2018 International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2018

Detection of Influential Nodes Using Social Networks Analysis Based On Network Metrics

Aftab Farooq
College of Electrical and Mechanical Engineering,
Department of Computer Engineering
National University of Sciences and Technology
Islamabad, Pakistan
E-mail: aftabfarooq2012@gmail.com

Muhammad Uzair
College of Electrical and Mechanical Engineering,
Department of Computer Engineering
National University of Sciences and Technology
Islamabad, Pakistan
E-mail: Muhammad.uzair15@ce.ceme.edu.pk

Gulraiz Javaid Joyia
College of Electrical and Mechanical Engineering,
Department of Computer Engineering
National University of Sciences and Technology
Islamabad, Pakistan
E-mail: ingrgulraiz@gmail.com

Usman Akram
College of Electrical and Mechanical Engineering,
Department of Computer Engineering
National University of Sciences and Technology,
Islamabad, Pakistan
E-mail: usmanakram@gmail.com

***Abstract*— Now a days, social network analysis has a great theoretical and practical significance, and this area has been explored in many fields. Social network is composed of the actors and their interaction (relationship between them). Due to increased usage of social network analysis in the practical applications, identifying the nodes from the social network has become the major problem. In any social network, the behavior and role of any individual in term of key player decides the significance and importance of social media. Visualization and the analysis of the social network includes different metrics measures i.e. Coefficient Clustering, Density, Closeness Centrality, Degree Centrality, Page Rank, Eigenvector Centrality etc. These features can be used to visualize the social networks. These factors help us in the determination of the influential nodes in any network. Detection of the influential nodes (Key Players) in any large scale social media is a complex task as thousands of new users join the network every day. This paper puts an effort to visualize, calculate the different metrics of social network and on the basis of these values determine the most influential node. The experimental results and a detailed quantitative analysis shows that this is the more efficient and effective way to detect the influential nodes in a social network.**

Keywords*— social network analysis(SNA); Data mining ; Community Detection ; online social network (OSN) ; online communities ;centrality measures ; performance ; efficiency

I. INTRODUCTION

Social media is means of communication and interaction among the people in which they create, share, trade and access the data and the thoughts thus creating groups (small nets) and networks (net of individuals and groups). A social network is a combination of the various social actors which is a representation of social network analysis. The analysis of the social. The analysis of the social network metrics which includes the different parameters that characterizes that the social media analysis is acquiring a more importance in modern era. For visualizing and analyzing the patterns of the complex social networks various different methods are used [1]. Social media analysis is a mathematical model that connects different dots to analyze nodes and their relationship [2].

Detection of the key player in an online social network is the most important area when analyzing the complex networks. We can say the key player is the one which is most influential in the social network. In recent decades the detection of the influential nodes from the social network had acquired the attention in the online social community. Network measures are the key theories in the detailed study of social networks. Centrality measure is most important for studying the organizational and team behavior [3].

The practical application of social network analysis include fraud detection, terrorist activities in covert networks, military surveillance for enemy activities. In private sector major companies are using the SNA techniques for customers support activities and analyze customer behaviors to enhance their business.

The paper consists of four sections. Section 2 explain the related work which highlights the main features which we used in social network analysis to detect the key players (influential nodes). Our proposed scheme is explained in the section 3. Section 4 presents the experimental and statistical results which are carried out to check the performance and validity of the proposed schema. We are using a case study that is followed by conclusion and discussion in the last section.

978-1-5386-1370-2/18/$31.00 ©2018 IEEE

II. LITERATURE REVIEW

A number of researchers had worked on the social network analysis for different kind of the applications. Kimura Metal [5] presents a cascaded method for detection of most key players in online social networks. Authors suggested different algorithms and techniques that extract network parameters which play a very vital role in detecting the key players in social network. Based on their statistical analysis we can say, their proposed algorithm is highly recommended than the greedy solution of the Leicht M et al and Newman community structure algorithm.

Salvatore Catanese et al [6] describes the friendship relation between users and how to access data of Facebook users. For achieving the purpose they used the web crawlers to access data directly from the website. By using gathered data and information the author constructed a sub group graph that represents the anonymous relationship among a sub group of significant users. An adhoc privacy complaint crawlers is studied to extract data from Facebook. Rejection sampling methods and Breadth-first-search (BFS) are used to minimize the biasness and for visualizing the structural characteristics of different samples that consist of huge no. of key players. Authors developed a visualizing tool for analysis of qualitative and quantitative characteristics of social media. To achieve the results they improved the efficiency of existing online social media analysis (OSN) techniques and adopted existing techniques methods and algorithms.

Pasquale De Moeetal [7] describes that how to analyze the behaviors of new users to predict whether the two nodes could be considered as a similar one. Author proposed a framework where the estimation can be performed to verify and check the similarities of two users that is based on visualizing the different activities that includes social events in which users are fully involved and on the basis of information of social relation i.e. common groups of users and friends.

Leidys del Carmen measured the clustering and association rules with famous CRISP-DM method to analyze the behaviors of the customers of the fashion industry in Instagram social network, which provided the industry with the handful of important information regarding their products and their trends and likeness among their customers. [21].

Numerous related research discussed the presence of influential nodes in a particular social networks (e.g. Facebook, Twitter, Micro blog). Cha et al [8] extracted Twitter data and did analysis to find the influence of Twitter users by comparing the network metrics values of Clustering Coefficient, Degree, Eigenvector Centrality and retweets. Xiong et al [9] introduced the model called "User Community Influencer Model" to detect the influential strength and identify the most influential nodes in a community.

Rossi, Vazirgiannis and Malliaros [10] proposed a framework that visualize the complex social networks and detect influential nodes in a network. Authors proposed a famous technique i.e. K-truss composition method that helps in visualizing and analyzing the social networks and detection of Influential nodes.

III. PROPOSED SCHEME

A. Dataset Description:

A covert network (illegal network) can also be visualized and analyzed as a social network, where each node represent the relationship of the communication. Hence terrorist network can be visualized as a social network. Noordin's top terrorist is one of the most famous covert network that was drawn from a 2006 publication of international Crisis Group "Terrorist attacks in Indonesia Noordin's Crisis Network" [11][12]. This network dataset is publically available. The data set is shown in the group of undirected relationship means if A has a relation with B then definitely B should have the relation with A and holds the complete information regarding the communication between nodes. This network consists of 79 nodes.

NodeXL tool of Microsoft was used to import our dataset and to visualize this covert network. Fig. 1, shows network generated by NodeXL tool for Noordin's top network. In this network various nodes represent the people while different edges represent the communication between the people.

This tool was used for extraction of features from the redefined or final dataset. Following Parameters are extracted by us during our implementation phase: Degree, Geodesic Distance (GD), Eigenvector Centrality (EC), Closeness Centrality (CC), Betweenness Centrality (BC), Clustering Coefficient (CC), Degree Centrality (DC), Graph Density (GD) and Page Rank.

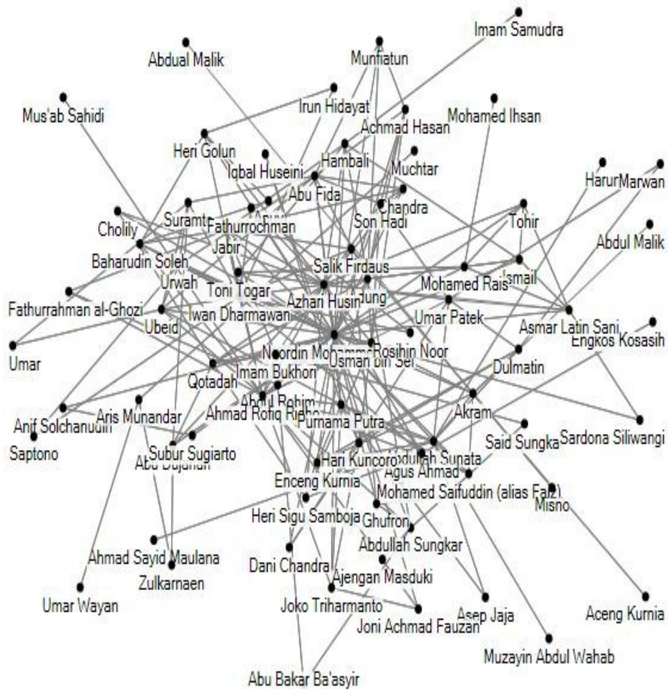

Fig. 1. The Noordin Muhammad Network.

1) Degree Centrality (DC)

The DC (Degree Centrality) is an index of exposure which represents how well connected the individual node is to network by using the number of direct contacts of a node [13]. The formula to calculate the degree centrality is given as

$$\boldsymbol{Dv} = aiv \, {}_{i=1}^{n} \tag{1}$$

Where, σ v is the Degree Centrality (DC) of v node and calculated with the help of adjacency matrix aiv.

2) Centrality

Centrality is helpful to measure the characteristics of individuals in a large network. Centrality also measure the extent of interaction & communication between individuals in a social network. The centrality is higher when there is more connections between members of a network. Individual nodes having high centrality in a network contributes more to the evolution of community in online social network [14].

a) Closeness Centrality (CC)

The Complex network graph the closeness or normalized closeness centrality (NCC) of a particular node is the length which is considered as average length of the shortest path between the two nodes of whole network. The Closeness Centrality (CC) is defined as reciprocal of farness [15] [16] that is:

$$Cx = \frac{1}{{}_{y}d(y,x)} \tag{2}$$

Where *d(y,x)* is measured distance between the x and y. Closeness Centrality (CC) can also be regarded as a measurement of how long it takes to spread any information from a particular node to all other nodes in a network.

b) Betweenness Centrality (BC)

Betweenness centrality quantifying that how long a user/node play role of bridge between two nodes in a large community network considering the shortest path. According to Linton Freeman [17] Betweenness Centrality (BC) is a quantifiable measures that fully control the node on its behavioral interaction between two other participating nodes in a social network. We can represent Betweenness Centrality as:

$$\mathrm{C_B}v = \frac{\sigma_{\mathrm{st}}(\mathrm{v})}{\sigma_{\mathrm{st}}} \tag{3}$$

$$s \neq v \neq t \in V$$

Where σ_{st} represents the total number of shortest paths from a node s to a node t and $\sigma_{st}(v)$ is the number of paths that passes through v.

Now let us consider the network in Fig. 2. Node A is high in Betweenness Centrality (BC) and node B is high in Degree Centrality

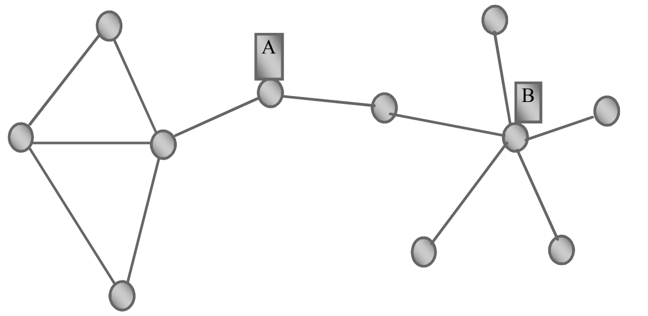

Fig. 2. Comparison between Degree Centrality and Betweenness Centrality measures.

c) Eigenvector Centrality (EC)

Measuring influence of a respective node in a social network refers as an Eigenvector centrality. Eigenvector centrality assigns a relative score to all participating Influential nodes in a complex network that is based on the concept of connection to high scoring participating nodes whose contribution is more to the score of the node in question than equality. Examples includes Katz centrality and Google's Page Rank are the variants of the eigenvector centrality.

Vertex *v* whose Eigenvector Centrality is given as [18] [19]:

$$ECv = v_x = \frac{1}{\lambda_{\max}(A)} \; {}_{j=1}^{n} a_{jx} v_j \tag{4}$$

Whereas $v = (v1, 2, \quad . \quad . \quad . \quad ,)$ refers an eigenvector for the maximum eigenvalue $\lambda_{\max}(A)$ of the adjacent matrix*A*.

d) Clustering Coefficient

Clustering Coefficient measures the likeliness of the nodes which are associates among each other. The Node who has high clustering coefficient indicates greater 'cliquishness'. In a graphical theory, a clustering coefficient measures the degree to which all participating nodes in a graph cluster together. We can calculate the average local clustering coefficients of all vertices n [20].

$$C = \frac{1}{n} \; {}_{i=1}^{n} C_i \tag{5}$$

IV. RESULTS AND DISCUSSION

The great experimental and quantitative analysis of our proposed schema is performed in this section. The performance of proposed scheme is measureing Degree, Eigenvector Centrality (EC), Closeness Centrality (CC), Betweenness Centrality (BC), Clustering Coefficient (CC), and Page Rank.

Fig. 2, Shows the metrics calculation of Noordin's network using NodeXL.

NodeXLGraph1

	Vertex	Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient
3	Abdual Malik	1	0.000	0.004	0.002	0.329	0.000
4	Akram	9	232.429	0.006	0.017	1.897	0.222
5	Abdul Rohim	4	38.236	0.006	0.012	0.803	0.333
6	Abu Bakar Ba'asyir	2	0.867	0.004	0.003	0.494	0.000
7	Ahmad Rofiq Ridho	19	284.321	0.007	0.042	3.114	0.211
8	Fathurrahman al-Ghozi	3	0.700	0.005	0.008	0.599	0.667
9	Noordin Mohammed Top	41	1496.550	0.009	0.076	6.485	0.113
10	Abdullah Sunata	12	275.067	0.006	0.022	2.374	0.258
11	Ahmad Sayid Maulana	1	0.000	0.004	0.002	0.318	0.000
12	Asep Jaja	2	0.000	0.004	0.004	0.463	1.000
13	Dulmatin	6	21.888	0.005	0.008	1.149	0.333
14	Enceng Kurnia	6	19.179	0.006	0.018	1.078	0.667
15	Harun	1	0.000	0.004	0.002	0.318	0.000
16	Mohamed Saifuddin (alias Faiz)	5	3.838	0.006	0.017	0.915	0.800
17	Muzayin Abdul Wahab	1	0.000	0.004	0.002	0.318	0.000
18	Purnama Putra	11	129.549	0.006	0.024	1.868	0.327
19	Umar Patek	8	86.952	0.006	0.017	1.389	0.214
20	Abdullah Sungkar	7	148.289	0.006	0.015	1.424	0.143
21	Adung	10	32.638	0.006	0.024	1.574	0.400

Edges / Vertices / Groups / Group Vertices / Overall Metrics

Fig. 3. Metrics calculation of Noordin Muhammad Network

Fig. 3, Shows that Noordin Mohammad Top (Row no 9) is the Key Player (Influential Node) in whole network and feature values of its metrics is high as compared to other nodes. Degree of this node is 41, Betweenness Centrality is 1496.550, Closeness Centrality is 0.009, Eigenvector Centrality is 0.076, Page Rank is 6.485 and Clustering Coefficient is 0.113.

Below Table 1, shows the metrics calculation of complete network. The NodeXL tool of the Microsoft was used to analyze and record the metric parameters of "Terrorist attacks in Indonesia Noordin's Crises Networks"[11] [12]. The table depicts that which network metrics were involved in the network and what were their measurement values after network analysis was performed.

TABLE I. METRICS CALCULATION OF COMPETE NETWORK

Sr. No.	Network Metric	Value
1	Vertices	75
2	Edges	397
3	Graph Density (GD)	0.072793
4	Maximum Geodesic Distance (MGD)	6
5	Average Geodesic Distance (AGD)	2.679111
6	Connected Components (CC)	1
7	Minimum Degree (MD)	1
8	Maximum Degree (MD)	41
9	Average Degree (AD)	5.387
10	Median Degree (MD)	4.000
11	Minimum Betweenness Centrality (MBC)	0.000
12	Maximum Betweenness Centrality (MBC)	1496.550
13	Average Betweenness Centrality (ABC)	63.467
14	Median Betweenness Centrality (MBC)	3.838
15	Minimum Closeness Centrality (MCC)	0.003
16	Maximum Closeness Centrality (MCC)	0.009
17	Average Closeness Centrality (ACC)	0.005
18	Median Closeness Centrality (MCC)	0.006
19	Minimum Eigenvector Centrality (MEC)	0.000
20	Maximum Eigenvector Centrality (MEC)	0.076
21	Average Eigenvector Centrality (AEC)	0.013
22	Median Eigenvector Centrality (MEC)	0.013
23	Minimum PageRank (MP)	0.289
24	Maximum PageRank (MP)	6.485
25	Average PageRank (AP)	1.000
26	Median PageRank (MP)	0.826
27	Minimum Clustering Coefficient (MCC)	0.000
28	Maximum Clustering Coefficient (MCC)	1.000
29	Average Clustering Coefficient (ACC)	0.462
30	Median Clustering Coefficient (MCC)	0.400

Below Fig. 4 depicts the graphical information of complete network metrics. To explain the network metrics this Fig. 3 is divided in six different sub figures, which are; degree, Closeness Centrality (CC), Betweenness Centrality (BC), Eigenvector Centrality (EC), Clustering Coefficient (CC) and Page Rank. Each one is labeled and has different parameters on x and y axis.

All the graphs (Sub Figures) were produced using the Microsoft NodeXL tool. The graphs shows the Network Metrics of "Terrorist attacks in Indonesia Noordin's Crises Networks"[11] [12].

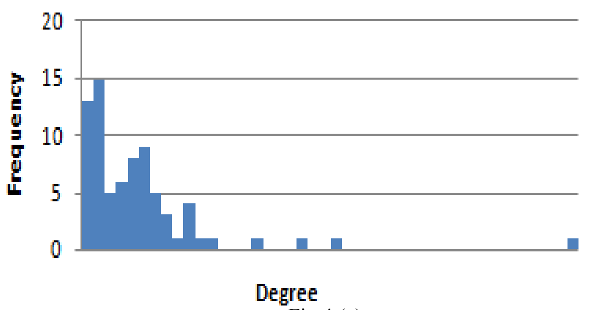

Fig 4 (a)

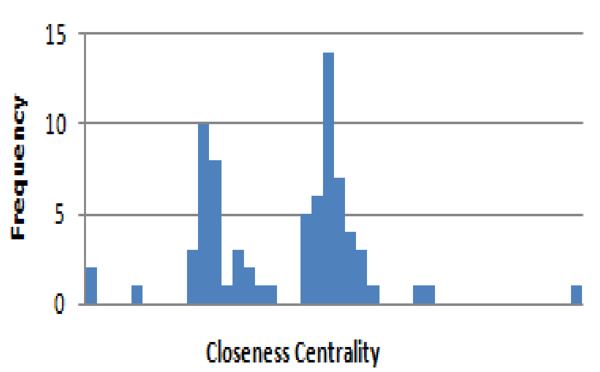

Fig 4 (b)

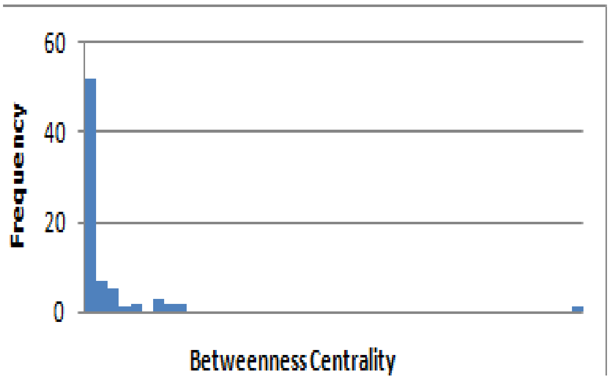

Fig 4 (c)

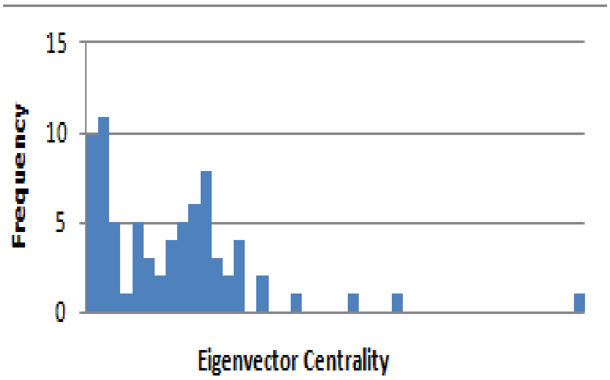

Fig 4 (d)

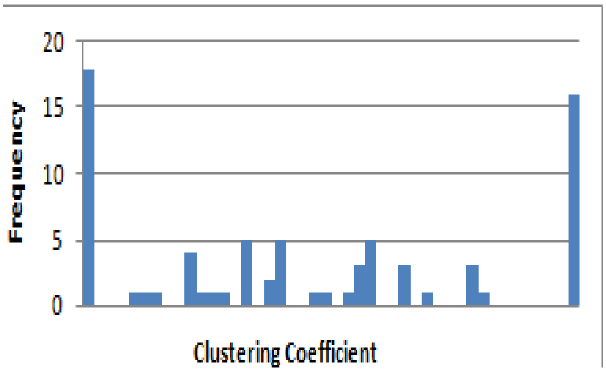

Fig 4 (e)

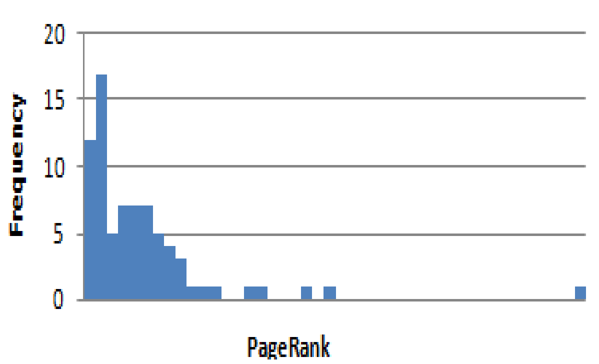

Fig 4 (f)

V. CONCLUSION

Every day more and more number of people are joining the social networks so detection of influential nodes in such a network is not an easy task. Thus our paper proposed a schema to identify the most influential nodes that are based on network measures that includes Degree, Betweenness Centrality, Closeness Centrality, Eigenvector Centrality, Page Rank and Clustering Coefficient. The proposed method has been tested using a test case. The experimental results and a detailed quantitative analysis show that this proposed schema is more efficient way for identifying influential nodes in a large social network.

REFERENCES

[1] Matei, S. A.: ‘Analyzing Social Media Networks with NodeXL: Insights from a Connected World by Derek Hansen, Ben Shneiderman, and Marc A. Smith’, Int. J. Hum. Comput. Interaction 27(4), 405–408 (2011)
[2] Krebs, Connecting the Dots—Tracking Two Identified Terrorists,2002.
[3] Tang, L., Liu, H.: Graph mining applications to social network analysis. In C. Aggarwal and H. Wang, editors, Managing and Mining Graph Data, chapter 16, pages 487-513. Springer(2010)
[4] Michael; Neumann, Gustaf (2015-07-01). "Multilayered analysis of co-development of business information systems" (PDF). Journal of Internet Services and Applications. 6 (1). doi:10.1186/s13174-015-0030-8.
[5] Kimura, M., Yamakawa, K. ; Saito, K. ; Motoda, H.," CommunityAnalysis of Influential Nodes for Infonnation Diffusion on a SocialNetwork" IEEE International Joint Conference on Neural Networks, pp.1358 - 1363, 2008.
[6] Salvatore Catanese et al., ―Extraction and Analysis of FacebookFriendship Relations,‖ Springer, 291-324, 2012.
[7] Pasquale De Meo et al., ―Finding Similar Users in Facebook,‖Information Science Reference, 304-323, 2012.
[8] LR CHA M Y, HADDADI H, BENEVENUTO F, et al. Measuring userinfluence in twitter: The million follower fallacy. Proceedings of the4th International AAAI Conference on Weblogs and Social Media,2010: 10-17.
[9] LR Xiong Z L, Jiang W J, Wang G J. Evaluating user community influence in online social networks. Proceedings of 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 2012: 640-64
[10] Maria-Evgenia G. Rossi, Fragkiskos D. Malliaros, MichallsVazirginnis. Spread it Good, Spread it Fast: Identification of InfluentialNodes in Social Networks. Proceedings of the 24th InternationalConference on World Wide Web, pages 101-102.
[11] N. Roberts and S. F. Everton, Roberts and Everton TerroristData: Noordin Top Terrorist Network (Subset), Machinereadabledata file.
[12] International Crisis Group (2006). Terrorism in Indonesia: Noordin’s Networks (No. Asia Report #114). Brussels, Belgium: International Crisis Group.
[13] Laat, Maarten de; Lally, Vic; Lipponen, Lasse; Simons, Robert-Jan (2007-

03-08). "Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis". International Journal of Computer-Supported Collaborative Learning. 2 (1): 87–103. doi:10.1007/s11412- 007-9006-4

[14] Alex Bavelas. Communication patterns in task-oriented groups. J. Acoust. Soc. Am, 22(6):725–730, 1950.

[15] indexof a graph". Psychometrika. 31: 581–603. doi:10.1007/bf02289527.

[16] Dunne, C., Shneiderman, B.: Improving graph drawing readability by incorporating readability metrics: A software tool for network analysts , University of Maryland, Human-Computer Interaction Lab Tech Report HCIL-2009-13 (2009)

[17] P. Bonacich and P. Lloyd, ―Eigenvector-like measures of centralityfor[1] asymmetric relations,‖ Social Networks, vol. 23, no. 3, pp.191–201, 2001.

[18] P. Bonacich and P. Lloyd, ―Eigenvector-like measures of centralityfor asymmetric relations,‖ Social Networks, vol. 23, no. 3, pp.191–201, 2001.

[19] Jungeun Kim and Jae-Gil Lee. 2015. Community Detection in Multi-Layer Graphs: A Survey. SIGMOD Rec. 44, 3 (December 2015), 37-48. DOI=http://dx.doi.org/10.1145/2854006.2854013.

[20] Greenhow, C. and Robelia, B.: Old Communication, New Literacies: Social Network Sites as Social Learning Resources. Journal of Computer-Mediated Communication, 14:130–1161.doi: 10.1111/j.1083-6101.2009.01484.x (2009)

[21] Leidys del Carmen, Analysis of the Behavior of Customers in the Social Networks Using Data Mining Techniques, Doi: 10.1109/ASONAM.2016.7752301