2017 3rd IEEE International Conference on Computer and Communications

# Finding Important Nodes Based on Community Structure and Degree of Neighbor Nodes to Disseminate Information in Complex Networks

Muluneh Mekonnen Tulu
State Key Laboratory of Integrated Services Networks
Xidian University
Xi'an, P.R. China
e-mail: mulumoke@yahoo.com

Ronghui Hou*
State Key Laboratory of Integrated Services Networks
Xidian University
Xi'an, P.R. China
e-mail: rhhou@xidian.edu.cn

Talha Younas
State Key Laboratory of Integrated Services Networks
Xidian University
Xi'an, P.R. China
e-mail: talha.younas@yahoo.com

***Abstract*—Applying effective methods to identify important nodes in a large and complex network is highly invaluable. Recently, in a large and complex network, finding a powerful node in the community to spread information throughout the network is the concern of many researchers. In this paper, to identify powerful nodes in a large and complex network, community-based important node (CBIN) method, which reflects on the community structure of the network and degree of neighbor nodes is proposed as a metrics. CBIN considers a random walk from a node to each community and degree of neighbor nodes. Then, CBIN describes how many influential friends the node has and how the node is essential to connect two or more than two communities of the network. The performance of CBIN is evaluated by Susceptible-Infected (SI) model. Simulation results on a real network show that the proposed method performs better than the existing methods to spread information in the network.**

***Keywords-community structure; degree of neighbor nodes; important node; complex network; community-based important node method; susceptible-infected model***

## I. INTRODUCTION

Networks can be represented in the real world in many systems. Human friendship, the physical proximity of animals, organizational structures, interconnectivity of infrastructures, Web hyperlinks, the similarity of data points are relationship represented in the network by links. The four types of networks generated from real-world data are social, biological, information, and technological networks [1]. A social network is a network connecting people to each other. It does not only belong to "online social networks" like Facebook, LinkedIn, and Twitter. Also, communication between people by telephone calls as well as collaboration, co-authorship, and co-appearance belong to social networks whereas the Protein-protein interactions, the network of blood vessels and metabolic pathways refer to biological network. A network of entities containing World Wide Web, word co-occurrence, and a network of citations represent information network while networks of airline routes, railways, roads, the electric power grid and the internet refer to technological network.

From the existence of network science to its recent vivid improvements [2], finding the powerful nodes to spread information in the complex network is a crucial issue for researchers [3].

Understanding and analyzing the network topology is playing a vital role to identify the important nodes which help to spread information quickly in the network [4], [5]. The significance of a node can have different meanings depending on its application. Betweenness centrality, degree centrality, and closeness centrality are the three common measures of node centrality formalized by Freeman [6]. Betweenness centrality counts the shortest path through the node. The number of neighbors or edges the node has in a network is simply expressed as a degree of a node, and it measures the importance of a node in the network. Closeness centrality is a node, which can disseminate information to others very effectively. It has shortest paths to all other nodes [6]. However, betweenness centrality has a shortage of information to describe the structure-property of important nodes and degree centrality is referred to only the direct connection to the target nodes.

Research conducted by L.L. Jiang *et al*. [7] shows that on the mutual behavior of social networks top-degree ranking nodes have positive effects. Still, it plays a less important role than the location of the node in the network. Considering this fact, Qi Zhang *et al.* [8] proposed a local structure (LE) to measure the importance of a node based on the degree of neighbor nodes. However, this method may fail to consider the structure of communities in the network and to rank the influential nodes within each community (or cluster) properly. To fill this gap, Zhiying Zhao *et al*. [9] proposed Community-based Centrality (CbC), which is used to identify the most information spreader nodes based on the network community structure. It is true that powerful node is a node that maximizes the spread of information in the network. However, CbC was not considered the impact of degree of neighbor nodes.

978-1-5090-6352-9/17/$31.00 ©2017 IEEE

Here in this paper, we proposed a method that selects important nodes in a large and complex network based on community structure and degree of neighbor nodes. Nodes selected by our proposed method connect communities to each other. Also, the presence of these nodes maximize, and the absence of it highly minimize the spread of information in the communities.

The rest part of the paper is organized as follows. A community and degree of neighbor node based important nodes selection method is presented in Section II. In Section III, the experiment results and analysis of complex network are described. Finally, the conclusion part is provided in Section IV.

## II. A Community and Degree of Neighbor Node Based Important Nodes Selection Method

Most networks naturally divided into modules or communities. Here in this part, strongly connected $n$ - node networks with weight matrix are considered [6], [10]. $A=[a_{ij}]$, i.e., $a_{ij}>0$ indicates the weight of the edge from node $i$ to $j$, when the network is binary (unweighted) it set to 1 while 0 if the edge does not exist. For all $i=1,2,...,n$, we assume $a_{ii}=0$. For a directed network the internal and external strength of node $i$ is denoted by $s_i^{in}=\sum_j a_{ji}$ and $s_i^{out}=\sum_j a_{ij}$, respectively, and the total strength is defined as $\delta_i=\sum_j (a_{ij}+a_{ji})$. For undirected network, it defined as $\delta_i=\sum_j a_{ji}=\sum_j a_{ij}$.

The internal density, external density, and degree of neighbor nodes are taken into consideration to measure the importance of the node to spread information in the network. We can assume in social networks if a person has many powerful friends in different communities, he can play a significant role in receiving and diffuse information around his circle to a large extent or more quickly than others [9]. As a result, our proposed scheme (CBIN) considers the external and internal density of the node, and powerful friends the node has in the network to calculate the impact of the node on receiving and diffusing information within and across the communities. CBIN of node $i$ is calculated via the following formula:

$$\rho_i=\sum_{h=1}^{c}\frac{d_{ih}}{N_h} \tag{1}$$

where $\rho_i$ is the total density of node $i$ in the community (or internal and external density of node $i$), $c$ is the number of communities in the network, $d_{ih}$ is the number of links between node $i$ and other nodes in community $h$, $N_h$ is the number of nodes in community $h$.

$$CBIN_i=\rho_i\times\frac{df_i}{\sum_{i=1}^{n}df_i} \tag{2}$$

where $df_i$ is the degree of node $i$'s neighbor. For example, node 6 in Fig.1 has one neighbor, but the degree of node 6 neighbor is five nodes excluding node 6, i.e., node 2, 3, 4, 5, and 7, $\sum_{i=1}^{n}df_i$ is the total number of degree of neighbors in the network.

Currently, different community detection algorithms have been proposed in [10]-[13]. Increasing modularity value is a confirmation of the good community detection partition since high modularity values resulted in the occurrence of cluster nodes with comparatively large intra-community edges. In this paper, α-partition proposed by C. Piccardi [10] is adopted because in α-partition, if a set of nominee partitions is available, setting the desired α-level allows one to choose the community with the best decomposition directly. To make it clear, we consider a network given in Fig. 1 with 16 nodes and 20 edges. Some centrality indices and CBIN of nodes in a network given in Fig. 1 is listed in Table I. The network is divided via α-partition into three communities (c = 3) by adjusting α value to 0.80, and its modularity value is 0.50625.

TABLE I. The Impact of Node Measured by Different Methods in a Network Given in Fig. 1

| Id | Degree | Betweenness | LE | CbC | CBIN |
|---|---|---|---|---|---|
| 1 | 6 | 0 | 2.5997 | 2.1875 | 0.062 |
| 2 | 2 | 0 | 1.371 | 0.75 | 0.02 |
| 3 | 2 | 0 | 1.371 | 0.75 | 0.02 |
| 4 | 2 | 0 | 1.371 | 0.75 | 0.02 |
| 5 | 2 | 0 | 1.371 | 0.75 | 0.02 |
| 6 | 1 | 0 | 0.5917 | 0.375 | 0.0083 |
| 7 | 3 | 9 | 1.9183 | 1 | 0.068 |
| 8 | 6 | 12 | 2.4248 | 1.875 | 0.06 |
| 9 | 1 | 0 | 0.5917 | 0.3125 | 0.01 |
| 10 | 4 | 5 | 2.2244 | 1.25 | 0.08 |
| 11 | 3 | 2 | 1.9878 | 0.9375 | 0.042 |
| 12 | 1 | 0 | 0.5917 | 0.3125 | 0.01 |
| 13 | 1 | 0 | 0.5917 | 0.3125 | 0.01 |
| 14 | 2 | 0 | 1.5305 | 0.625 | 0.02 |
| 15 | 3 | 0 | 1.9591 | 0.9375 | 0.036 |
| 16 | 1 | 0 | 0.5917 | 0.3125 | 0.01 |

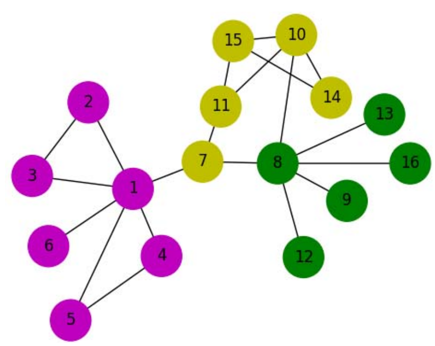



Figure 1. Community structure of a network, different colors indicate the community of the node.Nodes with the same clour belong to the same community.

TABLE II. THE MOST FOUR IMPOERTANT NODES IN A NETWORK GIVEN IN FIG. 1.

| Rank | Degree | Betweenness | LE | CbC | CBIN |
|---|---|---|---|---|---|
| 1 | 1 | 8 | 1 | 1 | 10 |
| 2 | 8 | 7 | 8 | 8 | 7 |
| 3 | 10 | 10 | 10 | 10 | 1 |
| 4 | 7 | 11 | 11 | 7 | 8 |

From Table II, node 10 and node 7 have small degree compared to node 1 and node 8. However, CBIN selects node 10 and node 7 as the top important nodes in a given network because node 10 and node 7 have strong internal density and its outer edges are connected to the node that has a high degree or has many friends in the network. Therefore, node 10 and node 7 have an excellent opportunity to disseminate information within and across the communities than other nodes. In addition, the all four important nodes which are selected by CBIN is a combination of powerful nodes selected by degree, betweenness centrality, LE, and CbC. However, the way it ranks the powerful nodes is completely different from other methods. This is the reason why our proposed method considers both community structure and the degree of neighbor nodes to indicate the impact of the node in complex network.

## III. EXPERIMENT RESULTS AND ANALYSIS

In this section, we compare the information dissemination ability of powerful nodes, which are selected by our proposed method and traditional methods over real-world data called dolphin network.

### *A. Dolphin Network*

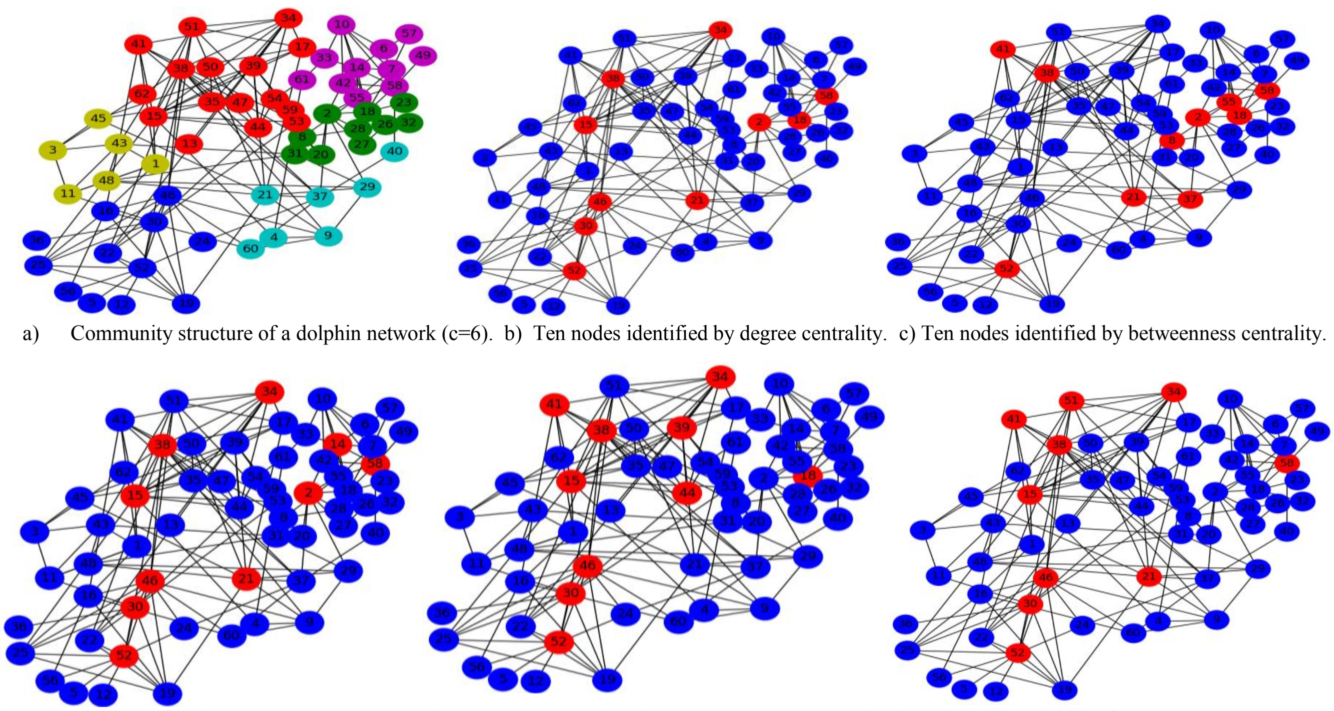
Dolphin network is an undirected social network of regular communications among 62 dolphins in a community living off Doubtful Sound, New Zealand [14].

### *B. Evaluation with Susceptible-Infected Model*

TABLE III. THE MOST TEN POWERFUL NODES IN DOLPHIN NETWORK

| Method | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Degree | 15 | 38 | 46 | 34 | 52 | 18 | 21 | 30 | 58 | 2 |
| Betweenness | 37 | 2 | 41 | 38 | 8 | 18 | 21 | 55 | 52 | 58 |
| LE | 15 | 46 | 38 | 34 | 21 | 52 | 30 | 58 | 2 | 14 |
| CbC | 15 | 34 | 38 | 46 | 52 | 30 | 39 | 44 | 41 | 18 |
| CBIN | 15 | 46 | 38 | 21 | 41 | 34 | 30 | 51 | 58 | 52 |

We applied SI model to infect the most important nodes distinguished by various methods in the dolphin network. Each infected node has one opportunity to infect other neighbor nodes in step with the probability $\beta$ (in our simulation we set the $\beta$ value to 0.08). The node is the most important as the more number of nodes are infected in the network while it used as a source of infection. Dolphin network is divided into six communities (c=6) via $\alpha$-partition community detection by adjusting $\alpha$ value to 0.50, and its modularity value is 0.5156. The top ten powerful nodes of dolphin network which are identified by degree centrality, betweenness centrality, LE, CbC, and CBIN is shown in Table III. The details visualization of top ten nodes selected by different methods is given in Fig. 2.



a) Community structure of a dolphin network (c=6). b) Ten nodes identified by degree centrality. c) Ten nodes identified by betweenness centrality.

d) Ten nodes identified by LE. e) The most ten nodes identified by CbC f) The most ten nodes identified by CBIN.

Figure 2. Applying different methods to identify important nodes in dolphin network: (a) The different colour indicates the community of the network and the nodes with the same colour belong to the same community. from (b ) to (f), nodes with blue colour are ordinary or normal nodes while nodes with red colour are powerful or imporant nodes . From (f) we can say that nodes selected by CBIN is a combination of nodes selected by other methods. Therefore, nodes selected by CBIN have a better probability to disseminate information in the network.

TABLE IV. THE ABILITY OF TOP 10 NODES SELECTED BY DIFFERENT METHODS TO INFECT NEIGHBOUR NODES AT EACH STEP (STEP 1 TO STEP 10)

| **Node ID** | **Step 1** | **Step 2** | **Step 3** | **Step 4** | **Step 5** | **Step 6** | **Step 7** | **Step 8** | **Step 9** | **Step 10** |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 0.01613 | 0.02419 | 0.03226 | 0.04113 | 0.05484 | 0.07903 | 0.09919 | 0.13306 | 0.17419 | 0.22419 |
| 14 | 0.01613 | 0.02339 | 0.03629 | 0.05242 | 0.07258 | 0.09194 | 0.11129 | 0.1371 | 0.16694 | 0.18871 |
| 15 | 0.01613 | 0.03065 | 0.05 | 0.08306 | 0.11935 | 0.15484 | 0.19274 | 0.24597 | 0.30323 | 0.35726 |
| 18 | 0.01613 | 0.02903 | 0.04193 | 0.05645 | 0.07339 | 0.09274 | 0.11613 | 0.14032 | 0.16774 | 0.19113 |
| 21 | 0.01613 | 0.02984 | 0.04113 | 0.05968 | 0.08226 | 0.11855 | 0.15323 | 0.20565 | 0.26129 | 0.325 |
| 30 | 0.01613 | 0.02903 | 0.04435 | 0.06129 | 0.0879 | 0.11694 | 0.15565 | 0.19516 | 0.22984 | 0.27984 |
| 34 | 0.01613 | 0.03226 | 0.05323 | 0.07258 | 0.10403 | 0.14113 | 0.1871 | 0.23548 | 0.28629 | 0.33306 |
| 37 | 0.01613 | 0.02419 | 0.03468 | 0.04919 | 0.07419 | 0.09919 | 0.13468 | 0.17581 | 0.23065 | 0.28629 |
| 38 | 0.01613 | 0.03387 | 0.05484 | 0.08065 | 0.11935 | 0.15323 | 0.19355 | 0.24355 | 0.28951 | 0.35 |
| 39 | 0.01613 | 0.02823 | 0.04354 | 0.06613 | 0.0871 | 0.11613 | 0.14758 | 0.1871 | 0.23952 | 0.28629 |
| 41 | 0.01613 | 0.02984 | 0.05 | 0.075 | 0.10887 | 0.15 | 0.19113 | 0.24516 | 0.29597 | 0.35161 |
| 44 | 0.01613 | 0.02662 | 0.04355 | 0.05968 | 0.08952 | 0.12258 | 0.15806 | 0.204 | 0.25403 | 0.29838 |
| 46 | 0.01613 | 0.03468 | 0.05403 | 0.07903 | 0.11139 | 0.15565 | 0.19758 | 0.23952 | 0.28548 | 0.32903 |
| **51** | **0.01613** | **0.02177** | **0.02742** | **0.04032** | **0.06048** | **0.09032** | **0.12984** | **0.17984** | **0.2355** | **0.29355** |
| 52 | 0.01613 | 0.03387 | 0.04758 | 0.06452 | 0.09194 | 0.11694 | 0.15726 | 0.20403 | 0.25081 | 0.29516 |
| 55 | 0.01613 | 0.02581 | 0.03871 | 0.05968 | 0.08387 | 0.11048 | 0.13468 | 0.16048 | 0.1871 | 0.221 |
| 58 | 0.01613 | 0.02667 | 0.0379 | 0.05645 | 0.06855 | 0.08871 | 0.10968 | 0.13306 | 0.15242 | 0.17258 |

TABLE V. THE RANK OF NODES BASED ON SPREADING INFORMATION IN DOLPHIN NETWORK.

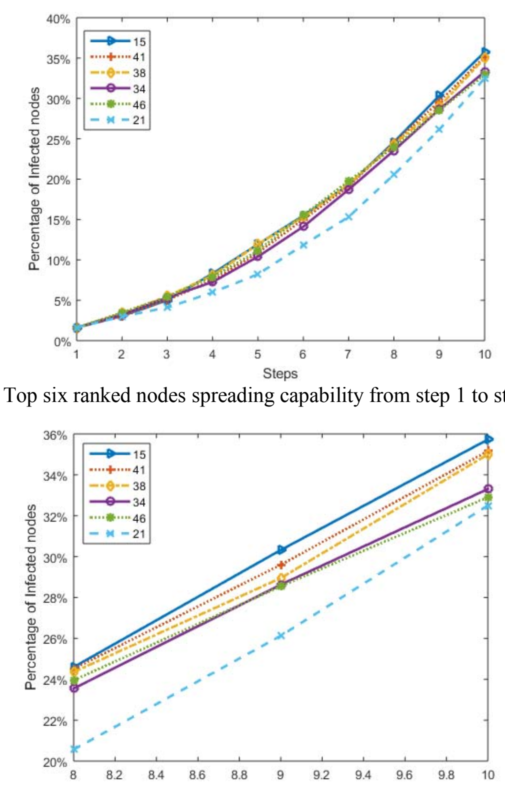
| **Rank** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Node Id** | 15 | 41 | 38 | 34 | 46 | 21 | 44 | 52 | 51 | 39 | 37 | 30 | 2 | 8 | 55 | 18 | 14 | 58 |

Table IV demonstrates the ability of individual top ten nodes selected by five different methods to spread information in the network. As we can observe, node 51 has a good impact to spread information in the dolphin network compared to some nodes selected by other methods starting from step 7. However, node 51 is identified only by CBIN, which other methods failed to recognize within the top ten nodes (Table III). We can rank the top ten nodes selected by betweenness centrality, degree centrality, LE, CbC, and CBIN based on the capability to spread information within the network from step 8 to step 10 (Table IV). The rank of the nodes is shown in Table V.

Top ten nodes to spread information in the dolphin network are the nodes rank from 1 to 10 (Table V). Out of these top-ranked nodes, betweenness centrality identifies only four nodes, degree centrality, and LE identify six nodes, CbC identifies seven nodes, and CBIN identifies eight nodes. Therefore, CBIN performs the best to select influential nodes to spread information in the network than others.

The performance of top six nodes to spread information in the network is shown in Fig. 3. We can see that out of top six nodes; betweenness centrality identifies only two of them, CbC identifies four of them, degree centrality and LE identifies five of them, and CBIN identifies all of them.

Fig. 4 shows the ability of top ten nodes that are selected by different methods to disseminate information in the network from step 40 to step 60. Since CBIN is based on the influential friends and the structure of the community, it is expected that at initial steps it may not performs to disseminate information as good as other methods.




a) Top six ranked nodes spreading capability from step 1 to step 10.

b) The details in step 8 to 10.

Figure 3. The spreading capability of top six selected nodes (from step 1 to step 10).

However, after some steps, it performs the best to spread information throughout the network quickly. As we can observe from Fig. 4, at step 50 CBIN spreads information to all nodes in the network. Therefore, it performs the best. LE and degree centrality perform well. CbC and BW perform next to degree centrality.

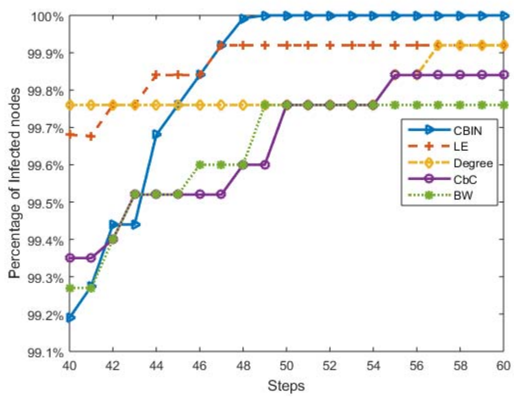



Figure 4. Information spreading capability of top ten nodes which are selected by different strategies.

## IV. CONCLUSIONS

In a large and complex network, identifying the powerful nodes to spread information throughout the network is challenging. In this paper, CBIN is proposed to select powerful nodes in complex networks. Our proposed method, rank the powerful nodes better than other methods; depend on their ability to spread information in the network. Nodes selected and ranked by CBIN is key nodes to connect communities and to receive and pass information in the network than other nodes. From the simulation results, we observed that node with high CBIN has a greater impact to spread information through the network than a node with high degree, betweenness centrality, LE, and CbC. Finally, CBIN outperforms the traditional methods to select powerful nodes in the complex network and rank them well by their capability to spread information in the network. In future, we aim to develop mobile community detection algorithm based on the pattern of mobile node, which will be cascaded to this method to identify important nodes in each community.

## ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (Grants Nos. 61571351, and 61401326), the important national science & technology specific projects 2015ZX03002006003, Natural Science Basic Research Plan in Shaanxi Province of China (Program Nos. 2016JM6028 and 2016JQ6054).

## REFERENCES

[1] B.Y. Ricardo, H. Carlos and M. Marcelo, “Query Clustering for Boosting Web Page Ranking,” in Advances in Web Intelligence. Springer, vol. 3034, pp.164–175,2004.

[2] B. Lars, B. Paolo, R. Marco, U. Johan, and V. Sebastiano, "Four Degrees of Separation," arXiv:1111.4570v3 [cs.SI], 2012.

[3] S. Aral and D. Walker, "Identifying Influential and Susceptible Members of Social Networks," Science, vol. 337, pp. 337–341, 2012

[4] M. Kas, C. KM, and C. LR, "Incremental closeness centrality for dynamically changing social networks " in Proceedings of the 2013 Ieee/Acm International Conference On Advances in Social Networks Analysis and Mining, Asonam, 2013.

[5] Z. Gao, Y. Shi, and S. Chen, "Measures of node centrality in mobile social networks," International Journal of Modern Physics C, vol. 26, p. 1550107, 2015.

[6] L. C. Freeman “Centrality in Social Networks” Social Networks, Vol.1, pp 215-239, 1979.

[7] L.L. Jiang, M. Perc, W.X. Wang, Y.C. Lai and B.H.Wang “Impact of link deletion son public cooperation in scale-free networks” EPL Vol. 93,2011.

[8] Zh. Qi, L. Meizhu, D. Yuxian and D. Yong “Local structure entropy of complex networks,” arXiv:1412.3910v1 [cs.SI], 2014.

[9] Z. Zhao, X. Wang, W. Zhang, and Z. Zhu, “A Community-Based Approach to Identifying Influential Spreaders,” Entropy, Vol.17, pp. 2228-2252, 2015.

[10] C. Piccardi, "Finding and testing network communities by Lumped Markov chains," PLoS One, vol. 6, p. e27028, 2011.

[11] S. Fortunato “Community detection in graphs” Phys. Rep. vol, 486, pp.75–174, 2010.

[12] J. Duch and A. Arenas “Community detection in complex networks using extreme optimization” Phys. Rev. E, vol.72, 027104. 2005

[13] M.E.J. Newman and M.Girvan “Finding and evaluating community structure in networks” Phys. Rev. E vol. 69, 026113, 2004.

[14] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, “the bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations,” behavioral ecology and sociobiology, vol.54, pp. 396-405, 2003.