share in comparison to legitimate post. We have collected total 949 posts from different pages to face book, and try to classify. For the purpose of implementation we use R and R studio programming and for predefine statistics analysis gephi is used.

FUTURE SCOPE

Here we have try to distinguished posts in terms of reaction and share, but for the purpose of deep understating of malicious post it is necessary to merge this approach to text mining and sentiment analysis. So that it will be possible to analyze post with positive and Negative attributes.

REFERENCES

[1]. Broder, A., Glassman, S., Manasse, S., and Zweig, G. "Syntactic clustering of the web." (WWW6'97) (Santa Clara, CA., April). PP 391–404, 1997.

[2]. Shivakumar, N. And Garica-Molina, H. "Finding near-duplicates of documents on the web." Web Databases (WebDB'98) (Valencia, Spain, March). PP 204–212,1998.

[3]. Heintze, N. "Scalable document fingerprinting." USENIX Electronic Commerce Workshop (Oakland, CA., November). PP 191–200,1996.

[4]. Sanderson, M. "Duplicate detection in the Reuters collection." Technical Report of the Department of Computing Science at the University of Glasgow, Glasgow G12 8QQ, UK,1997.

[5]. McCain, M., and William C. "Integrating Quality Assurance into the GIS Project Life Cycle", ESRI Users Conference 1998.

[6]. Goodchild, M., and Gopal, S. (Eds.), "Accuracy of Spatial Databases", Taylor & Francis, London, ISBN: 0-85066-847-6, 1989.

[7]. Scott L., "Identification of GIS Attribute Error Using Exploratory Data Analysis," Professional Geographer 46(3), PP 378.386. 23 FEB 2005.

[8]. FGDC Federal Geographic Data Committee, FGDC-STD- 001-1998. "Content standard for digital geospatial metadata (revised June 1998)," Federal Geographic Data Committee, Washington, D.C., 1998.

[9]. C. Policroniades and I. Pratt, "Alternatives for Detecting Redundancy in Storage Systems Data," USENIX Annual Technical Conference, Boston, MA, USA, June 2004.

[10]. A. Muthitacharoen, B. Chen, and D. Mazieres, "A Low-bandwidth Network File System," ACM Symposium on Operating Systems Principles (SOSP), Banff, Canada, PP 174-187, Oct. 2001.

[11]. P. Kulkarni, F. Douglis, J. LaVoie, J. M. Tracey, "Redundancy Elimination Within Large Collections of Files", USENIX Annual Technical Conference, Boston, MA, USA, June 2004.

[12]. Cooper, James W., Coden, Anni R., Brown, Eric W., "A Novel Method for Detecting Similar Documents", Hawaii International Conference on System Sciences,2004.

