

A Study on Graph Theory Properties of On-line Social Networks

G. H. J. Lanel*, H. S. S. P. Jayawardena**

* Department of Mathematics, University of Sri Jayewardenepura, Sri Lanka.

** ²Department of Mathematics, Texas Tech University, USA.

DOI: 10.29322/IJSRP.10.03.2020.p9929

<http://dx.doi.org/10.29322/IJSRP.10.03.2020.p9929>

Abstract- This report attempts to provide a descriptive analysis of online social media networks of Facebook, Twitter and LinkedIn based on graph theory concepts. The first part of the study focuses on identifying various activities of the above networks and constructing a suitable graph model for each social media to represent respective activities. The next part of this study addresses the evaluation of the constructed models of each social media separately in order to identify the applicability of them to analyze behavioral patterns and characteristics of the users. Final part of this study focuses on proposing a method to provide relevant information on online social networks to outsiders, which will be helpful for competent decision making without violating the user privacy.

Index Terms- Behavioral patterns, graph theory, scatter plots, social media networks.

I. INTRODUCTION

Graph theory is a mathematical concept which is based on mathematical structures made up of nodes and edges, called 'graphs', which are used to model pair-wise relations between objects. An online social network (OSN); which is used by people to build social networks or social relations with other people can be represented by a graph with nodes and edges and analyzed in order to determine many graph theory properties associated with them.

It is certain that human behavior plays a vital role in many fields such as management, marketing, trading, politics etc. Along with the advancement of information and communication technology, predicting human behavior and their requirements has been a challenging task. As OSNs become the tools of choice for connecting people, sociologists expect that their structure will increasingly mirror real life society and relationships. This has been proven mainly by comparing the power-law distribution and small-world property of real-world social network with that of OSN. Therefore, Graph theory concepts provide an effective and efficient method to analyze and evaluate OSNs which will provide a strong base of information to determine and predict human behavior in the network.

Tons of social media sites are available throughout the world with millions of active users at present. According to many rankings Facebook, Twitter and LinkedIn possess a higher position among

them. They are specifically designed for the people to interact with one another. On the other hand, they are distinct in the way their ties are established and with regard to the casualty and professionalism of them. Therefore, on-line platforms of Facebook, Twitter and LinkedIn are evaluated in this analysis.

II. MATERIALS AND METHODS

A. Construction of graph models

There are several activities inherited to on-line platforms of Facebook, Twitter and LinkedIn. The users may create new bonds with other users, they may post their photos on on-line platform so that their friends can see them, or they may send messages to their friends etc. The way in which these activities are carried out may vary with one network to another. So, determining each activity in the three sites is essential in order to model the social network of each of the social media site. Model of a social network describes a graph with several types of nodes and edges with respect to each activity on the online platform. According to each model, many sub graphs can be extracted and analyzed in order to predict behavioral patterns of the users within the network. The constructed models of social networks of Facebook, Twitter and LinkedIn are as follows.

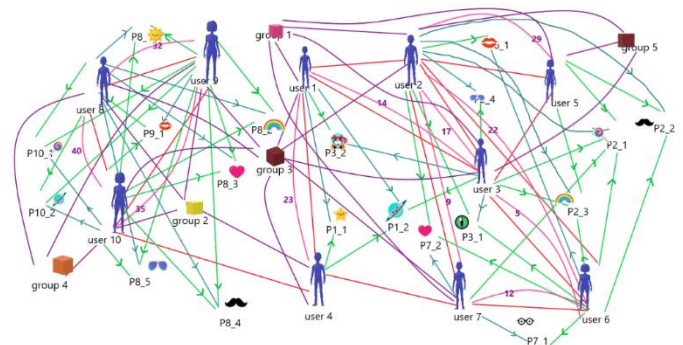


Figure 1-Facebook network model

In the Facebook network model, there are three types of nodes which are users, posts and groups. The links are friendship (red), reactions on posts (green), creation of posts (blue), messages (pink) and group membership (purple). Four sub graphs can be extracted from Facebook network model to analyze the user behavior of Facebook. They are; Friendship graph, Graph for reacting on posts, Groups graph, Messaging graph.

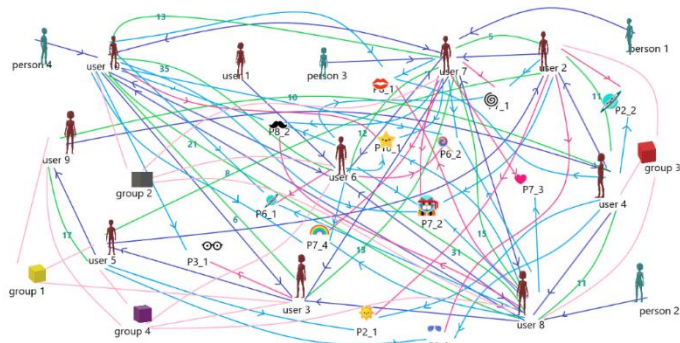


Figure 2-Twitter network model

In the Twitter network model, users, unregistered persons, posts and groups are the nodes. Links represent; following (dark blue), reactions on posts (light blue), creation of posts (dark pink), messages (green) and group membership (light pink).

This also has four subgraphs that can be extracted in which the characteristics may be different in some of the graphs from that of Facebook. The subgraphs of Twitter network model are; Followers graph, Graph for reacting on posts, Groups graph and Messaging graph.

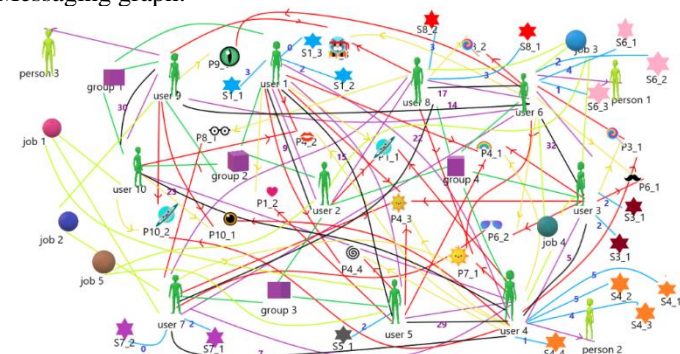


Figure 3-LinkedIn network model

In the LinkedIn network model users, unregistered persons, posts, groups, skills and jobs represent the nodes of the graph. Links and their representing colors are connections (purple), group membership (green), posting (gold), reactions on posts (red), messages (black), endorsement (blue), invitations (purple directed), recommending jobs (light green). There are six sub graphs that can be extracted from this network in order to predict meaningful results on the user behavior of LinkedIn. They are; Connections graph, Graph for reacting on posts, Groups graph, Messaging graph, Graph for endorsement of skills and Graph for recommending jobs.

B. Analysis of a sample

On-line social networks of samples of users of Facebook, Twitter and LinkedIn are generated by using the software, Gephi. The graphs are generated using Barabasi-Albert network construction method. It is a MATLAB code which produces an adjacency matrix to represent a network. Barabasi-Albert algorithm has been specifically written to generate scale-free networks. The produced adjacency matrix from this algorithm can be input to Gephi in order to develop respective graphs. Only one graph is constructed if there are common sub graphs in the three models of Facebook,

Twitter and LinkedIn. Then, graph theory properties of the constructed graphs are analyzed to determine behavioral patterns of the users within each of the social network. The generated graphs are illustrated below.

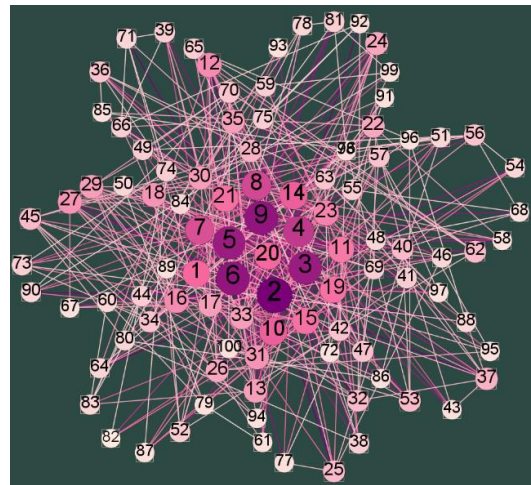


Figure 4-Fb friendship graph

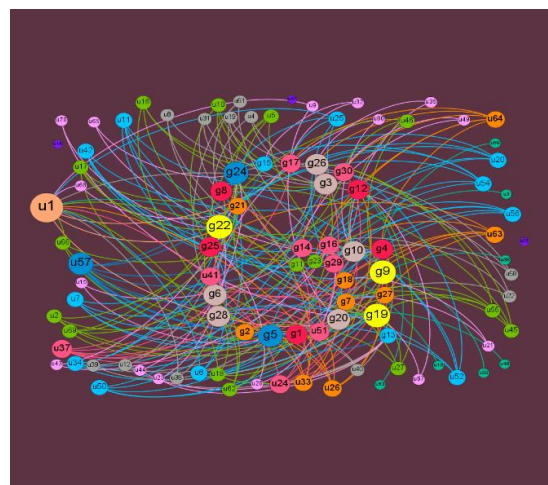


Figure 5-Groups' graph

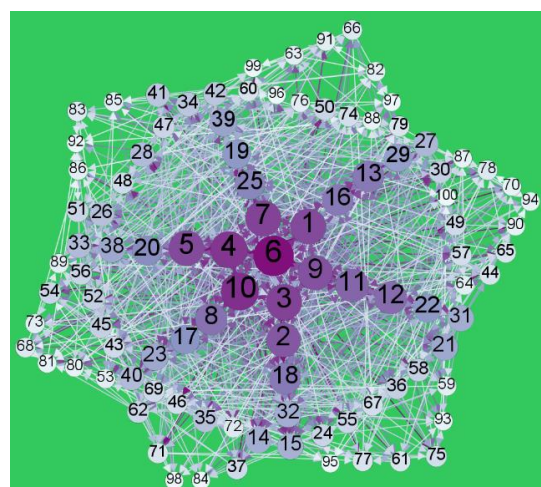


Figure 6-Twitter followers' graph

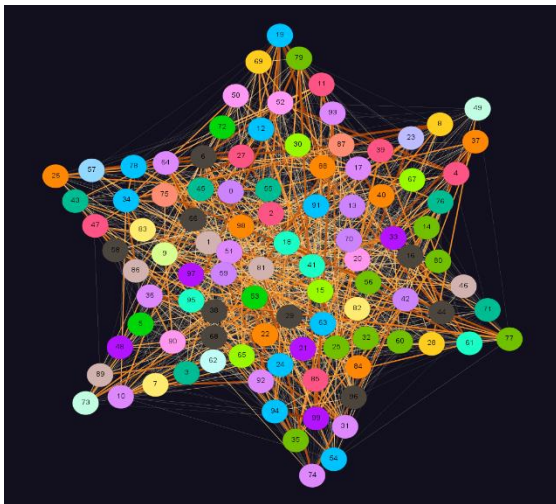


Figure 7-Messaging graph

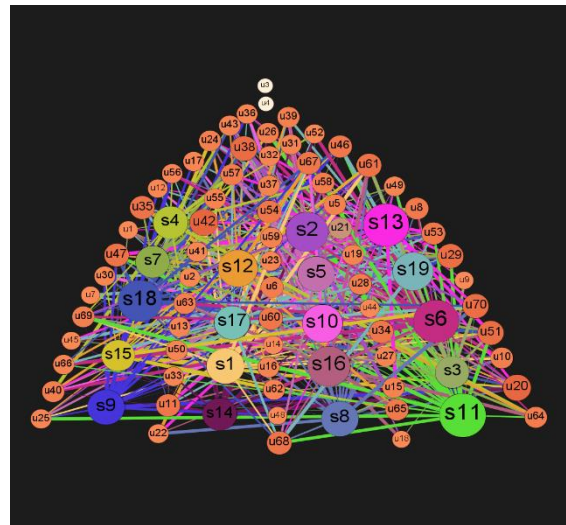


Figure 9-Graph for endorsement of skills in LinkedIn

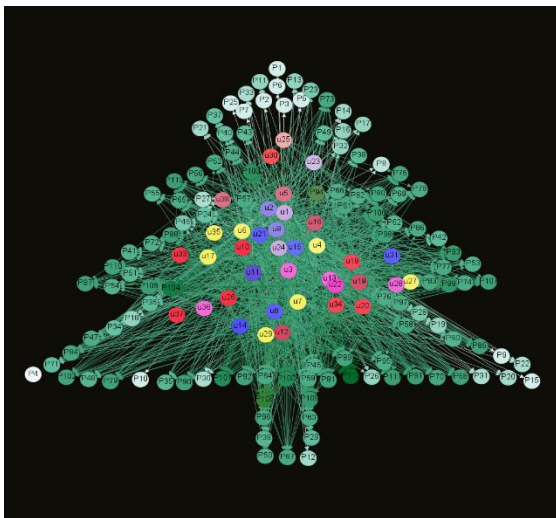


Figure 8-Graph for reacting on posts

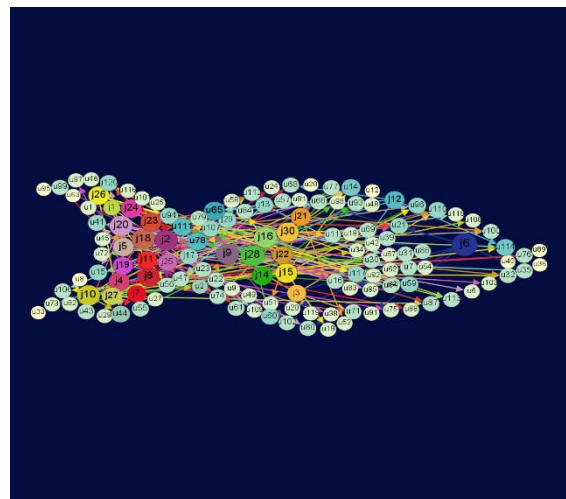


Figure 10-Graph for recommending jobs in LinkedIn

III. RESULTS AND DISCUSSION

This analysis is based on the results obtained from Gephi. Output for different graph theory properties are given as scatter plots where x-axis and y-axis represent the value of a particular property and number of users who have each value for the relevant property respectively.

A. Fb friendship graph

Graph density is 0.1. So, the general level of connectedness of this graph is much lower since it is very low than 1. Average path length is 2.257.

Figure 11 shows the power law distribution of degree of users in the network. In this network, average degree is 9.88. According to the above graph, most of the Fb users have degree 5. User 82 has the minimum degree which is 4. And user 2 has maximum degree which is 32. That is user 2 has the highest number of friends in the network thus having the greatest potential to influence other users within the network.

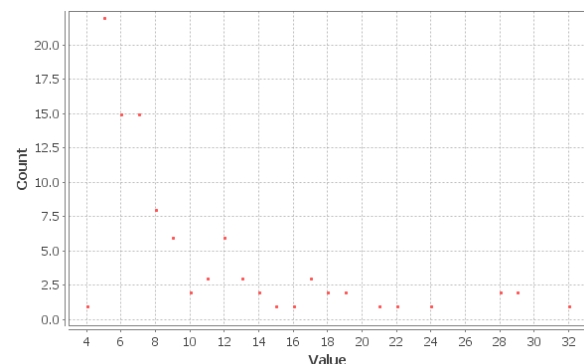


Figure 11-Degree distribution

According to Figure 12, every node has a different value for betweenness centrality since the count is 1 for each value. User 93 has the minimum betweenness centrality 3.323. And user 2 has the maximum betweenness centrality 446.906. Thus user 2 has the greatest control over the flow of information.

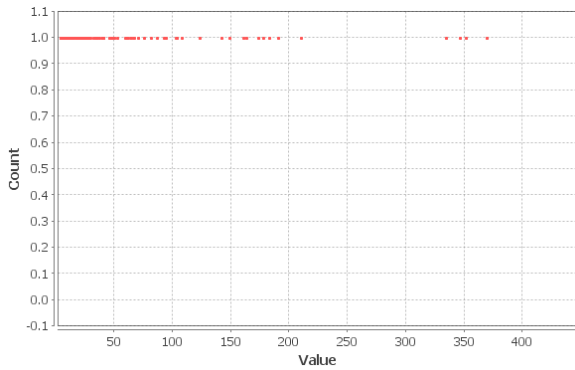


Figure 12-Betweenness centrality distribution

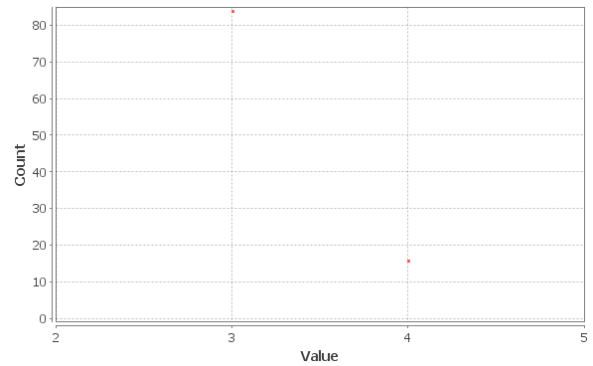


Figure 16-Eccentricity distribution

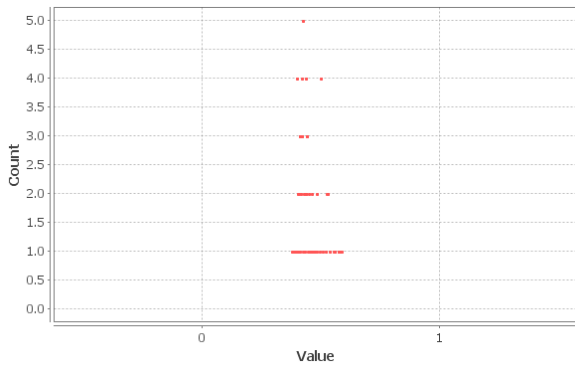


Figure 13-Closeness centrality distribution

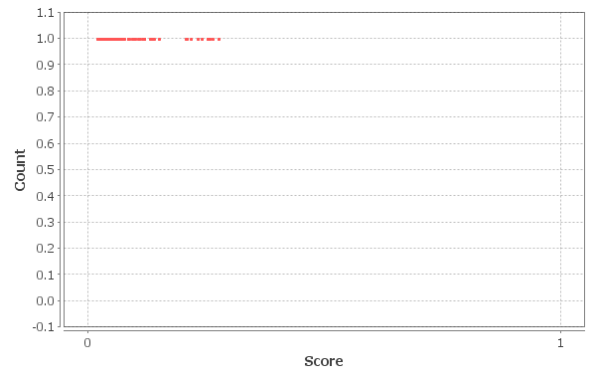


Figure 17-Authority distribution

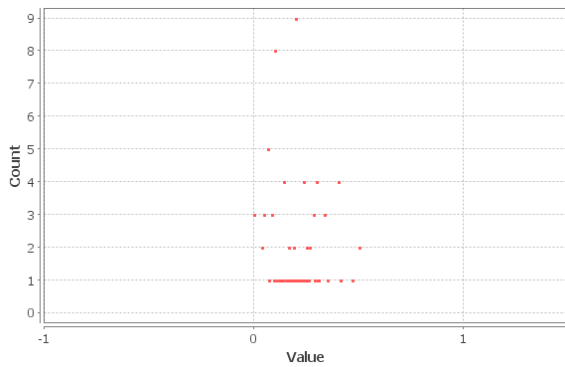


Figure 14-Clustering coefficient distribution

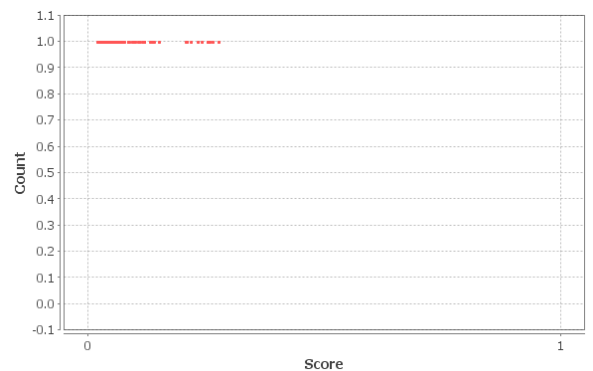


Figure 18-Hubs distribution

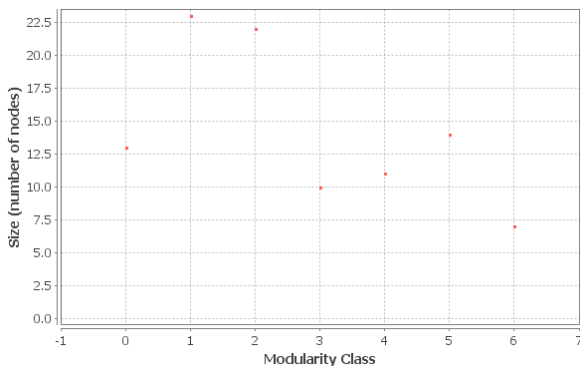


Figure 15-Size distribution

In Figure 13, five users have a closeness centrality of 0.423. Most of the users have different closeness centrality values so that they have different levels of access to information within the network. User 2 has the minimum closeness centrality of 0.586 and user 96 has the maximum closeness centrality of 0.378. Therefore, user 96 has the highest potential for independent communication.

As in Figure 14 all the friends have an eccentricity of either 3 or 4. Most of them have an eccentricity of 3. That is, the furthest user of most of the users in this network is 3 links away from them. As maximum eccentricity is 4, it is the diameter of the network and minimum eccentricity is 3, it is the radius of this network.

This friendship network has much less dense local neighborhood around each user since its average clustering coefficient is 0.196 which is a lower value as in Figure 15. There are 457 triangles in this network. Users 68, 72 and 96 have no triangles with their

friends. User 2 has the maximum number of triangles 101 with his friends. Thus, the probability that two friends of user 2 are also friends with each other, is the greatest compared to other users. This results the densest local neighborhood around user 2.

Figure 16 illustrates that different users have different authority values. That means they can drive action within the network in different levels. User 96 has the minimum authority value 0.01760381 and user 2 has the maximum authority value 0.27429184. Thus user 2 has the greatest strength to influence other users to drive action within this network.

Modularity of this network is 0.174 according to the Figure 17. Since this is a low value, users within each module of the network are less densely connected. There are 5 communities in this friendship graph. That is, the users are densely connected to each other and formed 5 communities within the network.

Since friendship network is an undirected graph, hubs distribution (Figure 18) is quite similar to authority distribution. In this network, user 96 has the minimum hub value 0.017603775 and user 2 has the maximum hub value 0.27429184. So, user 2 is pointed towards many highly authoritative users within the network. In other words, user 2 is mostly responsible for effective spreading of material on network.

LinkedIn connections graph is similar to Fb friendship graph since it is an undirected graph. It has LinkedIn users as the nodes and connections as the links. So, the analysis of LinkedIn connections graph would be very similar to the analysis of Fb friendship graph, but the results may vary with the user connections of the network.

B. Twitter followers' graph

Density of this network is 0.102. Like in Fb friendship graph, general level of connectedness of this network is much lower. Average path length is 2.373. That is one user in the network can communicate with any other user through approximately 2 steps on average.

As in Figure 19, in-degree defines a user being followed by other users in a network. Every user has the in-degree of either 9, 10, 11 or 12 according to this chart. Users 28, 8, 2 and 7 have the minimum in-degree of 9 and user 11 has the maximum in-degree of 12. So, user 11 has the maximum followers and he is the most popular within the network.

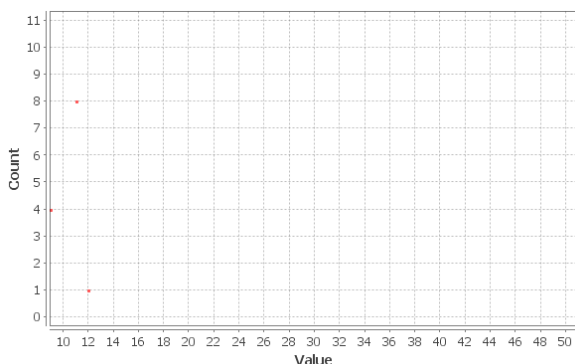


Figure 19-In-degree distribution

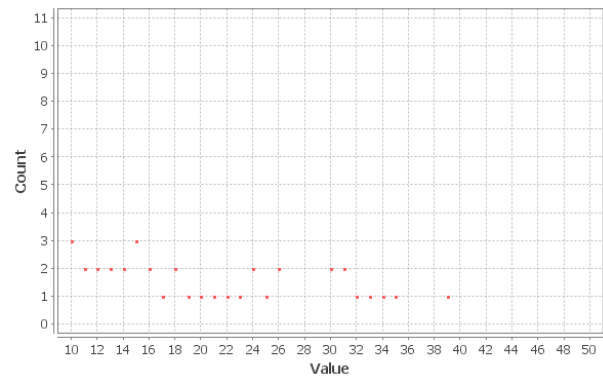


Figure 20-Out-degree distribution

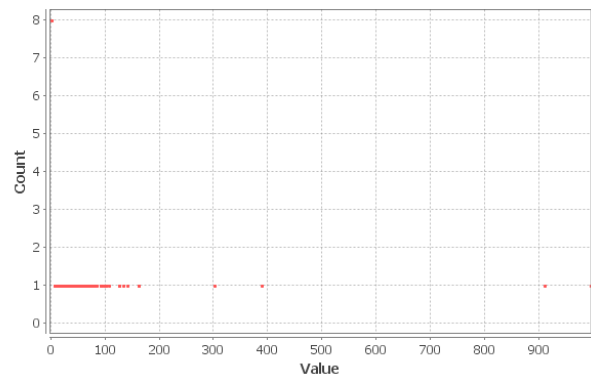


Figure 21-Betweenness centrality distribution

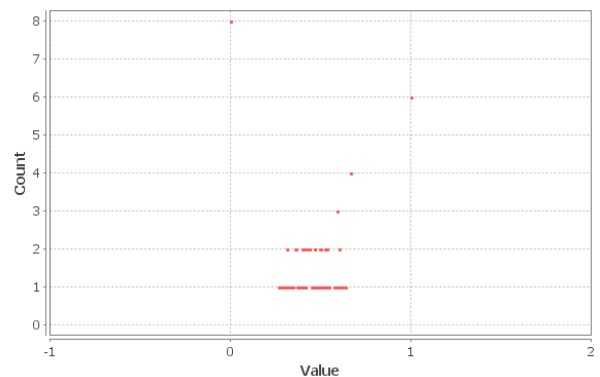


Figure 22-Closeness centrality distribution

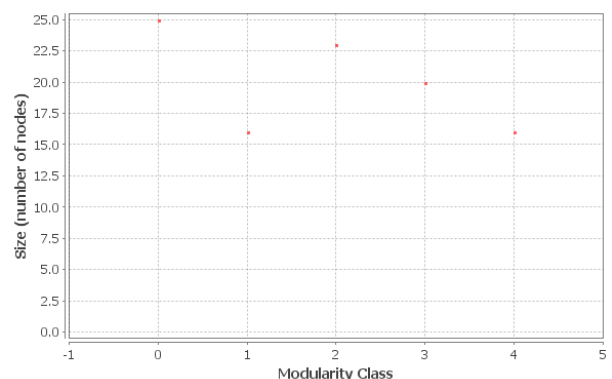


Figure 23-Size distribution

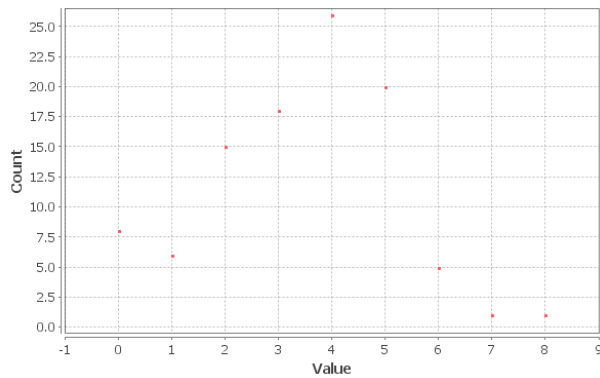


Figure 24-Eccentricity distribution

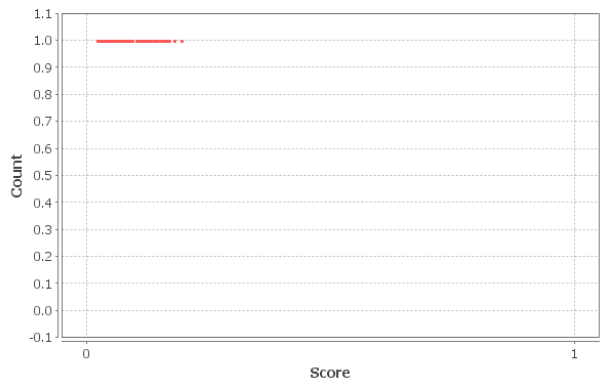


Figure 25-Authority distribution

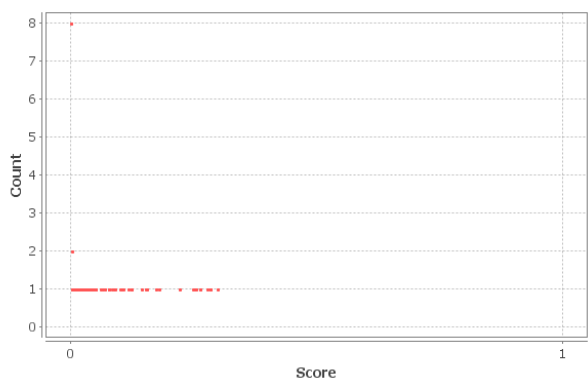


Figure 26-Hubs distribution

Out-degree (Figure 20) defines a user following other users in a network. Users 98, 96, 99, 100, 84, 92, 95 and 97 have the minimum out-degree of 0. That is, they do not like to follow others. And user 6 has the maximum out-degree of 39. So, he follows many of the other users within the network.

Figure 21 shows that users 92, 95 and 97 have the minimum betweenness centrality of 0. So, they have no control over the flow of information. And user 51 has the maximum betweenness centrality of 995.394. That is user 51 has the greatest impact over the communication throughout the network.

According to (Figure 22), some of the users in this network has no out-degree. They are users 98, 96, 99, 100, 84, 92, 95 and 97 who have the minimum closeness centrality of 0. Thus, they do not

have the potential for independent communication with other users resulting a zero-closeness centrality. And users 88, 93, 94, 91, 89 and 85 have the maximum closeness centrality of 1. So, they have the highest ability to communicate on their own throughout the network.

As average clustering coefficient of the network is 0.137, there is a less dense local neighborhood around each user within the network. User 94 has the minimum clustering coefficient of 0.036 and user 8 has the maximum clustering coefficient of 0.271. Thus user 8 has the greatest number of closed triplets among his links resulting a dense neighborhood.

Modularity (Figure 23) of this network is 0.174. Thus, users in each module are less densely connected with each other. There are 5 communities or modules in the network. That is users are grouped in 5 modules sharing their common interests.

As in Figure 24, users 98, 96, 99, 100, 84, 92, 95 and 97 have the minimum eccentricity of zero as they are not pointed towards any of the user within the network. So, network radius is zero. And user 59 has the maximum eccentricity of 8. That means the furthest away user from user 59 is 8 steps far from his position. Thus, network diameter is 8.

Different users have different levels of strength to drive others to act on particular information as they have different authority values (Figure 25). User 92 has the minimum authority of 0.021175425 and user 11 has the maximum authority of 0.19236776. Therefore, user 11 has greatest impact in driving action within the network.

Users 98, 96, 99, 100, 84, 92, 95 and 97 have the minimum hub value (Figure 26) of zero as they have zero out-degree. And user 6 has the maximum hub value of 0.2985324. So, user 6 is pointed towards many authoritative users in this network.

C. Graph for reacting on posts

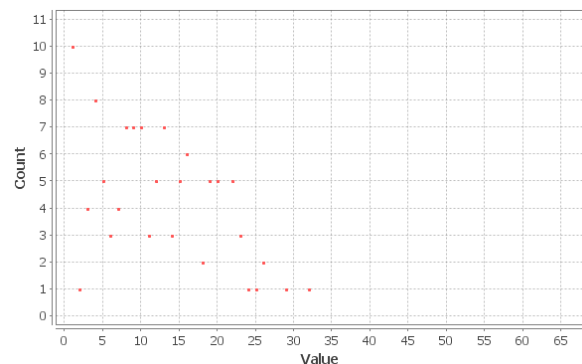


Figure 27-In-degree distribution

In-degree (Figure 27) defines the number of reactions on posts of different users. According to the results obtained by Gephi, post 4 has no reactions on it. And post 113 has the maximum reactions on it which is 32 reactions. So, post 113 can be considered as the popular post among the users of the network.

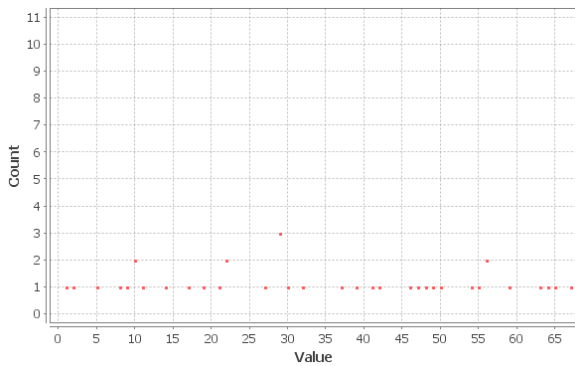


Figure 28-Out-degree distribution

As in Figure 28, out-degree is the number of reactions made by each user on posts within the network. User 38 has reacted on only 1 post which is the minimum number of reactions and user 4 has reacted on 67 posts which is the maximum number of reactions. Thus, user 4 is most active in reacting on posts of his friends.

D. Groups' graph

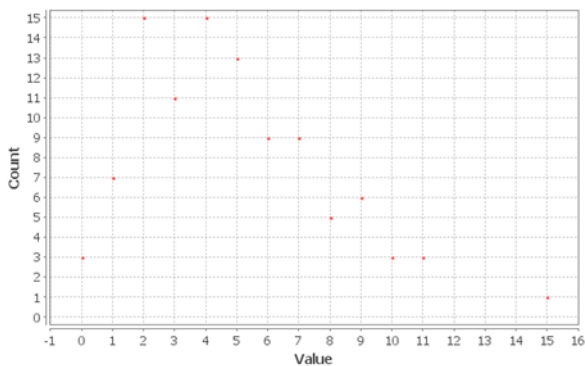


Figure 29-Degree distribution

Degree distribution defines either the number of members of a group or group memberships of a user. This chart shows that there are 15 users or posts with degree 2 or 4. That means most of the nodes in this network has a degree 2 or 4. It may be in terms of number of groups a user is attached to or group memberships of a particular group.

Users 14, 29 and 47 have no group membership of any of the groups and user 1 has the maximum number of group memberships which is 15. So, user 1 likes the most to share his common interests with other members in a group.

Group 11 has the minimum number of group members which is 4 and group 22 has the maximum number of group members which is 11. Therefore, group 22 is popular among the users within the network.

E. Messaging graph

Weighted degree distribution defines the number of messages shared by a particular user with his friends. According to this chart, weighted degree distribution lies in between 1000 and 8500. That means, the users share messages with each other within this range and different users share different number of messages as most of them have different weighted degree values. Average weighted degree is 3809.700.

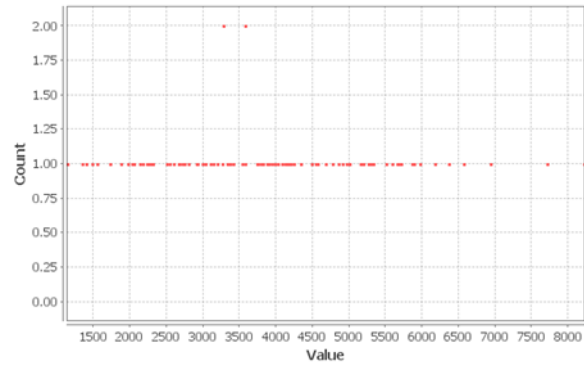


Figure 30-Weighted degree distribution

User 72 has the minimum weighted degree of 1150 and user 6 has the maximum weighted degree of 8221. So, user 6 is most active in sharing information through messages with his friends within the network.

Users 94 and 99 have shared the maximum number of messages in this network which is 985 messages. So, they communicate most frequently with each other through messages.

F. Graph for endorsement of skills in LinkedIn

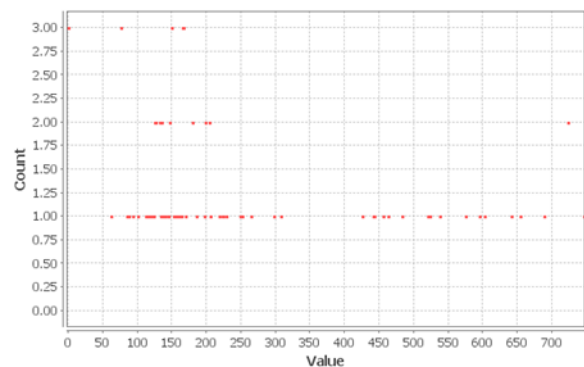


Figure 31-Weighted degree distribution

Weighted degree distribution defines the number of endorsements for skills of a user. Average weighted degree of the network is 239.244 in this network.

Users 3 and 4 have no skills endorsed. User 81 has the maximum weighted degree with endorsed skills. So, user 81 is the highly recommended person by his friends as the most skilled person in this network.

G. Graph for recommending jobs in LinkedIn

In-degree (Figure 32) is the number of jobs recommended to a particular user in this network. Users 8, 12, 28, 33, 36, 40, 63, 89, 95 have only one job recommended to them. User 94 has the maximum number of jobs which is 6, recommended to him. So, he can apply for 6 jobs if he likes.

Out-degree (Figure 33) defines the number of users a particular job has been recommended to. Job 1 is recommended to only one user. And job 6 is recommended to a maximum number of users in the network which is 11. So, the employer of job 6 may look forward towards receiving applications from 11 users.

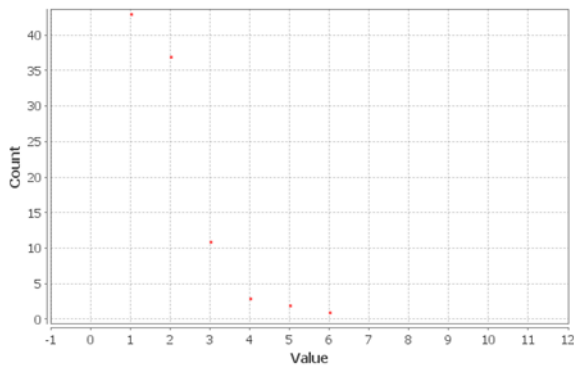


Figure 32-In-degree distribution

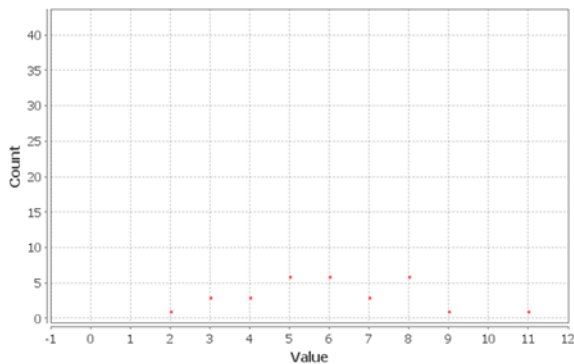


Figure 33-Out-degree distribution

IV. CONCLUSION

Various activities of Facebook, Twitter and LinkedIn can be identified and a suitable graph model for each social media to represent respective activities can be constructed. It is possible to evaluate graph theory properties of the constructed models of each social media separately in order to analyze behavioral patterns and characteristics of the users.

By evaluating graph theory properties of OSNs, valuable information on user behavior can be obtained which is beneficial for both the user himself and the company which is operating the on-line platform. The user can get his profile improved in order to increase his performance level within the network. On the other side, the operators can utilize relevant information to improve their respective on-line platforms and give away useful information to outside parties without disrupting the users' privacy.

REFERENCES

- [1] S. K. Devanarayana, G. H. J. Lanel, Develop a model to map client's people development requirements and the delivery of the service to achieve effective results, International Journal of Advanced Engineering Research and Science (IJAERS), Volume 4, Issue 3, (Mar. 2017), PP 162-165, DOI URL: <http://dx.doi.org/10.22161/ijaers.4.3.26>.
- [2] Ahn, Y, Han, S, Kwak, H, Moon, S & Jeong, H 2007, 'Analysis of topological characteristics of huge online social networking services', proceedings of the 16th international conference on World Wide Web, May 2007, pp. 835-844.
- [3] Benevenuto, F, Rodrigues, T, Cha, M and Almeida, V 2009, 'Characterizing user behavior in online social networks', proceedings of the 9th Internet Measurement Conference, November 2009, pp. 49-62.
- [4] Catanese, S, De Meo, P, Ferrara, E & Fiumara, G 2010, Analyzing the Facebook Friendship Graph, paper presented at the 1st Workshop on Mining the Future Internet, Berlin.
- [5] Chau, D, Pandit, S, Wang, S & Faloutsos, C 2007, 'Parallel crawling for online social networks', WWW'07: proceedings of the 16th international conference on World Wide Web, May 2007, pp. 1283-1284.

AUTHORS

First Author – Dr. G.H.J.Lanel, Department of Mathematics, University of Sri Jayewardenepura, Sri Lanka, ghjlanel@sjp.ac.lk.

Second Author – H.S.S.P.Jayawardena, Department of Mathematics, Texas Tech University, USA, sithma.jayawardena@ttu.edu.