IJICTR International Journal of Information & Communication Technology Research
ITRC

Volume 7- Number 3- Summer 2015 (43-51)

Discovering Influencers for Spreading in Weighted Networks

Zahra Rezaei
Department of Industrial Engineering, IT Group
K.N. Toosi University of Technology
Tehran, Iran
zrezaei@mail.kntu.ac.ir

Mohammad Jafar Tarokh
Department of Industrial Engineering, IT Group
K.N. Toosi University of Technology
Tehran, Iran
mjtarokh@kntu.ac.ir

Received: February 28, 2014-Accepted: June 24, 2015

***Abstract*— Identifying the influential nodes in networks is an important issue for efficient information diffusion, controlling rumors and diseases and optimal use of network structure. The degree centrality which considers local topology features, does not produce very reliable results. Despite better results of global centrality such as betweenness centrality and closeness centrality, they have high computational complexity. So, we propose semi-local centrality measure to identify influential nodes in weighted networks by considering node degree, edges weight and neighboring nodes. This method runs in $O(n(k)^2)$.So, it is feasible for large scale network. The results of applying the proposed method on weighted networks and comparing it with susceptible-infected-recovered model, show that it performs good and the influential nodes are generated by our method can spread information well.**

Keywords-component; influential nodes; degree semi-local centrality; centrality measures; weighted networks

I. INTRODUCTION

Identifying influential nodes that lead to faster and wider spreading in social networks, has attracted the attention of the researchers for years and still is one of the hot topics in the field of social networks (e.g. [1-4]). It shows new applications such as finding social leaders, designing viral marketing strategies (e.g. [5]), controlling rumor and disease spreading [6] and measure of information flow (e.g.[7]). The main issue is how to measure the ability of a node to spread a message to a sufficiently large portion of the network [4]. Many methods are proposed to measure the power of the nodes and identify the influential nodes in a social network (e.g. [8-13]). Literature review which is undertaken in this field shows that we can classify these methods as follow:

1. Network-Based Approaches:

Any method that analyzes the explicit relationship links or topological structure of a network and or evaluates the social interaction such as comments and citations, are considered in network-based approaches category (e.g. [4], [9], [11], [14-22]). These approaches are the most common method to identify the influential nodes.

2. Using diffusion models:

Simulation and modeling of information diffusion process between nodes and their neighbors by epidemic models, are another way to identify the influential nodes. The methods such as greedy algorithm that solve top-k influential nodes by taking into account diffusion mechanism, can also be considered in this category (e.g. [13], [23-25]).

3. Using factors associated with users and content analysis :

In social network, users are different in the characteristics, frequency and quality of content generated. So, some users' opinions will more influence other users' views and some will not. Thus, we can identify the influential nodes with analyzing the factors associated with users and or analyzing the content which users will send to network (e.g. [10], [12]).

4. Using combination of above approaches:

In this way, combination of above approaches are used (e.g. [8], [11]).

The simplest methods consider only local topological features of the nodes in a network graph which can point out degree centrality [26]. Despite its simplicity and low computational complexity, it is less relevant [2]. Closeness centrality and betweenness centrality are global measures that can give better results though they have high computational complexity [2]. So, it is difficult or even infeasible to apply them in large-scale networks. [2] is proposed a semi-local centrality measure, called nearest and next nearest neighbors, as a tradeoff between the low-relevant degree centrality and other time-consuming measures.

Most of the measures that are presented are designed for binary networks. Since many real networks are believed to be weighted, attempts have been made to extend them to weighted networks (e.g. [27-30]). All these attempts have only focused on edges weight, and not on the number of edges. Accordingly, [31] extended these measures for weighted networks by considering the edges weight and number of edges.

Other proposed measures are as follows. In laplacian centrality, the importance of a vertex v is reflected by the drop of the laplacian energy of the network to respond to the deactivation of the vertex from the network [32]. C_g index [33] considers not only node degree and edges weight but also strength of the neighboring nodes. Evidential centrality (EVC) [34] which is based on the Dempster–Shafer theory is obtained by the combination of degree and weight strength of each node. Since the EVC has ignored the global structure information of the network, evidential semi-local centrality (ESC) measure [6] is proposed. The ESC considers modified evidential centrality and the extension of semi-local centrality in weighted networks and obtained results which are more reasonable than the EVC. Weighted k-shell decomposition [35] is modified original k-shell decomposition method to identify the influential nodes. According to k-shell decomposition method, the most influential nodes are those located in core layers.

Spanning tree centrality (STC) [36] is proposed to measure the centralities of nodes in a weighted network. The STC score of a vertex *v* in *G* is defined as the number of spanning trees with the vertex *v* as a cut vertex [36].

We propose degree semi-local centrality (DSC) based on analysing topological structure of network with linear complexity, which identifies influential nodes by considering node degree, edge weight and importance of neighboring nodes. To evaluate the performance of the proposed method, we adopt susceptible-infected-recovered (SIR) model (e.g. [6], [9], [34], [37], [38]). Our experimental results on four networks, compared with the SIR model, show that this method can identify the influential nodes effectively. The rest of the paper is organized as follows. We briefly review previous studies in section 2 and describe our centrality measure with example network in section 3. In section 4, we apply the SIR model to evaluate the effectiveness of the proposed method in weighted networks. Then we present the Results and Conclusions in section 5 and 6.

II. RELATED WORKS

Many centrality measures for nodes ranking on a network are presented. Most of them are designed for unweighted networks. [15] is presented a new method based on TOPSIS approach which is a multiple attribute decision making technique to identify the influential nodes. This method is calculated the value of different centrality measures which are considered as the multi-attribute in the TOPSIS. Then, TOPSIS is utilized to identify influential nodes. [16] is proposed a multi-attribute ranking method based on TOPSIS to evaluate the node importance from many perspective such as DC, BC, CC and improved K-shell .Improved K-shell decomposition [16] is the indicator which gives a more precise distinction of local characteristic differences between nodes than K-shell decomposition. [17] is proposed new method by combining global diversity and local features to identify the most influential network nodes. In the first step, global node information is obtained using algorithms such as a community detection algorithm and k-shell decomposition algorithm. In the second step, local node information is acquired through the use of various types of local centrality, including degree centrality. Last, global diversity and local features are combined to determine node influence [17]. The most well-known of these measures are the degree centrality (DC), betweenness centrality (BC) and closeness centrality (CC). [39] is defined these three measures for unweighted networks.

Comparing with various measures developed for unweighted networks, little work has been done yet for weighted networks [36]. In [31], is attempted to extend Freeman's measures for weighted network as follows. The degree centrality of node i, denoted as $C_D^{w\alpha}(i)$, is defined as:

$$C_D^{w\alpha}(i) = k_i \times (\frac{s_i}{k_i})^\alpha \tag{1}$$

Where k_i is the degree of node i, s_i is the sum of edges weight located on node i and α is a positive tuning parameter which determines the relative importance of degree compared to edges weight. In this paper, the degree centrality is equation (1). The degree centrality considers only edge weight and degree of the node in the network. So, when two nodes have the same degree and strength but have different structural properties, the reasonable result is not achieved. For example in Fig.1, according to the equation (1), degree centrality of nodes A and B are equal to 3.2. While the structural position of node B is better than node A because node B has neighbors with degrees $k_C = 7$, $k_D = 6$, while node A has neighbors with degrees $k_E = 2$, $k_F = 3$.

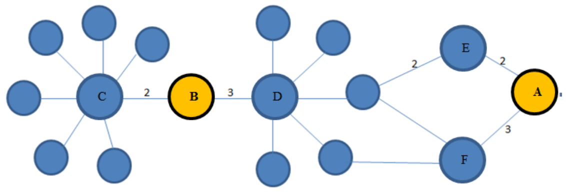


Fig.1. Same degree centrality of nodes A and B, but different structural properties

The betweenness centrality of node i, denoted as $C_B^w(i)$, is defined as:

$$C_B^w(i) = \sum_{i \neq j \neq k} \frac{g_{jk}^w(i)}{g_{jk}^w} \qquad (2)$$

Where g_{jk}^w, is the number of shortest paths between node j and k and $g_{jk}^w(i)$ is the number of those paths that go through node i. Computational complexity of betweenness centrality for weighted network by using Brandes' algorithm is O(n^2 log n+ nm) [40]. In addition to the high computational complexity, it has also another limitation. Betweenness centrality relies on the idea that, in social networks information flows only along shortest paths while messages generated in a source node and directed toward a target node in the network, may flow along arbitrary paths [26].

The closeness centrality of node i, denoted as (i), $C_C^w(i)$ is defined as :

$$C_C^w(i) = \left[\sum_j^N d^w(i,j)\right]^{-1} \qquad (3)$$

Where $d^w(i,j)$ is the shortest distance between node i and node j. The closeness and betweenness centrality measures rely on the identification of the shortest paths among nodes in a network. Calculating the shortest paths between all pairs of nodes in a network takes the complexity O(n^2)with the Floyd's algorithm.

The weighted k-shell decomposition [35] is modified k-shell decomposition method based on adding the degree of its two end nodes as edge weight. According to k-shell decomposition method, inner core layer nodes are more important than periphery layer nodes [17]. The experimental results show that this method is comparable with the local centrality and coreness centrality in identifying the influential nodes [35].

The STC [36] is proposed to measure the centralities of nodes. This centrality is based on that, if vertex v is central in the network, the probability that node v acts as cut-vertices in spanning trees are high. So, node v is an important node in the network, if it has high number of spanning trees of G with v as a cut-vertex.

The c_g index of node i is c_g if c_g is the highest rank so that the sum of the products of the edge strength of the top c_g node and the strength of corresponding neighboring node is at least c_g^2 [33]. As, c_g index of node i is calculated as follows: $a^2 \leq c_g(i) = \sum_{j \in \Gamma(i)} w_{ij} s_j \leq b^2$, the possibility that c_g index of several nodes are equal, is more. So, nodes that are ranked by this method don't have high accuracy.

The evidential semi-local centrality [6] which is based on the Dempster–Shafer theory is combination of modified evidential centrality and the extension of semi-local centrality in weighted networks**.** ESC value of node i, denoted as ESC (i), is defined as:

$$Q^w(u) = \sum_{v \in \Gamma(u)} N^w(v) \qquad (4)$$

$$\text{ESC}(i) = \sum_{j \in \Gamma(i)} Q^w(j) \qquad (5)$$

Where $\Gamma(i)$ is the set of nearest neighbors of node i and $N^w(v)$ is the sum of MEC of node v and its nearest and next nearest neighbors.

The semi-local centrality of node i $C_L(i)$ [2], is defined as :

$$Q(u) = \sum_{w \in \Gamma(u)} N(w) \qquad (6)$$

$$C_L(i) = \sum_{u \in \Gamma(i)} Q(u) \qquad (7)$$

Where $\Gamma(u)$ is the set of the nearest neighbors of node u and $N(w)$ is the number of the nearest and the next nearest neighbors of node w.

III. ROPOSED METHOD

As we mentioned, semi-local centrality C_L is proposed for unweighted networks. According to its good performance and linear complexity, we combined it with degree centrality [31], to be applied in weighted networks. This proposed measure is called degree semi-local centrality (DSC). The DSC of node v is defined as:

$$Q^w(u) = \sum_{j \in \Gamma(u)} N^w(j) \qquad (8)$$

$$\text{DSC}(v) = \sum_{u \in \Gamma(v)} W_{vu} Q^w(u) \qquad (9)$$

Where $\Gamma(u)$ is the set of the nearest neighbors of node u and $N^w(j)$ is the sum of degree centrality of node j and its nearest and next nearest neighbors. In the other words, $N^w(j)$ is the sum of degree centrality of node j and all neighbors of node i with 1-hop and 2-hop distance. According to c_g index that considers edge weight between node i and its neighbor node j and Considering the fact that individuals are more likely to be influenced by their stronger ties, the coefficient W_{vu} is added to the equation (9). In equation (8), information of nearest and next nearest neighbors are considered. In a word, the DSC measure assigns high score to the node based on the adjacency of that node with nodes which have high-Q^w and are connected with high weight to it.

For example, the DSC of node 5 is calculated as follow. According to Fig.2, node 5 has five nearest neighbors including nodes from 1 to 4 and 6 ($\Gamma(5)$ ={1,2,3,4,6}). So, the DSC of node 5 is:

DSC(5) is:

$$(2\times Q^w(1))+(2\times Q^w(2))+(4\times Q^w(3))+(2\times Q^w(4))+(3\times Q^w(6))$$

Thus, we first calculate $Q^w(1)$, $Q^w(2)$, $Q^w(3)$, $Q^w(4)$ and $Q^w(6)$. According to (8), $Q^w(1)$ is:

$Q^w(1) = N^W(5)$ (node 1 has one neighbor $\Gamma(5)$ ={5}). Node 5 has five nearest neighbors including nodes from 1 to 4 and 6 and two next nearest neighbors including Nodes 12 and 13. Thus, the N^W of node 5 is:

$$N^W(5) = C_D^{w\alpha}(1) + C_D^{w\alpha}(2) + C_D^{w\alpha}(3) + C_D^{w\alpha}(4) + C_D^{w\alpha}(5) + C_D^{w\alpha}(6) + C_D^{w\alpha}(12) + C_D^{w\alpha}(13) = 30.6632$$

Similarly, we can calculate the values of $Q^w(2)$, $Q^w(3)$, $Q^w(4)$ and $Q^w(6)$. Finally, degree semi-local centrality of node 5 is:

$$DSC(5) = (2\times Q^w(1)) + (2\times Q^w(2)) + (4\times Q^w(3)) + (2\times Q^w(4)) + (3\times Q^w(6)) = 649.842$$

The values of DSC(v) for other nodes are presented in the fifth column of Table 1. The values of evidential semi-local centrality and c_g index of the nodes in Fig.2, are shown in the sixth and seventh columns of Table 1 respectively. Also, the values of two global centrality measures, namely the closeness centrality and betweenness centrality of the nodes for comparing performance of proposed semi-local measure with global centrality measures, are represented respectively in the eighth and ninth columns of Table 1.

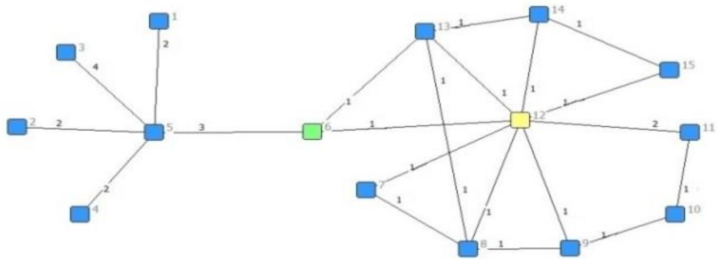


Fig.2. Weighted example network with 15 nodes [6]

Table 1. Scores of 15 nodes of weighted network of Fig.2 based on various centrality methods

v	$C_D^{w\alpha}$	N^w	Q^w	DSC	ESC	C_g	CC	BC
1	1.4142	18.1779	30.6632	239.649	1.9031	5	0.0254	0
2	1.4142	18.1779	30.6632	239.649	1.9031	5	0.0254	0
3	2	18.1779	30.6632	479.297	1.9031	7	0.0288	0
4	1.4142	18.1779	30.6632	239.649	1.9031	5	0.0254	0
5	8.0623	30.6632	119.8243	649.842	13.3217	6	0.0332	46
6	3.873	47.1127	114.4034	807.901	22.9086	7	0.0361	45
7	2	32.8078	77.6779	442.185	17.2707	3	0.0238	0
8	4	34.8078	151.3558	623.719	23.9851	4	0.0260	1.5
9	3	34.8078	97.6127	511.801	19.2095	3	0.0251	1
10	2	19.9348	69.6156	160.418	6.7144	2	0.0206	0
11	2.4495	34.8078	62.8049	651.275	13.4645	4	0.0305	11
12	8.4853	42.8701	290.83	916.484	45.96	5	0.0402	53.5
13	4	40.8701	157.5984	673.137	27.7556	4	0.0312	6.5
14	3	32.8078	116.548	524.106	20.1432	3	0.0246	0.5
15	2	32.8078	75.6779	407.378	16.3013	3	0.0238	0

Nodes 3 and node1 have the same nearest and next nearest neighbors but the edge weight of node 3 is twice the edge weight of node 1. Thus in rating list, it is better that node 3 is located above node1. While in the ranking results of ESC in Table 1, this case is not observed. The result of SIR model on this network also shows that node 3 is located above node1 in rating list.

The DSC measure which uses information of nearest and next nearest neighbors, is likely to be more effective to identify influential nodes than degree centrality and c_g index. Because it considers more information about nodes than degree centrality and c_g index.

Since to calculate $N^W(j)$ need to calculate the nearest and next nearest neighbor of node j, the computational complexity of this measure is $O(n(k)^2)$ where n is the number of nodes in the graph and k is the average degree of the network. This computational complexity is much lower than the betweenness complexity O(n^2 logn + nm) and closeness complexity O(n^3).

IV. EVALUATE THE PERFORMANCE OF PROPOSED METHOD

To evaluate the performance of the proposed method, two classical weighted data sets Zachary's karate club network [41] and Freeman's EIES network [42] are used. In addition, we used weighted example network of Fig.2 and Fig.3. These two networks are chosen to compare the results of ESC with DSC.

The data set of Zachary's Karate Club Network was collected from the 34 members of a university karate club by Wayne Zachary over two years. In the weighted network which is shown in Fig.4, each node represents each member in the club, each edge represents a relationship of friendship between two members outside of club activities and the weight assigned to each edge is relative strength of the relationship.

The dataset of Freeman's EIES was collected in 1978 and contains three different network relations among researchers working on social network analysis. The first network is the inter-personal relationships among the researchers at the beginning of the study. The second network is the inter-personal relationships among the researchers at the end of the study. In these two networks, the edges weight are proportional to the intensity of the relationship between researchers.

The third network is different from the two other networks. The edges weight in the third network are defined as the number of messages sent among 32 researchers on an electronic communication tool. In this paper, the third network is considered. In Table 2, statistical characteristics of the networks are given.

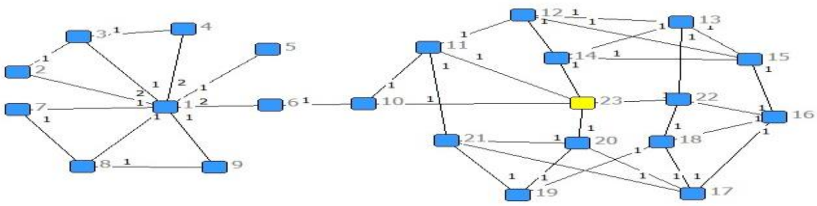


Fig.3. Weighted example network with 23 nodes [6]

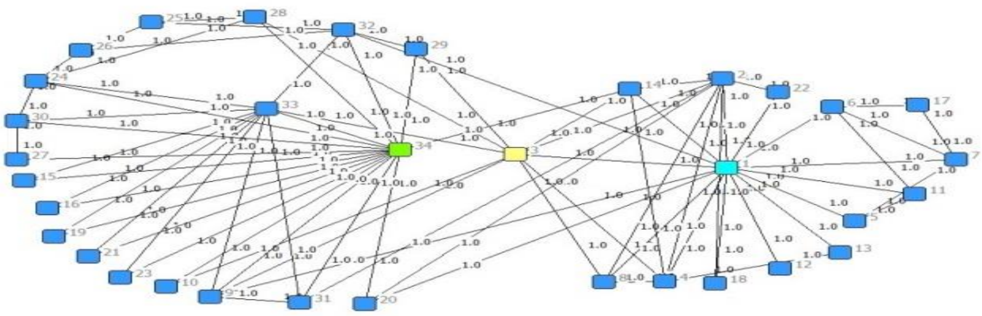


Fig.4. Zachary's karate club network

Table 2. Statistical characteristics of the networks, N is the number of nodes, $\lambda_{\text{threshold}}$ is spreading threshold for network [37] and β is constant used in model SIR according to equation (10)

Networks	N	$\lambda_{\text{threshold}}$	β
Weighted network with 15 nodes (Fig.2)	15	0.25	0.55
Weighted network with 23 nodes (Fig.3)	23	0.25	0.8
Freeman's EIES network	32	0.05	0.9
Zachary's karate club network	34	0.13	0.8

To assessment the performance of proposed method, we use the susceptible-infected-recovered (SIR) model to check spreading influence of nodes [6], [37], [38]. The spreading influence of node v $\delta(v)$ is defined as the number of infected nodes averaged over a sufficiently number of simulations [4]. In this model, initially we set a node v which we want to investigate the ability to spread, be infected and all other nodes are set to be susceptible. Then, each infected node after attempts to infect its susceptible neighbors with infection probability λ_{ij} , is recovered. In weighted networks, infection probability is:

$$\lambda_{ij} = \left(\frac{w_{ij}}{w_M}\right)^{\beta} \qquad \beta > 0 \tag{10}$$

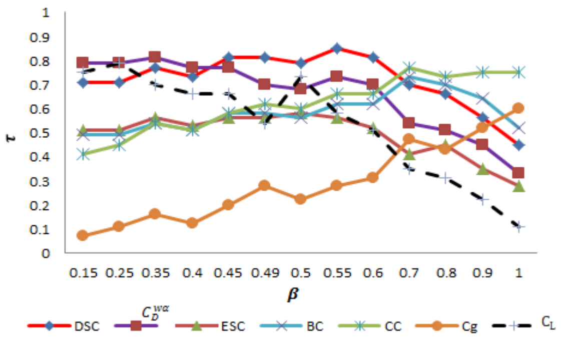
[43], at which susceptible node i acquire the infection from the infected neighbor j, β is a positive constant and w_{ij} is the edge weight between node i and j. Since $\frac{w_{ij}}{w_M} \leq 1$, the smaller value of β causes the infection spread more quickly. This spreading process is repeated until there remains no infected node in the network. The number of infected nodes at the end of a spreading process over a enough large number of simulations is an indicator to estimate the ability to spread of the initial infected node v. We set the number of simulations to be 10000.

In addition, we use correlation coefficient, the kendall's tau τ, to measure the correlation between ranked list of nodes by our propose method and the one generated by the SIR model. The higher of the kendall's tau value shows the higher accuracy of the method. Whatever, the correlation coefficient τ is closer to 1, the method is more corresponded to the SIR model.

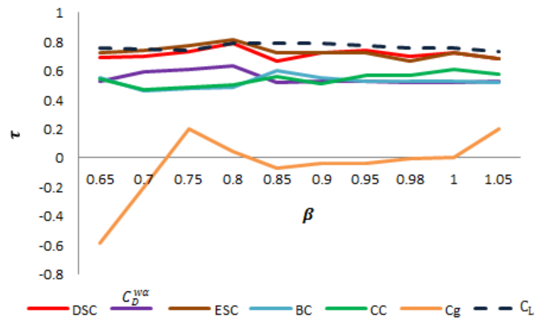
V. EXPERIMENTAL RESULTS

For the Freeman's EIES and Zachary's karate club networks, the kendall's tau τ for the DSC, degree centrality $(C_D^{w\alpha})$, BC, CC, c_g index, semi-local centrality (C_L) and STC are shown in Fig.5. Also, for the weighted networks of Fig.2 and Fig.3, the kendall's tau τ for the DSC, $C_D^{w\alpha}$, ESC, BC, CC, c_g index and C_L , are represented in Fig.5.

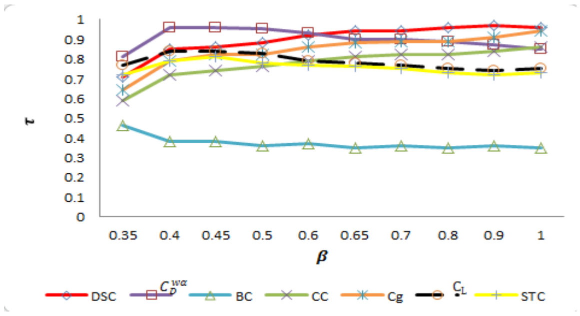
It can be seen from Fig.5(a), the DSC presents the best performance among other methods where the infection probability λ_{ij} is larger than the epidemic threshold $\lambda_{\text{threshold}} = \frac{k}{k^2}$ [37].



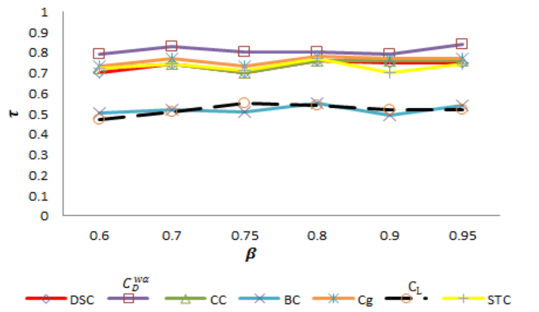
(a) Example network with 15 node



(b) Example network with 23 node



(c) Freeman's EIES Network



(d) Zachary's Karate Club

Fig. 5. The correlation coefficient, the Kendall's tau τ, is plotted by change the value of β and consequently change the value of λ in weighted networks of Fig.2, Fig.3, Freeman's EIES and Zachary's karate club networks. The results are averaged 10000 runs with different value of β.

As described earlier, according to equation (10) by choosing small values of β we can increase the infection probability λ_{ij}. For example in Fig. 2, w_M is 4 and w_{ij} changes between 1, 2 and 4. If for example, we consider β =0.55, the λ_{ij} will be much larger than $\lambda_{\text{threshold}} \approx 0.25$ and DSC has maximum correlation coefficient (τ=0.85) with the SIR model. So, in this network, nodes with higher DSC have higher influence. In Fig.5(b), the C_L, ESC and DSC present better result in this network. Although, the C_L and ESC perform slightly better than DSC. To clarify the performance of three methods in this network, we chose top-10 node of influential nodes that are ranked by the DSC, $C_D^{w\alpha}$, ESC, BC, CC, c_g and C_L. Then, the average value of $\delta(v)$ over top-10 nodes on each method is calculated. The results show that the $\delta(v)$ average value of DSC is 14.85 $(\delta_{DSC} = 14.85)$. The $\delta(v)$ average value of other methods are $\delta_{ESC} = 14.96$, $\delta_{C_L} = 14.96$, $\delta_{BC} = 13.35$, $\delta_{CC} = 12.01$, $\delta_{Cg} = 11.09$ and $\delta_{C_D^{w\alpha}} = 14.81$. It means that, if top-10 nodes of DSC are infected, they can infect 14.85 nodes in average.

In Fig.5(c), the DSC and $C_D^{w\alpha}$ have better performance in the Freeman's EIES Network with correlation 0.97.

In Zachary's Karate network, although, the $C_D^{w\alpha}$ performs best and has maximum correlation coefficient but the DSC measure also has good performance. In order to clarify the performance of the DSC measure, we present the top-10 nodes in Table 3 and Table 4 as ranked by BC, CC, c_g index, $C_D^{w\alpha}$, C_L, STC and DSC in the Zachary's Karate Club and Freeman's EIES Network. The value of $\delta_{DSC}(v)$ that is represented in eighth column is the number of total infected nodes averaged over 10000 implementations by nodes that are ranked by DSC. Table 5 shows the average value of $\delta(v)$ over top-10 nodes on the centralities that are mentioned.It can be seen from Table 5 and Fig.5(c), in the Freeman's EIES network, DSC measure and degree centrality have maximum correlation coefficient (τ=0.97) with the SIR model and if, top-10 of influential nodes that are ranked by DSC or degree centrality are infected in this network, they can infect 12.32 nodes in average that is greater than the $\delta(v)$ average value of other measures. So, they have better performance in this network. They perform slightly better than the STC, CC and c_g index. The performance of BC compare to the other centrality is not good.In the Zachary's Karate network, although the degree centrality is more corresponded to the SIR model than other, but DSC also identified influential node well. As top- 10 influential nodes that are ranked by DSC are infected, they can infect 24.57 nodes in average. It is equal to the $\delta(v)$ average value of degree centrality ($\delta_{C_D^{w\alpha}} = 24.57$). Also, according to [6], node 3 is most influential node or top-1 node in this network that is identified by the DSC, CC and c_g.

Table 3. The top-10 ranked nodes of the Zachary's Karate Club by the BC, CC, c_g, $C_D^{w\alpha}$, C_L, STC , DSC.

BC	CC	C_g	$C_D^{w\alpha}$	C_L	STC	DSC	$\delta_{DSC}(v)$
1	3	3	34	1	1	3	25.043
3	34	33	1	3	34	33	24.852
34	9	34	33	34	33	34	24.915
33	1	1	3	33	2	2	24.85
9	14	2	2	9	3	1	24.847
32	2	9	32	14	4	9	24.356
2	32	14	4	2	32	14	24.275
4	33	24	24	32	6	32	24.519
14	28	32	9	4	24	24	24.493
7	24	4	14	31	7	4	23.528

Table 4. The top-10 ranked nodes of the Freeman's EIES Network by the BC, CC, c_g index, $C_D^{w\alpha}$, C_L, STC , DSC.

BC	CC	C_g	$C_D^{w\alpha}$	C_L	STC	DSC	$\delta_{DSC}(v)$
1	1	1	1	1	1	1	15.1923
29	29	29	29	24	29	29	15.4903
8	2	8	8	31	2	8	14.4152
2	8	2	31	5	31	32	13.6742
32	32	32	2	29	8	2	13.4666
4	11	11	24	8	24	11	12.5139
24	24	31	32	2	11	31	12.4711
10	15	24	11	4	32	24	10.7769
3	31	15	10	10	10	10	7.6576
5	30	30	4	11	27	30	7.5186

Table 5. Average value of $\delta(v)$ over top-10 nodes on seven centrality

Network	DSC	C_L	STC	BC	CC	C_g	$C_D^{w\alpha}$
Freeman's EIES Network	12.32	11.36	12.28	10.39	12.24	12.24	12.32
Zachary's Karate network	24.57	24.30	23.65	24.06	24.49	24.57	24.57

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we focused on providing a method to identify the influential nodes in weighted networks so that it is applicable in large-scale networks and has low computational complexity. So, we proposed a semi-local centrality measure, the degree semi local centrality (DSC), which is based on analysing topological structure of network. It considers not only node degree and edges weight, but also neighboring nodes. We applied this method on four weighted networks. Then, we evaluated the effectiveness of our method by comparing it with the SIR model. The experimental results show that, our method performs good and the influential nodes are generated by our method can spread information well. Considering that, the DSC measure considers node degree and edges weight and utilizes information of nearest and next-nearest neighbors of each node, is likely to be effective to identify the influential nodes than the degree centrality and C_g index. Because the degree centrality considers only local topology features (node degree, edges weight) and does not consider adjacency importance of one node to other nodes. The C_g index utilizes information of nearest neighbors of each node, While DSC measure considers information of nearest and next-nearest neighbors. For example in Fig.6, although node 22 has only four neighbors and weak edges weight, but its neighbors have connections with other network nodes. So, if node 22 is infected, it can infect more nodes through its neighbors. For this reason, the $\delta(v)$ value of node 22 is high and equal to 15.0304. Our method identified this node as influential node while the C_g index and degree centrality can't identify it.

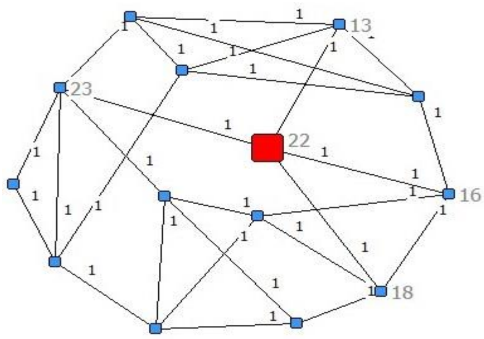


Fig.6. The local structure surrounding node 22 in the weighted example network with 23 nodes

In semi local centrality C_L, edges are treated equally while it is important to take into consideration the edges weight when the centrality measures are designed. The experimental results show it.

In addition, its computational complexity $O(n(k)^2)$ is less than the computational complexity of the betweenness centrality O(n^2 logn+nm) and the closeness complexity $O(n)^3$ .So, it is feasible for large scale networks.Generally, since different measures such as the ones which are mentioned focus on different aspects of network structure, we can't say which measure is always the best [36]. The selection of centrality method depends on not only the network's structure, but also the user's aim or goal [36].

The current work has limitations and can be improved in the future. the DSC and all previous measures which are mentioned, assume that the direct relation between two nodes is symmetrical. Nevertheless, it is easy to find situations in which the connections are directed, having a specific sense. Therefore, we propose, this method will be extended for directed weighted networks. In discovery of influential users, content analysis and activeness based factors such as number of post by user in a time interval, login information to the site over time, are important factors. So, we will take them into consideration in future work.

REFERENCES

[1] Kiss, C., and Bichler, M., " Identification of influencers — Measuring influence in customer networks.", Decision Support Systems, Vol.46 No.1, pp.233–253, 2008.

[2] Chen, D., Lü, L., Shang, M.-S., Zhang, Y.-C., and Zhou, T, " Identifying influential nodes in complex networks", Physica A: Statistical Mechanics and its Applications, Vol.391 No.4, pp.1777–1787, 2012.

[3] Nettleton, D. F., " Data mining of social networks represented as graphs", Computer Science Review, Vol.7, pp.1–34, 2013.

[4] Bae, J., and Kim, S., " Identifying and ranking influential spreaders in complex networks by neighborhood coreness", Physica A: Statistical Mechanics and its Applications, Vol.395, pp. 549–559, 2014

[5] Domingos, P. Richardson, M., " Mining the network value of customers", In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining , pp.57–66, 2001

[6] Gao, C., Wei, D., Hu, Y., Mahadevan, S., and Deng, Y., " A modified evidential methodology of identifying influential nodes in weighted networks", Physica A: Statistical Mechanics and its Applications, Vol.392 No.21, pp.5490–5500, 2013.

[7] Grassi, R., " Vertex centrality as a measure of information flow in Italian Corporate Board Networks", Physica A: Statistical Mechanics and its Applications, Vol.389 No.12, pp.2455–2464, 2010.

[8] Agarwal, N., Liu, H., Tang, L., and Yu, P. S., "Identifying the influential bloggers in a community", Proceedings of the international conference on Web search and web data mining ", pp. 207–217, 2008.

[9] Kitsak, M., Gallos, L., Havlin, S., and Liljeros, F., " Identification of influential spreaders in complex networks", Nature Physics, Vol.6 No.11, pp.6–11, 2010.

[10] Li, Y.-M., Lin, C.-H., Lai, C.-Y., Identifying influential reviewers for word-of-mouth marketing. Electronic Commerce Research and Applications, 9(4), 294–304, 2010.

[11] Li, Y.-M., Lai, C.-Y., Chen, and C.-W., " Discovering influencers for marketing in the blogosphere", Information Sciences, Vol.181 No.23, pp.5143–5157, 2011.

[12] Aral, S., and Walker, D., "Identifying influential and susceptible members of social networks", Science New York, N.Y., Vol.337 No.6092, pp. 337–41,2012.

[13] Zhang, X., Zhu, J., Wang, Q., and Zhao, H., "Identifying influential nodes in complex networks with community structure", Knowledge-Based Systems, Vol. 42, pp. 74–84, 2013.

[14] Lü, L., Zhang, Y., Yeung, C. H., Zhou, T., "Leaders in social networks", the Delicious case. PloS one, 6(6), e21202, 2011.

[15] Du, Y., Gao, C., Hu, Y., Mahadevan, S., & Deng, Y, "A new method of identifying influential nodes in complex networks based on TOPSIS", Physica A, Vol.399, pp.57–69, 2014.

[16] Liu, Z., Jiang, C., Wang, J., & Yu, H., "The node importance in actual complex networks based on a multi-attribute ranking method", Knowledge-Based Systems, Vol.84, pp.56–66, 2015.

[17] Fu, Y., Huang, C., & Sun, C., "Using Global Diversity and Local Topology Features to Identify Influential Network Spreaders". Physica A, Vol.433, pp.344-355, 2015.

[18] Gao, C., Zhong, L., Li, X., Zhang, Z., & Shi, N., "Combination methods for identifying influential nodes in networks", International Journal of Modern Physics C, Vol.26 No.06, 1550067, 2015.

[19] Zeng, A., & Zhang, C., "Ranking spreaders by decomposing complex networks". Physics Letters A, Vol.377 No.14, pp.1031–1035, 2013.

[20] Basaras, P., Katsaros, D., & Tassiulas, L., "Detecting Influential Spreaders in Complex", Dynamic Networks. Computer, Vol.46 No.4, pp.24–29, 2013.

[21] Silva, A., Guimarães, S., Meira, W., & Zaki, M., "ProfileRank", Proceedings of the 7th Workshop on Social Network Mining and Analysis - SNAKDD '13, pp. 1–9,2013 .

[22] Li, Q., Zhou, T., Lü, L., and Chen, D., Identifying influential spreaders by weighted LeaderRank. Physica A: Statistical Mechanics and its Applications, Vol.404, pp. 47–55, 2014.

[23] Zhu, Z., "Discovering the influential users oriented to viral marketing based on online social networks", Physica A, Vol.392 No.16, pp.3459–3469, 2013.

[24] Wang, Y., Cong, G., Song, G., & Xie, K., "Community-based greedy algorithm for mining top-K influential nodes in mobile social networks", Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining ,pp. 1039, 2010.

[25] Bauer, F., & Lizier, J. T., "Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: A walk counting approach". EPL (Europhysics Letters), Vol.99 No.6, 68007, 2012.

[26] De Meo, P., Ferrara, E., Fiumara, G., and Ricciardello, A., " A novel measure of edge centrality in social networks", Knowledge-Based Systems, Vol.30, pp.136–150, 2012.

[27] Newman, M. E. J., " Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality", Physical Review E, Vol.64 No.1, pp.016132, 2001.

[28] Newman, M. E. J., " Analysis of weighted networks", Physical Review E, Vol.70 No.5, pp.056131, 2004.

[29] Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A., "The architecture of complex weighted networks", Proceedings of the National Academy of Sciences of the United States of America, Vol. 101 No.11, pp.3747–52, 2004.

[30] Opsahl, T., Colizza, V., Panzarasa, P., and Ramasco, J., " Prominence and Control: The Weighted Rich-Club Effect", Physical Review Letters, Vol.101 No.16, pp.168702, 2008.

[31] Opsahl, T., Agneessens, F., Skvoretz, J., "Node centrality in weighted networks : Generalizing degree and shortest paths", Social Networks, Vol.32 No.3, pp.245–251, 2010.

[32] Qi, X., Fuller, E., Wu, Q., Wu, Y., and Zhang, C.-Q., " Laplacian centrality: A new centrality measure for weighted networks", Information Sciences, Vol.194, pp.240–253, 2012.

[33] Yan, X., Zhai, L., and Fan, W.," C-index: A weighted network node centrality measure for collaboration competence", Journal of Informetrics, Vol.7 No.1, pp. 223–239, 2013.

[34] Wei, D., Deng, X., and Zhang, X., " Identifying influential nodes in weighted networks based on evidence theory", Physica A: Statistical Mechanics and its Applications, Vol.392 No.10, pp.2564–2575, 2013.

[35] Wei, B., Liu, J., Wei, D., Gao, C., & Deng, Y., "Weighted k-shell decomposition for complex networks based on potential edge weights". Physica A: Statistical Mechanics and its Applications, Vol.420, pp.277–283, 2015.

[36] Qi, X., Fuller, E., Luo, R., & Zhang, C., "A novel centrality method for weighted networks based on the Kirchhoff polynomial". Pattern Recognition Letters, Vol.58, pp.51–60, 2015.

[37] Chu, X., Zhang, Z., Guan, J., and Zhou, S., "Epidemic spreading with nonlinear infectivity in weighted scale-free networks", Physica A: Statistical Mechanics and its Applications, Vol.390 No.3, pp.471–481, 2011.

[38] Liu, J.-G., Ren, Z.-M., Guo, Q., Ranking the spreading influence in complex networks. Physica A: Statistical Mechanics and its Applications, 392(18), 4154–4159, 2013.

[39] Freeman, L., "Centrality in social networks conceptual clarification", Social networks, Vol.1 No.3, pp. 215–239, 1979.

[40] Brandes, U., "A faster algorithm for betweenness centrality", The Journal of Mathematical Sociology, Vol.25 No.2, pp.163–177, 2004.

[41] Zachary, W. W. ," Flow Modelfor Conflict An Information Fission in Small Groups" Anthropological Research,Vol. 33 No.4, pp.452–473, 1977.

[42] Freeman S.C., and Freeman L.C., " The networkers network: a study of the impact of a new communications medium on sociometric structure", Social Science Research Reports S46, University of California, Irvine, CA, 1979.

[43] Gang, Y., Tao, Z., Jie, W., Zhong-Qian, F., and Bing-Hong, W. ,"Epidemic Spread in Weighted Scale-Free Networks", Chinese Physics Letters, Vol.22 No.2, pp.510–513, 2005.

Zahra Rezaei is a M.Sc. student in Electronic Commerce at K.N. Toosi University of Technology, Tehran, Iran. She received her B.Sc. degree in information technology from Sheikh Bahaei University, Isfahan, Iran. Her current primary research interests are in social networks and e-marketing.

Mohammad J. Tarokh is an associate professor in the department of Industrial Engineering at K.N. Toosi University of Technology, Tehran, Iran. He received his B.Sc. from Sharif University of Technology in Tehran, his M.Sc. from University of Dundee in UK and his Ph.D. from University of Bradford, UK. His main research interests are in knowledge management, business intelligence, customer relationship management and supply chain management.

IJICTR

This Page intentionally left blank.