

# A New Method of Identifying Influential Nodes in Complex Network,- Diffusion Centrality

Tingping Zhang<sup>1,2</sup>, Bin Fang<sup>1</sup>, Xinyu Liang<sup>3</sup>

1 College of Computer Science  
Chongqing University  
Chongqing, China

2 School of Information Science & Engineering  
Chongqing Jiaotong University  
Chongqing, China

3 College of Ocean & Civil Engineering  
Dalian Ocean University  
Dalian, Liaoning, China

**Abstract** — How to identify influential nodes in weighted complex network has drawn much attention in recent years. According to empirical observations, an individual in the real world usually influences its nearest neighbors and next nearest neighbors. The influence of the individual spreads from the center to the surrounds and follows a gradient descent rule. In this paper, we propose a new method to identify influential nodes in complex social networks and name it Diffusion Centrality. This measure describes the influence spread rules of the real world well. Numerical examples are provided to demonstrate the efficiency of the proposed method.

**Keywords** - Complex networks, Centrality measures, Influential nodes, Semi-local centrality

## I. INTRODUCTION

How to identify influential nodes in complex networks has attracted increasing attention in recent years [1-9]. Concretely, we need to focus on how to measure the ability of a node to spread information or disease to a large crowd of individuals in the network. Until now, lots of centrality measures have been proposed, such as three classical centrality measures – Degree centrality (DC), the Betweenness Centrality (BC) [10] and the Closeness Centrality (CC) [11], which are studied and compared with the proposed method in this paper.

First, CC considers the assumption that the information or disease between two nodes spreads on their shortest path, which is regarded as the most efficient transport manner. Therefore, the definition of CC is that the inverse of the average shortest path distances from a node to all other nodes in the network. Recently, Kitsak et al. [12] argue that the node spreading influence is determined by its location in a network. By decomposing a network with the k-shell decomposition method, they found that the most influential nodes, namely the network core, could be identified by the largest k-core values. In 2011, Chen et al. [13] proposed an effective Semi-local Centrality which can give a better result in low computational complexity than other methods, such as BC and CC.

In this paper, we propose a novel centrality measure to identify influential nodes in complex networks – Diffusion Centrality. Inspired by the influence (or power strength) spreading rules of most individuals, one not only influences

its nearest friends or family members but also members of his friends or family members. In a word, the diffusion rule of an individual is limited to two layers of the center in social networks. Therefore, according to this rule, we define a novel influence centrality measure to identify influential nodes in complex networks.

The rest of the paper is organized as follows. Section 2, we introduce the basic concepts of the three classical measures and the disease spreading model – SI model [14]. We proposed our method Diffusion Centrality in Section 3. Then, some numerical examples are used to demonstrate the validity and efficiency of our proposed model in Section 4. Section 5 concludes the paper.

## II. BASIC THEORY

### A. Three Classical Centrality Measures.

Degree centrality, closeness centrality and between-ness centrality have been well defined in Ref. [11] to identify influential nodes in a complex network.

Definition 2.1. The DC of node  $i$ , denoted as  $C_D(i)$ , is defined as

$$C_D(i) = \sum_j^N x_{ij} \quad (1)$$

where  $i$  is the focal node,  $j$  represents all other nodes,  $N$  is the total number of nodes, and  $x_{ij}$  represents the connection between node  $i$  and node  $j$ . The value of  $x_{ij}$  is defined as 1 if node  $i$  is connected to node  $j$ , and 0 otherwise.

The closeness and betweenness centrality measures rely on the identification and length of the shortest paths among nodes in the networks. Therefore, a large number of approaches have been great interest in the shortest distances among nodes in networks [2].

Definition 2.2. In a binary network, the shortest path is found by minimizing the number of intermediary nodes, and its length is defined as the minimum number of ties linking the two nodes, either directly or indirectly [15]:

$$d(i, j) = \min(x_{ih} + \dots + x_{hj}) \quad (2)$$

where  $h$  are intermediary nodes on paths between node  $i$  and  $j$ .

Definition 2.3. The CC of node  $i$ , denoted as  $C_c(i)$ , is defined as

$$C_c(i) = \left[ \sum_j^N d_{ij} \right]^{-1} \quad (3)$$

where  $d_{ij}$  denotes the distance between node  $i$  and node  $j$ .

Definition 2.4. The BC of node  $i$ , denoted as  $C_B(i)$ , is defined as

$$C_B(i) = \left( \sum_{s \neq i \neq t} \frac{g_{st}(i)}{g_{st}} \right) \frac{n(n-1)}{2} \quad (4)$$

where  $g_{st}$  is the number of binary shortest paths between node  $s$  and node  $t$ , and  $g_{st}(i)$  is the number of those paths that go through node  $i$ . This formula  $\frac{n(n-1)}{2}$  is used to normalize the betweenness centrality value, where  $n$  is the number of the nodes.

### B. SI Model.

To evaluate the performance of the proposed model, we adopt the SI model to examine the spreading influence of nodes ranked by different methods [16, 17]. In this model, there are two compartments, namely susceptible  $S(t)$  and infected  $I(t)$ .  $S(t)$  is expressed as the number of individuals susceptible to (not yet infected) the disease, and  $I(t)$  represents the number of individuals that have been infected and are able to spread the disease to susceptible individuals. At each step, each infected node spreads the disease to all its neighbors with probability  $P$ . Node  $j$  is infected by node  $i$

with probability  $P_{ij} = \left( \frac{\omega_{ij}}{\max(\omega) + 1} \right)^\alpha, \alpha > 0$ , where  $\omega_{ij}$  is the weight of edge  $E_{ij}$  and  $\alpha$  is a positive number in the SI model [14]. Since  $\frac{\omega_{ij}}{\max(\omega) + 1} < 1$ , the smaller the  $\alpha$  is, the more

quickly the infection spreads. The total number of infected nodes at time  $t$ , denoted by  $F(t)$ , can be considered as an indicator to evaluate the influence of the initially infected node at time  $t$ . Clearly,  $F(t)$  increases with  $t$ , and stops when there is no reachable susceptible node to be infected.

### III. PROPOSED METHOD

Consider an individual in real world, he or she would probably influence his or her friends directly. Then, his or

her friends would accept the influence and spread it to their neighbors. Without doubt, personal influence is presented as a form of diffusion. However, the influence accepted by the next nearest friends is smaller than the nearest friends. Therefore, the influence in social networks of the real world here takes on some form of gradient descent.

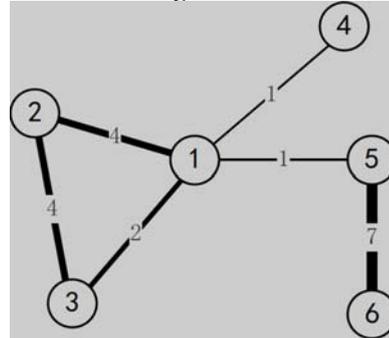


Figure 1. An example network consisted of 6 nodes and 6 edges.

In order to describe the above influence diffusion rule, we propose a new centrality measure to identify influential nodes in social networks - Diffusion Centrality. It considers both the nearest and the next nearest neighbors with descent weight factor. The Diffusion Centrality  $C_D(v)$  of node  $v$  is defined as

$$C_D(v) = (1 - \alpha) \sum_{i \in \Gamma(v)} W(i) + \alpha \sum_{j \in \Psi(v)} W(j), \quad (5)$$

where  $\Gamma(v)$  is the set of the nearest neighbors of node  $v$  and  $\Psi(v)$  is the set of the next nearest neighbors of node  $v$ .  $W(i)$  is the weight of edge  $L_{iv}$  and  $W(j)$  is the sum of edge weights of node  $j$ . As we know, the influence of an individual has the property of direction. Therefore, undirected edge actually means mutual influence between two adjacent individuals. Take Fig. 1, node 1 has five nearest neighbors and one next neighbors and thus the value of the first influence layer  $\sum_{i \in \Gamma(1)} W(i) = 8$  and the second influence

layer  $\sum_{j \in \Psi(1)} W(j) = 23$ . Set the descent weight factor  $\alpha = 0.5$ , then, we get  $C_D(1) = 15.5$ .

Diffusion Centrality measure is likely to be more effective to identify influential nodes than degree centrality measure as it coincides with the social influence spreading rule.

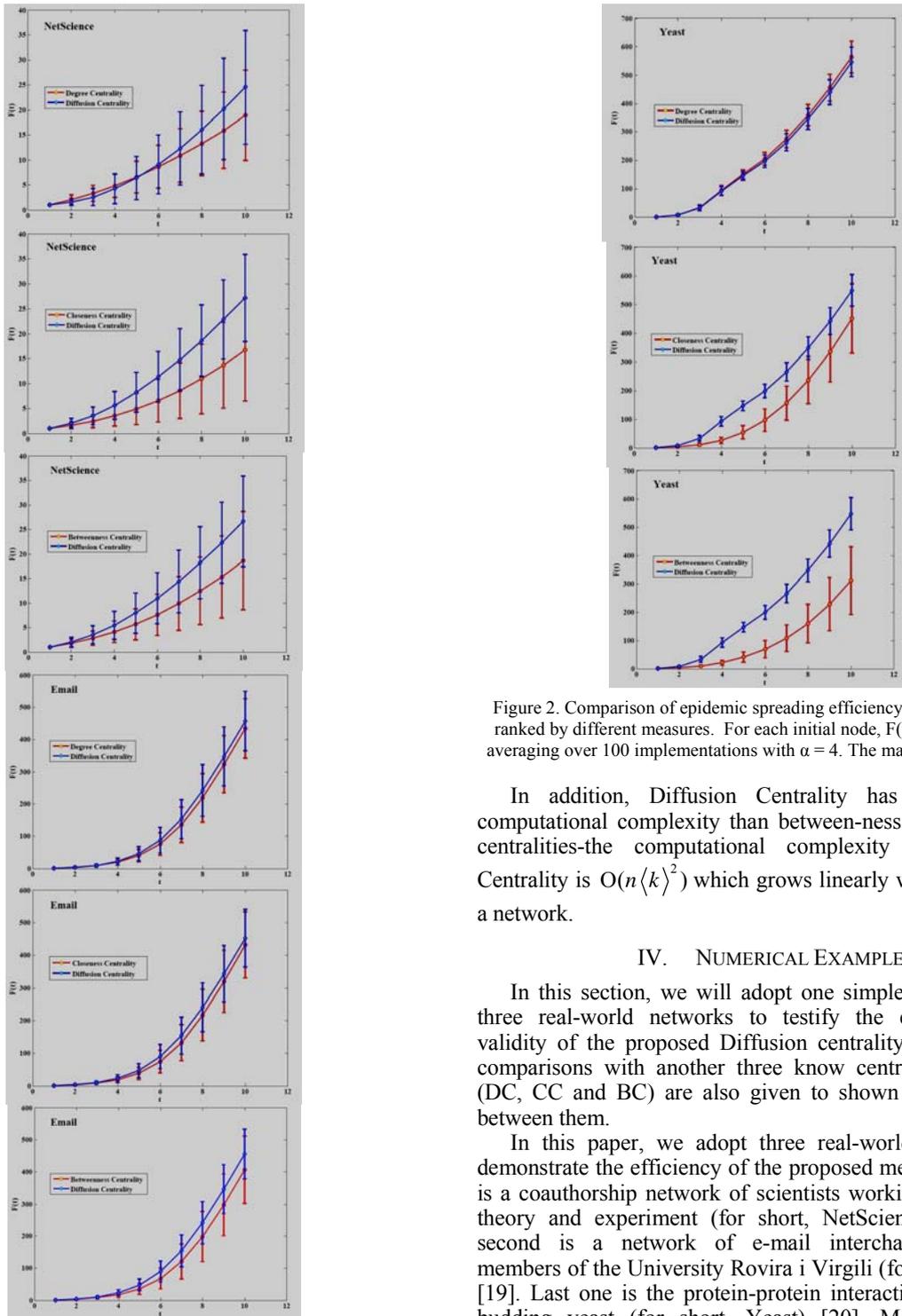


Figure 2. Comparison of epidemic spreading efficiency of top-10 nodes ranked by different measures. For each initial node,  $F(t)$  is obtained by averaging over 100 implementations with  $\alpha = 4$ . The max time step is 10.

In addition, Diffusion Centrality has much lower computational complexity than between-ness and closeness centralities-the computational complexity of Diffusion Centrality is  $O(n\langle k \rangle^2)$  which grows linearly with the size of a network.

#### IV. NUMERICAL EXAMPLE

In this section, we will adopt one simple networks and three real-world networks to testify the efficiency and validity of the proposed Diffusion centrality. Meanwhile, comparisons with another three know centrality measures (DC, CC and BC) are also given to shown the difference between them.

In this paper, we adopt three real-world networks to demonstrate the efficiency of the proposed method. The first is a coauthorship network of scientists working on network theory and experiment (for short, NetScience) [18]. The second is a network of e-mail interchanges between members of the University Rovira i Virgili (for short, Email) [19]. Last one is the protein-protein interaction network in budding yeast (for short, Yeast) [20]. Moreover, some properties of these networks are also outlined in Table 1.

As shown in Fig. 2, we compare Diffusion Centrality with other three traditional centrality measures by comparing the influence of nodes that appear in the top-10 list either by DC (or BC, CC) or the proposed method, but do not appear in both lists. The upper three panels show that Diffusion

Centrality always performs better than BC and CC in the network NetScience. Note that DC is better than Diffusion Centrality in the initial steps. Later, the infected nodes become less than Diffusion Centrality as times go on. Similarly, we can also obtain that the Diffusion Centrality

measures the influential nodes in a better way than BC and CC in the networks Email and Yeast, except DC better than Diffusion Centrality in Yeast. In sum, the proposed method Diffusion Centrality shows great performance than DC, BC and CC.

TABLE I THE BASIC TOPOLOGICAL FEATURES OF THE THREE REAL NETWORKS. N AND M ARE THE TOTAL NUMBERS OF NODES AND LINKS, RESPECTIVELY.  $\langle k \rangle$  AND  $k_{max}$  DENOTE THE AVERAGE AND THE MAXIMUM DEGREE. C AND  $\langle l \rangle$  ARE THE CLUSTERING COEFFICIENT AND AVERAGE PATH LENGTH, RESPECTIVELY.

Data	n	m	$\langle k \rangle$	$k_{mac}$	C	$\langle l \rangle$
Yeast	2375	11693	9.847	118	0.306	5.096
Netscience	379	914	4.823	34	0.741	6.42
Email	1133	5451	9.622	71	0.220	3.606

IV. CONCLUSION

In this paper, a novel centrality measure Diffusion Centrality is proposed based on social influence spreading rule to identify influential nodes in complex networks. By considering two layer influence spreading rule, the proposed method is more effective than the traditional methods - DC, BC and CC. The numerical examples show that proposed method approach can effectively identify influential nodes in complex networks. In the near future, we may try to apply Diffusion Centrality to evaluate the real influence of each person in some fields of real world, such as academic researchers, company leaders and so on.

ACKNOWLEDGMENT

The work is partially supported by National Basic Research Program of China (973 Program) (Grant No. 2013CB329100).

REFERENCES

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol. 393, pp.440-442, 1998.

[2] M. E. Newman, "Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality", *Physical Review E*, vol. 64, pp. 106-132, 2001.

[3] O. Sporns, C. J. Honey, and R. Kötter, "Identification and classification of hubs in brain networks ", *PloS one*, vol. 2, pp. 1049-1052, 2007.

[4] D. Wei, X. Deng, X. Zhang, Y. Deng, and S. Mahadevan, "Identifying influential nodes in weighted networks based on evidence theory", *Physica A: Statistical Mechanics and its Applications*, vol. 392, pp.2564-2575, 2013.

[5] C. Gao, D. Wei, Y. Hu, S. Mahadevan, and Y. Deng, " A modified evidential methodology of identifying influential nodes in weighted networks", *Physica A: Statistical Mechanics and its Applications*, vol. 392, pp.5490-5500, 2013.

[6] S. H. Strogatz, "Exploring complex networks", *Nature*, vol. 41, pp.268-276, 2001.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web", vol. 12, pp.356-459, 1999.

[8] V. Latora and M. Marchiori, "A measure of centrality based on network efficiency", *New Journal of Physics*, vol. 9, pp. 188-197, 2007.

[9] K. E. Joyce, P. J. Laurienti, J. H. Burdette, and S. Hayasaka, " A new measure of centrality for brain networks", *PLoS One*, vol. 5, pp. 122-130, 2010.

[10] L. C. Freeman, S. P. Borgatti, and D. R. White, " Centrality in valued graphs: A measure of betweenness based on network flow", *Social networks*, vol. 13, pp.141-154, 1991.

[11] L. C. Freeman, "Centrality in social networks conceptual clarification", *Social Networks*, vol. 2, No. 07, pp. 54-65, 1979.

[12] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks", *Nature Physics*, vol. 6, pp. 888-893, 2010.

[13] D. Chen, L. Lü, M. S. Shang, Y. C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks", *Physica A: Statistical Mechanics and its Applications*, vol. 4, pp. 1777-1787, 2012.

[14] Y. Wang and G. Xiao, "Epidemics spreading in interconnected complex networks", *Physics Letters A*, vol. 42, pp. 2689-2396, 2012.

[15] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths", *Social Networks*, vol. 3, pp. 245-251, 2010.

[16] Y. Gang, Z. Tao, W. Jie, F. Zhong Qian, and W. Bing Hong, "Epidemic spread in weighted scale-free networks", *Chinese Physics Letters*, vol. 2, pp. 510-513, 2005.

[17] S. Gao, j. Ma, Z. Chen, G. Wang, and C. Xing, "Ranking the spreading ability of nodes in complex networks based on local structure", *Physica A: Statistical Mechanics and its Applications*, vol. 1, pp. 130-147, 2014.

[18] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices", *Physical review E*, vol. 3, pp. 36-104, 2006.

[19] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions", *Physical review E*, vol. 68, pp. 65-103, 2003.

[20] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, et al., "Topological structure analysis of the protein-protein interaction network in budding yeast", *Nucleic acids research*, vol. 9, pp. 2443-2450, 2003.