The centrality ranking algorithm based on the network structure

KAI ZHANG,YINGHONG MA
Shandong Normal University
School of Management Science and Engineering
Jinan
CHINA
zhangkai80224@163.com,yhma@sdnu.edu.cn

Abstract: Centrality is an important measurement of network structure. The node centrality ranking algorithm is to calculate the significant degree of nodes in social networks. An improved algorithm of ranking node centrality is present in this issue. Numerical simulation showed the feasibility and validity of the node centrality ranking algorithm, the convergence of the algorithm is proved by Perron-Frobenius Theorem. Comparing with the previous centrality algorithms, it exhibited more efficiency in computing complexity. Meanwhile, an edge centrality ranking algorithm and its efficiency is discussed too.

Key–Words: centrality; ranking algorithm; network structure

1 Introduction

social networks have permeated into everywhere of our society,such as online social networks, research networks, and traffic networks, etc. With the development of information science, big data is used in biological, social, and technological systems etc. The structured data often represent by complex networks, the node is the research object and the edge is the interaction or relationship between objects. Not only from a scientific perspective but also for commercial or strategic motivations, the identification of the centrality actors inside a network is very important. For example, when the critical node is attacked, the network will be paralyzed. The centrality of nodes, or the identification of which nodes are more central than others, has been a key issue in network analysis, which has many implications in information flows, bargaining power, infection transmission, influence and other sorts of important behaviors on a network.

Measures of the node importance ranking can be categorized into four main groups depending on the network structure which they are based. Many authors suggested the local social network topology to compute centrality values. Degree measured the involvement of the node in the network, and its simplicity is advantage: only the local structure around a node must be known for it to be calculated [1,2];The second groups are eigenvectors, neighbors characteristics and closeness [3-5], although these measures take the global network structure into consideration ,and can be applied to networks with disconnected components, it is not without limitations; Some authors introduced a novel node centrality measure known as K-kernel centrality [6,7];In the spirit of information propagation models, some authors suggested to perform random walks on the social network to compute centrality values [8].

Some centrality metrics are based on PageRank algorithm and HITS algorithm [9]. The PageRank of a node can be interpreted as the weight of the node respecting to the stationary distribution of an associated homogeneous Markov chain or in other words, the average portion of time spent at the node by an infinite random walk. Node centrality is attracting an increasing attention by the scientific community, in particular during the latest years, such as predicting node degree centrality with the preferential attachment and triadic closure [18],application of degree centrality[19] and centrality-Newman for collaborative relationship distribution[20].

An attention also requires defining an importance measure (also referred to as centrality) to weight edges. At present, the edge ranking algorithm based on information flow is not rare. Fortunato et al. extended it to an edge and quantified the importance of an edge of a graph G [10]. The information centrality of the edge is defined as the relative drop in the network efficiency caused by the removal of the edge from G. Newman defined the betweenness of an edge as the number of shortest paths between pairs of vertices that run along it [11]. Meoa et al. proposed the concept of K-path edge centrality [12]. The index is based on the importance of the information dissemination ability to calculate the edge of the network.

PageRank algorithm was initiated ordering web-search results. Calculating PageRank is usually done using the power method which can be implemented very efficiently, even for very large systems [13]. PageRank is a method in which we can rank nodes in different link structures such as social networks in order of "importance" given the link structure of the complete system [14]. Opsahl et al. pointed out that PageRank algorithm took into consideration the global topological properties of the network. However, lack of some other practical factors [15].

We improve the PageRank algorithm, and present CentraRank algorithm including betweenness and closeness. The convergence of the CentraRank is proved to be true. The results of the performed experimentation keep the effective and feasibility of CentraRank algorithm. And the initial value can significantly increase the likelihood of convergence of the algorithm. An edge ranking algorithm (EdgeRank) based on node CentraRank is proposed. The arrangement is as follows: Some previous measurements on centrality are introduced in section 2; Section 3, we optimize the node centrality, and propose the properties of the algorithm in detail; In section 4, the effectiveness feasibility are proved by empirical data and simulation data; In section 5, EdgeRank algorithm derived by node centrality ranking is present; Finally, the conclusion and discussion is given.

2 Previous centrality measurements

Considering a network represented by a graph G= (V, E), where the V and E are sets of all the nodes and edges, respectively. Measures of centrality can be categorized into four main groups depending on the types of statistics on which they are based [16]. Degree - how connected a node is, closeness - how easily a node can reach other nodes, betweenness - how important a node is in terms of connecting other nodes, neighbors characteristics - how important, central, or influential a nodes neighbors are.

Perhaps the simplest measure of the position of a given node in a network is simply to keep track of its degree. The degree centrality of a node is simply $d_i\left(g\right)/(n-1)$ so that it ranges from 0 to 1 and tells us how well a node is connected, in terms of direct connections. Of course, degree centrality is clearly missing many of the interesting aspects of a network. In particular, it completely misses any aspect of how well located a node is in a network.

This second class of measures keeps track of how close a given node is to each other node. One obvious closeness-based measure is just the inverse of the average distance between given node and any other node [16].

$$m_i = \frac{(n-1)}{\sum_{k\neq j} P(k,j)} \tag{1}$$

Despite the closeness can measure how easily a node can reach other nodes, but the algorithm takes into account only the local node proximity.

Betweenness centrality is based on how well situated a node is in terms of the paths that it lies on. Averaging across all pairs of nodes, the betweenness centrality of a node is as follows [16]:

$$Ce_i = \sum_{k\neq j:i\notin(k,j)} \frac{\frac{P_i(kj)}{P(kj)}}{\frac{(n-1)(n-2)}{2}} \tag{2}$$

Betweenness of the node can reflect its position in the internet communication and it has a more significant effect for partial measure of the position of the nodes.

Extrapolating, it is easy to see that the distribution of degrees of a node found by choosing a link uniformly at random from a network that has degree distribution P and then picking either one of the end nodes with equal probability is as follows: $\widetilde{P}(d) = \frac{P(d)d}{\langle d\rangle}, where\,\langle d\rangle = \sum_d P(d)d$, which provides a good approximation of the degree of a neighbor.

PageRank is one of the traditional algorithm, its main mathematical model is as follows: $PR\left(A\right) = (1-d)+d\left(\sum\limits_{i=1}^{n}\frac{PR(T_i)}{C(T_i)}\right)$,Where $PR\left(A\right)$ Ithe PageRank - value of pages;$PR\left(T_i\right)$ - the PageRank value of pages which links to T_i;$C\left(T_i\right)$ - the out-degree of T_i;d - damping factor,$0<d<1$.It is typically set to 0.85. PageRank is important to incorporate both degree and neighbors when studying the centrality of a node.

3 Optimizing node centrality ranking

The approach based on the structure of the network provides two principal advances in our understanding of centrality: verifying the correctness and effectiveness of the algorithm.

3.1 The optimization of algorithm

Optimization problems carry two major operations of proposed algorithm in detail, including model optimization and selection of initial values.

3.1.1 The optimization of algorithm model

In this paper, the computer generated networks are selected for current cases to check the PageRank algorithm. Fig.1 and table 1 provides an example to illustrate how the PageRank works.

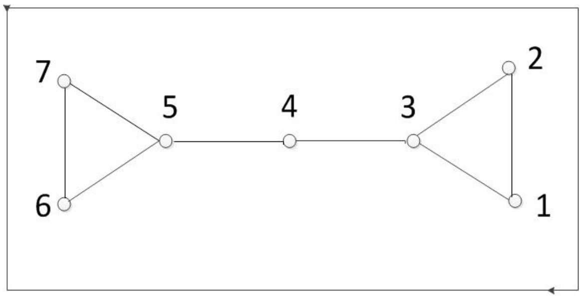

Figure 1: A network with 7 nodes (Figure is taken from [16]

Centrality	Nodes 1,2,6,7	Nodes 3 and 5	Node 4
Degree	0.33	0.5	0.33
Closeness	0.4	0.55	0.6
B-T	0	0.53	0.6
PR	0.33	0.5	0.33

Table 1: Centrality Comparisons for Fig.1 (Table is taken from [16])

In table1, B-T represents Betweenness,which is similar to the following table. First degree algorithm does not permit causal identification at the centrality level. In Fig.1 the degree of nodes 3 and 5 are third, and the degree of node 4 is only second. Arguably, node 4 is at least as central as nodes 3 and 5, and far more central than the other nodes that each has two links(nodes 1, 2, 6, and 7). There are several senses in which we see a powerful or central role of node 4. If one deletes node 4, the component structure of the network changes. This might be very important if we are thinking about something like information transmission, where node 4 is critical in path-connecting nodes 1 and 7. So there will be deviation when we measure the centrality with degree.

Second PageRank algorithm takes into account the degree and the degree of a neighbor, so the result is same with the degree centrality.

Finally, closeness and betweenness can measure how easily and how important a node can reach other. Table 1 show that PageRank algorithm is lack of these indicators, and we also can see that PageRank algorithm, closeness and betweenness are irrelevant, so that combining linearly these two indicators and PageRank algorithm can achieve complementary effects.

Next, we investigate why we make improvement for the PageRank algorithm. Essentially speaking, four centrality indicators can characterize each individual node by far clearly and discriminative among these measures in that different ranks have different scores. They only reflect local or global feature are the most important feature of the improvement. Comparing with the local indices, the global ones ask for the whole topological information. So we note that the combination of global and local measures to design a structure-based measure became necessary.

Therefore, this paper proposes a new algorithm model, it not only retains the advantages of the PageRank algorithm, also joins closeness and betweenness linearly. We can combine the advantages of different indicators by a simple linear addition. The new algorithm named CentraRank (also referred to as CR algorithm). Algorithm model is as follows:

$$CR_i = \mu \sum_{j=1}^{n} W_{ji} CR_j + (1-\mu)\, C_i \tag{3}$$

From equation (3), each parameter is as follows:

① CR_i represents centrality value of nodes, it is the final indicator to measure the centrality.

② $\mathbf{W_{ji}}$ is the adjacency matrix. The network adjacency matrix contains only 0 and 1. $w_{ji} = \begin{cases} 1,\ (v_j, v_i) is\ the\ edge\ of\ G; \\ 0, otherwise. \end{cases}$

③ μ is a tuning parameter that can set according to the research setting and data. If this parameter is between 0 and 0.5, then the closeness and betweenness are taken as favorable, whereas if it is set above 0.5, the former is favorable.

④ C_i is the mean of the closeness and betweenness values.

$$C_i = (m_i + Ce_i)/2 \tag{4}$$

m_i is the closeness value , which can be calculated from the formula (1); Ce_i is the betweenness value, which can be calculated from the equation (2).

3.1.2 the selection of initial value

Another important aspect to be elucidated is that, in general, the algorithm may depend not only on the model but also on the selection of initial value. Although the initial value does not affect the final result, but it can reduce the number of runs to some extent, and enhance the performance of the algorithm. Table 2 illustrates that our method is highly correlated with the In-degree algorithm, compared with the other methods such as the betweenness centrality, the closeness centrality and the degree method. Using the experiments on the Fig.2, we have found that the Indegree centrality significantly outperforms the other measures with initial value.

Correlation	B-T	I-D	O-D	D	C
PR	0.140	0.622	0.146	0.564	0.120

Table 2: The pearson correlation of measures

In table2,B-T represents Betweenness,O-D represents Out-Degree,I-D represents In-Degree,D represents Degree and C represents closeness, which are similar to the following table.

The complexity of data is a very important factor of the algorithm. PageRank algorithm takes the random number between 0 and 1 to improve the computational complexity and reduce the number of running. So the paper adopts the standardized degree as the initial value of CR algorithm and then proves the advantages of this algorithm.

3.2 The design of algorithm

Our algorithm is designed as follows:

①Import the adjacency matrix $\mathbf{P1}$ and then transfer the matrix $\mathbf{P1}$ to standardized matrix $\mathbf{P}$;

②Calculate the m_i and Ce_i of each node according to the equation (1) and (2)and then Calculate the C_i of each node according to the equation (4) ;

③Select the a standardized degree the initial value CR' and determine the value of μ;

④Calculate the CR_i of each node according to the equation (3);

⑤Determine the convergence of the result. If$\max\left(abs\left(CR' - CR\right)\right) > 0.0001$, output$CR$; otherwise repeat ④.

3.3 the convergence of the algorithm

Algorithm depends on the equation (3), so we just need to prove that the convergence of equation (3). The convergence of equation (3) is equivalent to the convergence of $CR_i = \mu \sum\limits_{j=1}^{n} W_{ji}CR_j$.

Proof: For each node, there is $CR_i = \mu \sum\limits_{j=1}^{n} W_{ji}CR_j$. Therefore, for all nodes on the network, we can know the formula $CR = \mu \cdot \mathbf{W} \cdot CR$.For a given matrix $\mathbf{W}$, CR is the corresponding eigenvectors. The convergence of equation (3) is equivalent to the uniqueness of eigenvectors.According to the Perron-Frobenius Theorem [16], if $\mathbf{W}$ is a nonnegative (column) stochastic matrix, so that the entries of each of its columns sum to 1, then there will exist a nonnegative right-hand eigenvector as solution with 1 being the corresponding eigen-value.The same is true of row stochastic matrices and left-hand eigenvectors.

If $\mu = 1$,there exist an unique nonnegative eigenvector CR, convergent of equation (3) hold on.

If $u \neq 1$,set $s = CR/\mu$, the original equation is converted to $s = \mathbf{W}s$, so the algorithm is convergent.

The convergence of equation (3) is proved by the above process.

4 Analysis to node centrality algorithm

4.1 Datasets and settings

We employ two datasets of real networks. The first one is the Sina microblogging network. We collected data in December 2014. Sina microblogging is constructed by the following relationship and visualized by Pajek [21](see Fig.2).The nodes in the network represent the users, if node 1 follows node 2, so the connection is formed with .There are 253 nodes and 510 edges in the network. We can conclude from the Fig.2 that the interaction between the nodes is relatively small, and this is consistent with the feature of Sina. The degree of most nodes is relatively small. It is in line with the actual network. So it is suitable for the algorithm analysis. Second one is the Les Misrables relationship network(see Fig.3). D. E. Knuth finished the relationship between the characters of the network according to the novel Les Mis [17]. The nodes represent the characters in the story and the edges represent two roles in the same scene or act in. The network is also a directed network which has 77 nodes and 508 links.

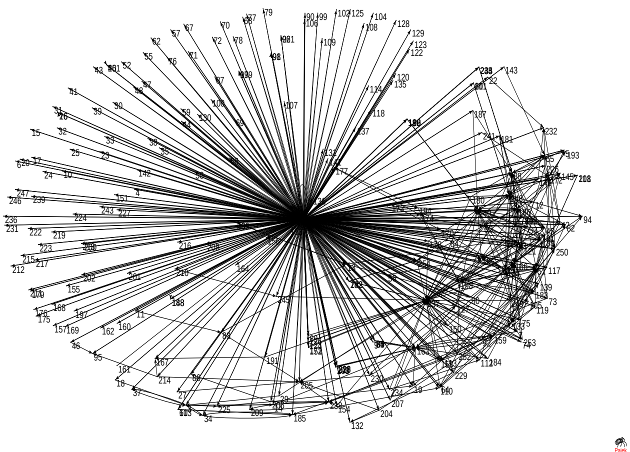

Figure 2: Visualization of Sina microblogging

4.2 The value of μ

It is a difficulty to determine the tuning parameter μ both for PR or CR algorithm. In order to illustrate the effect presented earlier for relationship with different μ, consider the error formed by fitting method,

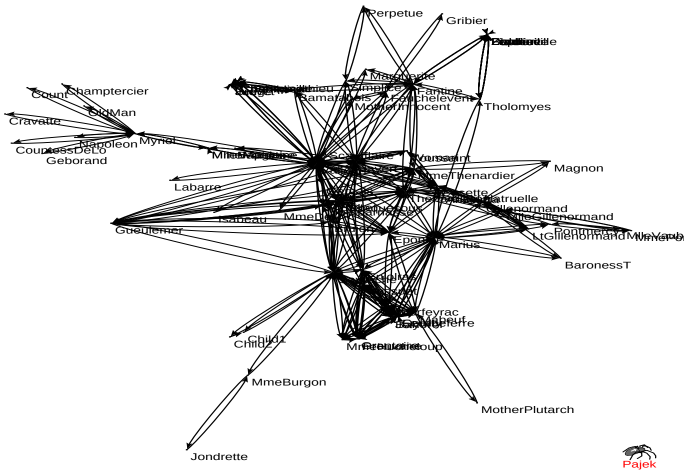

Figure 3: Visualization of Les Misrables

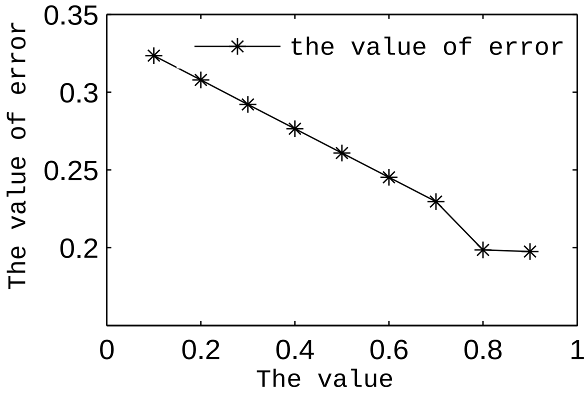

Figure 5: Error of Les Misrables

illustrated in fig.4 and fig.5. We observe the errors from sina microblogging and Les Misrables.

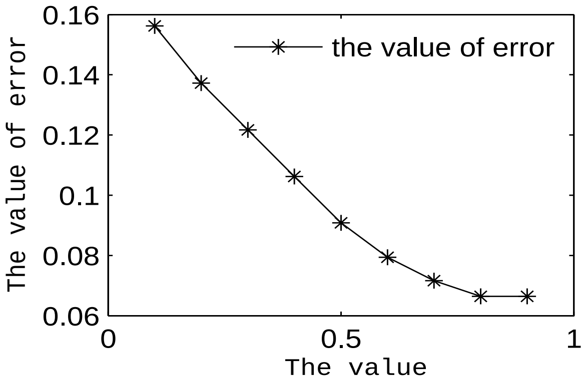

Figure 4: Error of Sina microblogging

The error in this paper is $(x_i - x_{i-1})^2$, where x_{i-1} and x_i represent respectively two adjacent ranking results of CR algorithm. Since the results of CR algorithm are not unique, this section takes the total error by many times of running. We set $\mu = \{0.1, 0.2...0.9\}$ and simulate in matlab 100 times.Fig. 3 tells us the value of error can determine the best μ.when μ=0.8 and μ=0.9,error is gradually reduced. In addition, most μ maintain relatively stable value across the diverse cases. So setting μ=0.85 as the tuning parameter is very close to PR algorithm.

4.3 Comparisons of centrality algorithms

The objective for this developed algorithm aims at verifying the validity of algorithm using the top 8 nodes from Fig.2 and three other centrality algorithms (see Table 3). In table 3, A-D represents All-

Number	A-D	I-D	B-T	PR	CR
237	2	1	6	1	1
240	4	3	2	2	2
163	5	4	4	4	3
166	3	2	43	3	4
186	8	10	1	12	5
134	9	93	3	8	6
51	7	5	8	6	7
248	6	7	7	7	8

Table 3: the ranking result of five centrality measures

Degree,I-D represents In-Degree,B-T represents Betweenness,which is similar to the following table.

By combining the PR algorithm with the betweenness algorithm, CR algorithm has better complexity than the PR algorithm. And results in CR and PR algorithm of number 163 and number 166 are opposite: number 163 is ranked No. 3 in the CR algorithm, but ranked No. 4 in the PR algorithm, while the ranking result of number 166 is opposite. Betweenness is very likely to stay in CR algorithm, which results in the change of number 163. The same situation is also appeared in the number 51 and 248. We can see the accuracy of results from the above conclusion.

4.4 The accuracy of algorithm

A well established method to measure the precision is the error method. So we compare the precision of five different centrality algorithms. The formula of error in this section is $(x_i - x_{i-1})^2$, where x_{i-1} and x_i represent respectively two adjacent ranking results of CR algorithm. In Fig.6, the abscissa represents

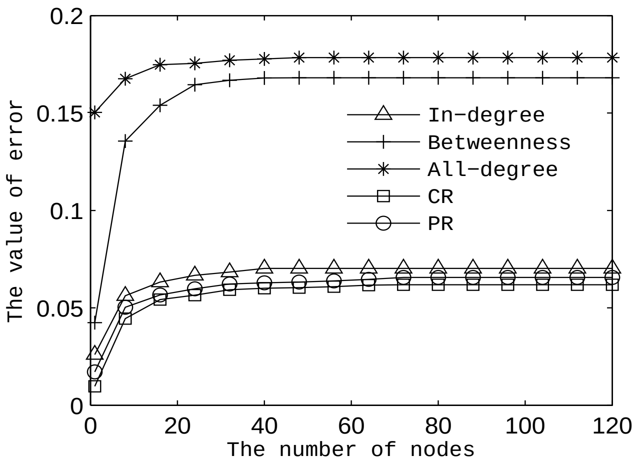

Figure 6: Error of five different centrality algorithms

the number of nodes. The ordinate represents value of the errors, the smaller of the errors, the better of the results. It is easy to find the CR curve is the smallest in the five algorithms, and PR algorithm is the second smallest, In-degree is the third.

We get the following conclusions By Fig.6:

(1) Accuracy of the algorithm is higher than the other four centrality ranking algorithm (error is smaller than other algorithm), and the advantage of the algorithm is quite obvious when the errors become stability. The maximum of average error is 0.0618 in CR algorithm, which is far less than the PR algorithm. (2)When it arrives at the top 10 nodes, the error changes slightly and when it arrives at the top 70 nodes, the error becomes stability. This is because the relationship in Sina microblogging is weak, so the two-way information exchange is relatively small, which means that there is small number of active users. In actual social networks, the number of users at the important position is very small, and most of the nodes are in a state of non-interaction. The experimental results agree with current situation.

4.5 Complexity of the algorithm

Seven datasets of networks are employed by this section. The top 6 is the randomly generated networks and the seventh is the network from Fig.2. The simulation is run by MATLAB7.0.

In Fig.7, the length of the matrix is denoted by n,$n = \{4, 8, 16, 32, 64, 128, 254\}$ represents the length of each adjacency matrix of the seven data used in this section. In Fig.7, it is easy to observe the following conclusions :

(1) The size of matrix does not matter with the number of runs. With the increase of size, the num-

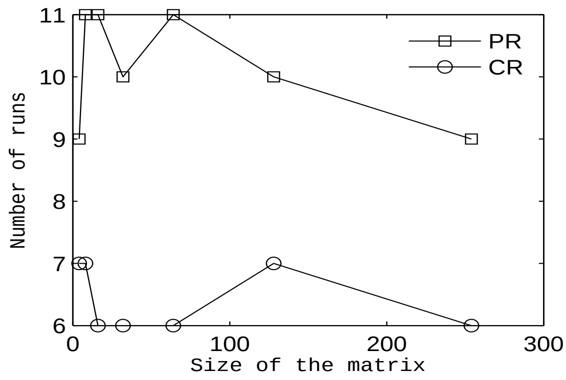

Figure 7: The number of runs in different sizes of matrix of PR and CR algorithms

ber of runs reduces firstly, and then increases, which indicates that two parameters are not correlated.

(2) With same size of networks, CR algorithm runs much faster than PR algorithm. For example, the random network with $n = 32$, the average number of runs is 10 while CR algorithm is 6. The results show the feasibility of CR algorithm. And we also find the number of runs of the algorithm does not increase with the size of networks.

4.6 The running time of algorithm

An important property of algorithms is running time against other centrality algorithms.Consequently, it is now vitally important to evaluate the running time of such networks in terms of the services supported by the network in question.We compared the CR algorithm and the PR algorithm with the same network datasets (data is taken from 4.5). It is as shown in the following Fig.8.

In Fig.8,the size of the network is denoted by n and the size of the networks matrix is also denoted by n.By Fig.8, when n=4,8,16,32,64,128,254, some conclusions are obtained:The running time of algorithms are positive proportion to n. The two algorithms running times are coincidence completely.One reason is algorithm time CR_i in equation (3) is equal to the time of PR_i if C_i is given before, if is given before, another reason is the size of the networks are relatively small.So the algorithms times are same.

(2) When graphs with millions of nodes, the algorithm time of C_i should be consider into CR_i ,because the algorithm time of C_i is $O(C_i) \approx n^3$, while $O(PR_i) < n^3$. Therefore, the size of datasets is large enough, $O(CR_i)$ is much larger than $O(PR_i)$.

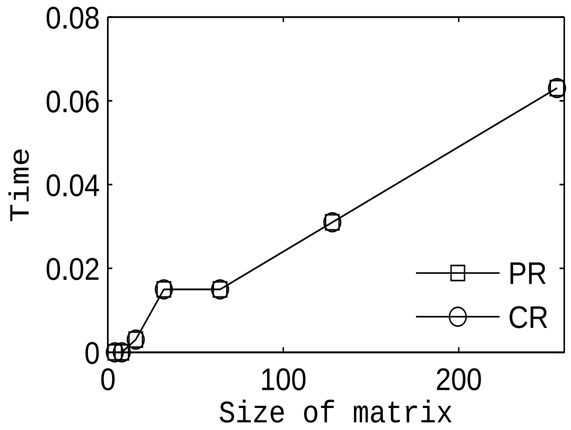

Figure 8: The time in different sizes of matrix of PR and CR algorithms

5 EdgeRank algorithm derived by node centrality ranking

From the perspective of the network structure, an edge centrality algorithm based on node centrality is proposed. The new edge centrality algorithm converts edges to nodes, called EdgeRank algorithm.

5.1 The EdgeRank algorithm

EdgeRank algorithm is generated as following:

①The edges in the network are converted to nodes, then we use the CR algorithm to solve the problem of edge ranking, the converted method is as follows: Fig.9(a) represents the conversion of directed

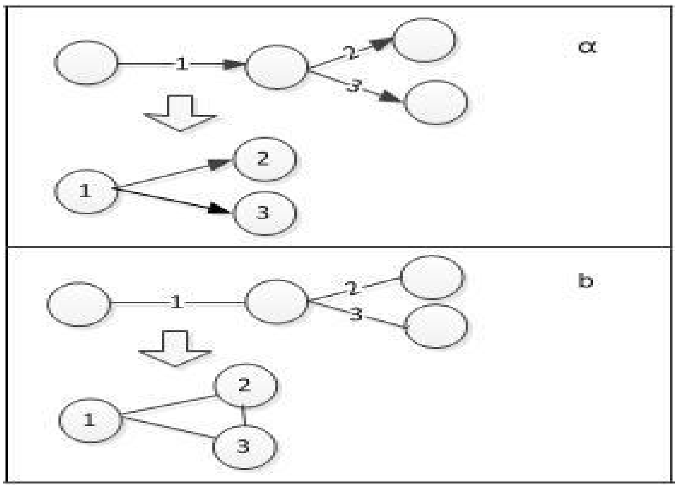

Figure 9: The conversion model of directed and undirected network.

network. The edges 1,2 and 3 in the first part of Fig.9(a) is converted to the nodes 1,2 and 3 in the last part,before and after the conversion are directed network. Fig.9(b) represents the conversion of undirected network. The edges 1,2 and 3 in the first part of Fig.9(b) transformed to the nodes 1,2 and 3 in the last part, the conversion hold the edges undirected in networks.

To demonstrate the equivalence of the network before and after conversion, we use the degree and betweenness of the networks to illustrate the equivalence. The degree of edge 1 in Fig.7 is 2, and the edge 2 and edge 3 is parallel in Fig.9(a). We can reach the edge 2 and edge 3 through the edge 1, the same with the conversion. The edges in the undirected network are fully connected and they are completely network after conversion.

It can be seen from the above that this transformation not only ensure the equivalence, but also ensure that the positions of all edges in the network remains unchanged. And the model can be simplified after transformation, which can calculate various properties of edges.

②Use the Centrarank algorithm to get the results of edge centrality ranking.

5.2 Efficiency of the EdgeRank algorithm

5.2.1 The edge centrality of undirected network.

Michaels strike network was an excellent example to test edge centrality [17], the social network of a forest product manufacturing factory contained 24 nodes and 38 edges as described by Judd H.Michael in 1997. The two union negotiators (Sam and Wendle) were responsible for explaining the changes, but they failed to do so, and a strike broke out. Bob and Norm C who were at the overlap of the three communities of the factory sociogram, convince them about changes. By following this strategy, the management solved the problem soon, and the strike ended. The simulation of NetworGame of the choice of Bob and Norm showed that they could convince everybody to stop the strike in 100 % of the simulations. Simulation of the choice of Sam and Wendle led to the poor result of convincing the others to stop the strike in 8 % of the simulations with the same settings, which corresponds well with outcome of the real-world events.

The abscissa in Fig.10 represents the number of edges after the conversion; the ordinate represents the results of ER algorithm.We can see in Fig.10, the ranking result of edge 17 is 1, which corresponds exactly to the edge Bob-Norm. And the ranking result of edge 37 is 37, which corresponds exactly to the edge Sam-Wendle. The results are not only consistent with NetworGame, but also fully in line with the actual sit-

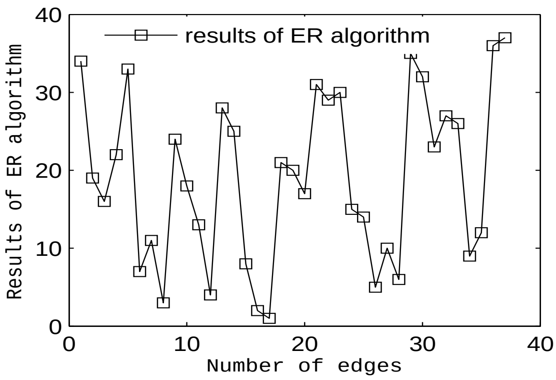

Figure 10: The edge centrality of undirected network.

uation, which can be able to explain the correctness of the algorithm.

5.2.2 Accuracy analysis of the directed network

The analysis is based on Les Misrables data in this section, 77 nodes and 508 edges in it. The top 6 results are summarized in Table 6 before and after transforming.

Ranking	Edge	Node	User name
1	36	12-49	Valjean-Gavroche
2	39	12-56	Valjean- Marius
3	23	12-28	Valjean- Javert
4	19	12-24	Valjean- Fantine
5	21	12-26	Valjean- Thenardier
6	40	12-59	Valjean- Enjolras

Table 4: The top 6 results of the Les Misrables

In Table 4, it is notable that the 12-49 (the top ranked) and the 12-56 (the second ranked) in the ER algorithm are also key persons of CR algorithm: Valjean is the top 1, Gavroche is the top 3 and Marius is top 4 in the CR algorithm. They are the main characters in the movie. Again, we observe the same tendencies for the top 6.

In order to better illustrate the pros and cons performance of the algorithm and explain the actual meaning of the edge centrality, we adopt the robust to test the edge centrality. We define a new indicator to compare the importance of edge and nodes. Denote total number of nodes by $|V|$, and the number of nodes in largest connected component by $|C|$, then

$$Damage \quad rate \ = 1 - \frac{|C|}{|V|} \tag{5}$$

Equation (5) shows that the damage rate of the network is proportional to the centrality of edge and node. The higher of damage rate indicates that the edge or node in the network is more important.

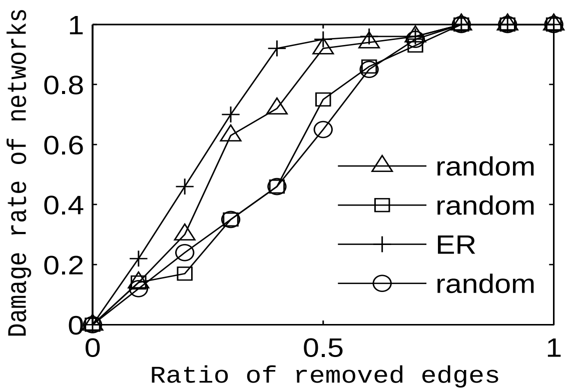

Figure 11: The damage rate of edges in directed network under attack

The abscissa in Fig.11 represents the ratio of removed edges; the ordinate represents the damage rate; ER algorithm represents to attack the ranking number of edges in ER algorithm, while the random represent random attack. The simulation is under Windows XP system environment and MATLAB7.0 software.

In Fig.11, in order to reduce the probability of randomness, we select three datasets from the top 10. The damage rate of ER algorithm is far greater than the random data. The random remains the second best, but it gets closer to ER as the abscissa becomes bigger. Because the network contains 254 edges, so the damage rate of the network achieves convergence in 0.8. The notable difference is that the flat region increases for ER and random as for the case of $x = 0.4$ .We can conclude that ER algorithm can characterize each individual edge by far clearly, which indicates the accuracy of the ER algorithm.

5.3 Comparison between edges and nodes

We select the Les Misrables and the top 8 nodes in the Fig.2 as the data in this section, which contains 8 nodes and 13 edges. The ranking results before and after transformations are as follows:

The abscissa in Fig.12 represents the number of edges after conversion; the ordinate represents the results of ER algorithm. As can be seen from Fig.12, the ranking result of edge 7 is 1 and the ranking result of edge 6 is 13. Next, we investigate the relationship among the edges and nodes under attack.

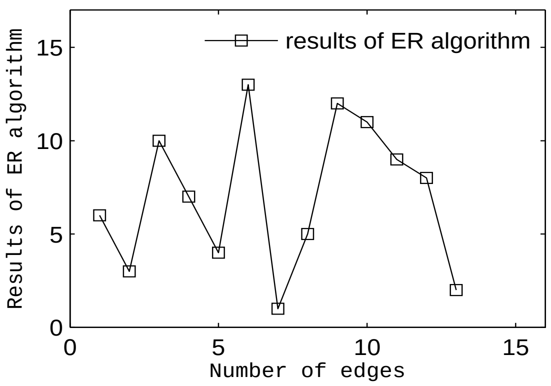

Figure 12: The edge centrality of the top 8 nodes in the fig.2 after transformation

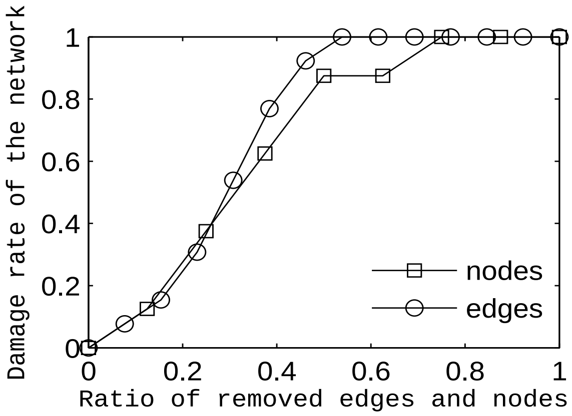

Figure 13: The damage rate of nodes and edges in the top 8 nodes after transformation

The abscissa in Fig.13 and Fig.14 represents the ratio of edges and nodes; the ordinate represents the damage rate; the simulation is under Windows XP system environment and MATLAB7.0 software.

As it can be seen from Fig.13 and Fig.14,with the increase of ratio of edges and nodes, the damage rate is decrease, which can indicate the accuracy of ER algorithm. We observe the different tendencies for the two networks. The attacking of edges can increase the damage rate of the network in Fig.13, while in Fig.14 is on the contrary. This is intuitively understandable. When the Les Misrables is considered, there are many important nodes, in which case the nodes play a key role; when the top 8 nodes in the Fig.2 is considered, there are many randomly linked nodes, in which case the edges play a key role. So we judge the importance of edges and nodes through the concrete cases.

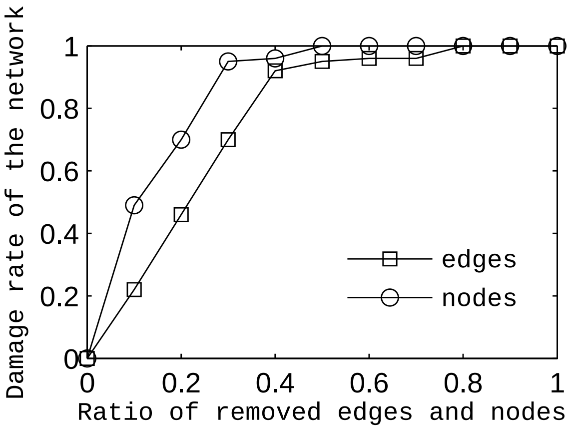

Figure 14: The damage rate of nodes and edges in Les Misrables after transformation

6 Conclusion and discussion

CentraRank algorithm is to identify and characterize the influential nodes in a social network based on the structure of the network, it measures the centrality of nodes. Empirical and numerical simulations prove the feasibility of the algorithm. CentraRank algorithm is superior to previous algorithm comparing the errors and complexities; based on node centrality, an edge centrality algorithm is proposed, and the correctness of the ER algorithm is discussed.

In CR algorithm, the structure of the node or the edge is considered, it is the improving the previous ranking algorithm. However, there are still a few defections in our CR algorithm. For example, it omits the social meanings of the nodes, and the conversion arouses the increasing of adjacency matrix. Those problems are deserved to further research.

Acknowledgements: This work is supported by Natural Science Foundation of China (grant No.71471106) and Specialized Research Fund for the Doctoral Program of Higher Education(grant No.20133704110003).

References:

[1] R. ZHUO MING,S. FENG,etal,Node importance measurement based on the degree and clustering coefficient information,*Phys.* 12,2013,pp. 12-18

[2] LC. FREEMAN,A set of measures of centrality based on betweeness,*Sociometry.* 40,1977,pp. 35-41

[3] P. BONACICH,Factoring and weighting approaches to status scores and clique identification,*Journal of Mathematical Sociology.* 2,1972,pp. 113-120

[4] P. BONACICH,Power and centrality:a family of measures,*American Journal of Sociology.* 92,1987,pp. 1170-1182

[5] JACKSON,A.M WOLINSKY,A Strategic Model of Social and Economic Networks,*Journal of Economic Theory.* 71,1996,pp. 44-74

[6] S. CARMI,S. HAVLIN,S. KIRKPATRICK,A model of Internet topology using k-shell decomposition,*Proc. Natl. Acad. Sci.USA.*,2007,pp. 11150

[7] Li Ying,etal,Dynamical Influence of Nodes Revisited: A Markov Chain Analysis of Epidemic Process on Networks,*Chinese Phys. Lett.* 29,2012

[8] ROM ERO DANIEL MO,GALUBA WOJCIECH,etal,Dynamical Influence of Nodes Revisited: A Markov Chain Analysis of Epidemic Process on Networks,*New York: [s.n.].*2011pp. 113-114

[9] S. FORTUNATO,V. LATORA,M. MARCHIORI,Method to find community structures based on information centrality,*Phys. Rev. E.* 70,2004

[10] M. NEWMAN,M. GIRVAN,Finding and evaluating community structure in networks,*Phys. Rev. E.* 69,2003

[11] D.P MEOA,E. FERRARAB,A novel measure of edge centrality in social networks,*Knowledge-Based Systems.* 30,2012,pp. 136-150

[12] Y.I SOON,N.S KOH,Speech enhancement using 2-D fourier trans-form,*IEEE Transactions on Speech and Audio Processing.* 11,2003,pp. 717-724

[13] L. Linyuan,etal,Leaders in Social Networks, the Delicious Case,*PLoS ONE.* 6,2011,pp. 1-9

[14] T. Opsahl,F. Agneessens,J. Skvoretz,Node centrality in weighted networks: Generalizing degree and shortest paths,*Social Networks* 32,2010,pp. 245-251

[15] O.M Jackson,Social and Economic Networks,*Princeton University Press.*2008

[16] E.D Knuth,Social and Economic Networks,*The stanford graph base: a platform for combinatorial computing.*Addison-wesley, Reading, Ma, 1993

[17] T. Opsahl,F. Agneessens,J. Skvoretz,Node centrality in weighted networks: Generalizing degree and shortest paths,*Social Networks* 32,2010,pp. 245-251

[18] Yang Yang, Yuxiao Dong,V. Nitesh,Predicting Node Degree Centrality with the Node Prominence Profile,*Nature Scientific Reports* 4,2014

[19] Guijie Zhang, Luning Liu, etal,Cext-N index:a network node centrality measure for collaborative relationship distribution,*Scientometrics* 101,2014,pp. 291-307

[20] Sinha, Anupam,Nagarajaram, etal,Effect of alternative splicing on the degree centrality ofnodes in protein-protein interaction networks of Homo sapiens,*Journal of proteome research* 12,2013

[21] Wouter de Nooy,Andrej Mrvar,Vladimir Batagelj,Exploratory Social Network Analysis With Pajek,*Poetics and Social Networks*,2010

Responds to the reviewers' comments:

Dear editors and reviewers:

Thank you for your letter and for the reviewers comments concerning our paper entitled 'The centrality ranking algorithm based on the network structure' (ID:5809-743).These comments are all valuable and very helpful for revising and improving our paper. So we have substantially revised our paper after reading the comments provided by the two reviewers.

Responds to the reviewers' comments:

—-REVIEWER1 —-

Comments: It will be much more better if the title can be shorter.

Revision: We should like to thank the reviewers for their helpful comments, but we think that the title may ignore key content if shorter.

Special thanks to you for your good comments.

—-REVIEWER 2 —-

1.The paper should be corrected by a native speaker.

Revision:

We greatly appreciate both your help and that of the reviewers concerning improvement to this paper. We do with a grammar check from a native English speaker.

2.Comments about reference

(1)Most references are before 2012. The author should also add newer references.

Revision:

Three further papers have been added to the text and Reference section. Thank you very much for the comments of the reviewers, we joined the content in section 1 paragraph 3 line 10.

Node centrality is attracting an increasing attention by the scientific community, in particular during the latest years, such as predicting node degree centrality with the preferential attachment and triadic closure [18], application of degree centrality [19] and centrality-Newman for collaborative relationship distribution [20].

References:

[18]Yang Yang, Yuxiao Dong,Nitesh V, Predicting Node Degree Centrality with the Node Prominence Profile,*Nature Scientific Reports*, Vol.4, 2014.

[19]Guijie Zhang, Luning Liu, etal,Cext-N index:a network node centrality measure for collaborative relationship distribution,*Scientometrics*, 101,2014,pp. 291-307

[20]Sinha, Anupam,Nagarajaram, etal,Effect of alternative splicing on the degree centrality ofnodes in protein-protein interaction networks of Homo sapiens,*Journal of proteome research*, 12,2013

(2)Equations (1) and (2) need references.

Revision:

we join the references in section 2, paragraph 3 (line 5).

One obvious closeness-based measure is just the inverse of the average distance between given node and any other node[16].

we join the references in section 2, paragraph 4 (line 4).

Averaging across all pairs of nodes, the betweenness centrality of a node is as follows [16]:

[16]Jackson M O, Social and Economic Networks, *Princeton University Press*, 2008

(3) Pajek: a reference is needed.

Revision:

we join the references in section 4, paragraph 1 (line 5).

[21]Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj,Exploratory Social Network Analysis With Pajek,*Poetics and Social Networks*,2010

3.Comments about time

(1) How much time does it take to run the algorithm on the used datasets ? What about PageRank ?

(2) What about graphs with millions of nodes ? What happens with execution times ?

Revision:

As suggested by reviewers, a discussion of the time of algorithm on the used datasets (data is taken

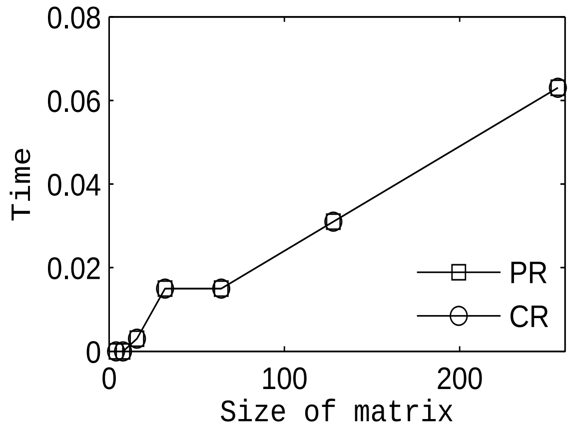

Figure 15: The time in different sizes of matrix of PR and CR algorithms

In Fig.15,the size of the network is denoted by n and the size of the networks matrix is also denoted by n.By Fig.15, when n=4,8,16,32,64,128,254, some conclusions are obtained:The running time of algorithms are positive proportion to n. The two algorithms running times are coincidence completely.One reason is algorithm time CR_i in equation (3) is equal to the time of PR_i if C_i is given before, if is given before, another reason is the size of the networks are relatively small.So the algorithms times are same.

(2) When graphs with millions of nodes, the algorithm time of C_i should be consider into CR_i ,because the algorithm time of C_i is $O(C_i) \approx n^3$, while $O(PR_i) < n^3$. Therefore, the size of datasets is large enough, $O(CR_i)$ is much larger than $O(PR_i)$.

4.Comments

-Fig.1 The network with 7 nodes, the data from the literature [16] = Fig.1 A network with 7 nodes (Figure is taken from [16])

Revision:

We greatly appreciate you concerning improvement to this paper. We have made correction according to the comments. We also have re-written table1 according to the suggestion.

I greatly appreciate both your help and that of the referees concerning improvement to this paper. = Table1 Centrality Comparisons for Fig.1 (Table is taken from [16])

5.Comments about format

(1) The paper should be sent in pdf format.

Revision:

We are very sorry for our negligence of the format. The paper is sent in PDF format.

(2)There are some empty spaces in the paper. They should be erased.

Revision:

Considering the reviewers suggestion, we have made correction. Some empty spaces are erased in PDF format.

(3)Most Figures are of low quality and cannot be seen very well. Resolution and quality should increase.

Revision:

We are very sorry for our negligence of the figures. We redraw some figures to increase the resolution and quality. Special thanks to you for your good comments.

We try our best to improve the paper and make some changes in the paper. I hope that the revised paper is now suitable for publication.