Team Control Number

28199

Problem Chosen

C

2014 Mathematical Contest in Modeling (MCM) Summary Sheet

Problem Clarification

In this paper, we build several networks, study their properties, and build models to measure players' influence. Off-the-shelf measures like degree are not optimal for influence estimation, because they do not consider all aspects of the network. We seek to define more accurate measures based on previous research for various networks and explore their utility in real life.

Model Design

Conventional measures such as centrality usually neglect some aspects of networks, for example, the weights of edges. Stephen P. Borgatti (2006) introduces a cohesion measure called KPP-NEG, but the order of nodes is still ignored, which causes inaccuracy.

We combine the Shapley approach with the cohesion measure to obtain the Shapley value of each node, which represents its contribution to the whole network's cohesion. It involves weight, structure, and order of nodes at the same time, which makes it superior to both conventional measures and KPP-NEG. However, this new approach can't be applied to directed networks.

Therefore, we use Pagerank algorithm derived from website networks to deal with directed networks like the citation network. But the basic method is not suitable for citation networks due to the differences between websites and papers and the limited network size. So we optimize it by adding a virtual node representing papers outside the network in two steps. The first is to adjust the weight of edges according to references. The second involves Times Cited of the target paper.
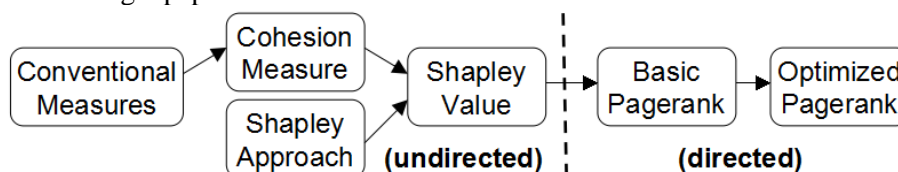

Figure 1. Model Design Process

Results

Erdos network is a small-world network according to its path length and clustering coefficient. Meanwhile, its degree distribution shows a feature of free-scale network.

For undirected networks, the Shapley value of Edros Network shows that the most influential co-author is ALON, NOGA M., which is different from that ranked by betweenness. The Shapley value implemented in the movie star network shows that the most influential star is Matt Damon, which is the same as that ranked by betweenness due to the relatively small size of the network.

For directed networks -- the citation network of 16 papers provided, the most influential paper remains the same whether in the extensive network or not, which is No.14. But the ranks of some papers change with different versions of Pagerank method.

Sensitivity analysis suggests that ranks based on Shapley value is sensitive to sample size, while ranks based on Pagerank is not sensitive to damping factor.

Influence Measures in Networks

Introduction

Problem Clarification

The study of complex networks is a young area of scientific research inspired largely by the empirical study of real-world networks such as computer networks and social networks.   With the development of society and technology, more and more networks spring up in various aspects of our life. Properties of networks have been widely explored, and models are needed to measure the influence and impact of players in different networks. However, few researches considerate weight, structure and order of notes at the same time, so we seek to define more accurate measures based on previous research for networks with and without directions or weight.

Model Design

The most typical conventional measurement is centrality, which includes degree, betweenness and closeness. However, both degree and closeness neglect the weight of each connection. In addition, Stephen P. Borgatti (2006) argues that the optimality of betweenness in identifying the node whose removal from the network would most reduce cohesion is not guaranteed under typical definitions of cohesion, and introduces a cohesion measurement called KPP-POS to overcome the defects of conventional approaches, but the order of nodes are ignored which causes inaccuracy.

We combine the Shapley approach and the KPP-POS measurement to obtain the Shapley value of each node as a influence measure, which represents its contribution to the whole network's cohesion. It considers the weight, the KPP-POS problem and the order of nodes at the same time, which makes it superior to both conventional measures and KPP-POS.

This new approach has many merits, but it can't be applied to directed networks. Therefore, we use Pagerank algorithm to deal with directed networks such as the citation network. The conventional Pagerank method is applied to website networks, but there are differences between website networks and paper citation networks. For instance, once built, the citation link or network is static while the website links are dynamically adjustable. Besides, the citation data related to papers outside the network is not taken into consideration, which makes the Pagerank value

meaningless in small networks. So we optimized this method by adding a virtual node which represents papers outside the network in two steps. The first step is to adjust the weight allocated to the cited paper according to the total number of its references. In the second step, different weight is given to inside papers by the virtual node based on their Times Cited since publication.

Erdos Network

Assumptions

- Co-author networks are undirected. The relationships between pairs of node are not directed, that is, the two nodes linked by the same line are equal.

- The influence of the number of joint publications and the date of the first cooperation can be neglected. Due to the lack of relative information, numbers of joint publications of different pairs are assumed to be the same. Therefore, our Erdos network model is unweighted, but this assumption is released in our movie-star network model.

- Take no consideration of those with Edros number 2. In order to avoid too much work, authors whose Erdős numbers are greater than 1 are not involved in this network.

Adjacent Matrix

The adjacent matrix of the co-author network can be figured out with MATLAB, in which $r_{ij} = 1$ if i can reach j and $r_{ij} = 0$ otherwise. W represents the adjacent matrix and N is the number of nodes in the network.

$$W = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & \cdots & r_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ r_{N1} & r_{N2} & \cdots & r_{NN} \end{bmatrix} \qquad (1)$$

Network Visualization

Given the adjacent matrix, we can visualize the network with PAJEK, which is represented in **Figure 2**. Visualized network can show relations among clusters (global view) and extract nodes that belong to the same clusters (detailed local view). The nodes in the center part of the network are closed to each other while there are 37 isolates at the edge. The graph shows that the node representing ALON, NOGA M. has a degree of 52, which is the highest degree in Erdos network.
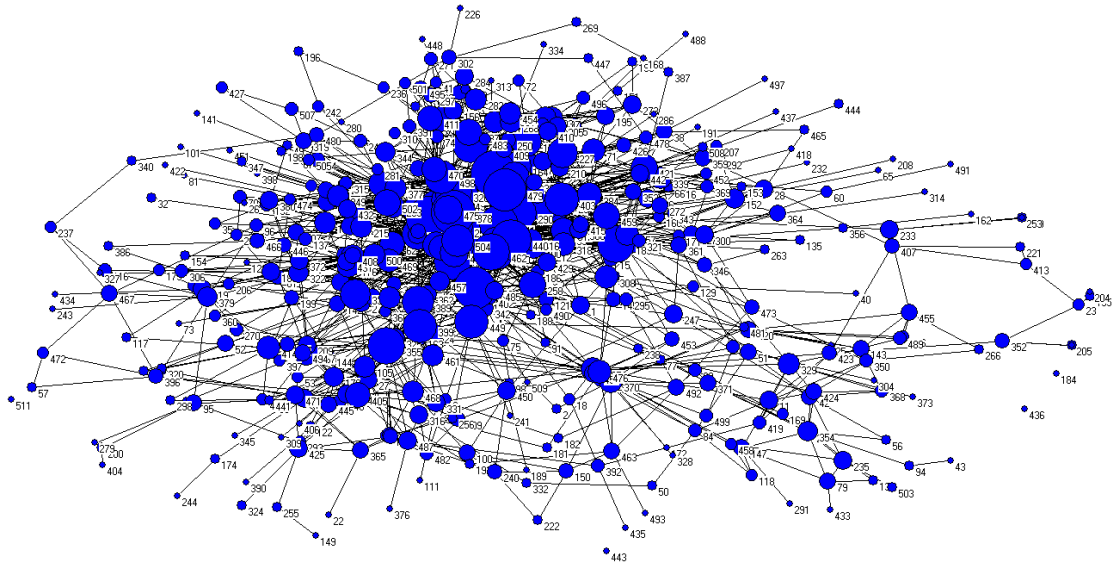
Figure 2. Node size corresponds to the degree of each author, a simple measure of influence. Numbers beside the nodes represent the corresponding author.

Properties of Network

Complex networks include small-world networks and free-scale networks. Small-world networks are between regular networks and random networks. To determine which type Erdos network belongs to, we analyze properties of it with three measures [Albert,R. and Barabasi,A-L.,2002]: degree distribution, average path length and clustering coefficient.

Firstly, degree distribution is the probability distribution of degrees over the whole network. To demonstrate better the feature of the degree distribution of this network, we plot the degree frequencies in **Figure 3**. A random network with same number of nodes is also analyzed as a comparison.
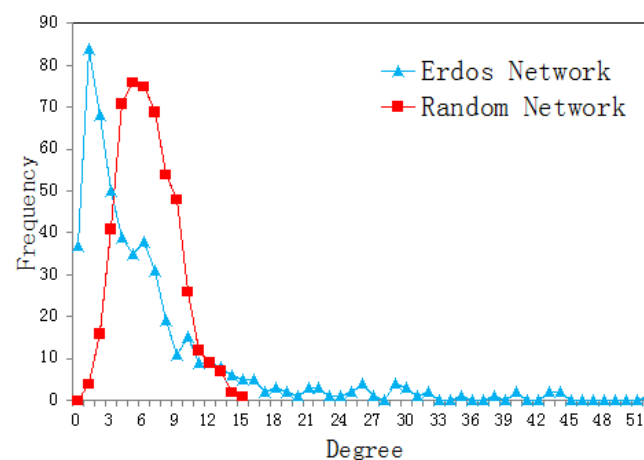

Figure 3. Degree Distribution of Erdos Network & Random Network with Same Number of Nodes

As is shown in **Figure 3**, the random network has a pronounced peak at average degree 6.41, while the degrees of Erdos network follow the power-law distribution, showing a feature of scale-free networks.

Secondly, the average path length of Erdos network is 3.82, which is defined as the average number of steps along the shortest paths for all possible pairs of network nodes, measuring the efficiency of information or mass transport on the network. The result shows that the average path length of Erdos Network is very low.

Thirdly, the Watts-Strogatz clustering coefficient is 0.3428, so the nodes in Erdos network tend to cluster together in a relatively deep extent.

The two preceding measures are calculated as follows (N is the number of nodes, E is the number of edges, an edge e_{ij} connects node v_i with node v_j):

$$l_G = \frac{1}{N(N-1)} \sum_{i \neq j} r_{ij} \tag{2}$$

$$CC_i = \frac{2 \left| \{ e_{jk} : v_j, v_k \in N_i, e_{jk} \in E \} \right|}{k_i (k_i - 1)} \tag{3}$$

$$\overline{CC} = \frac{1}{N} \sum_{i=1}^{n} CC_i \tag{4}$$

In order to describe properties more specifically, we use PAJEK to generate a random network with same number of nodes, which is compared with Erdos network in **Table 1.** This random network has the same number of nodes and average degree with Erdos network.

Table 1. Empirical results of networks

L_{actual}	L_{random}	CC_{actual}	CC_{random}
3.82	3.64	0.3428	0.0153

According to the above data, we can see that Erdos network shows the small-world phenomenon: $L \approx L_{random}$ but $CC >> CC_{random}$. Therefore, Erdos network is a small-world network.

The properties of Erdos network are summarized in **Table 2**. In **Table 2** the diagonal elements describe the small-world Erdos network and it is significant to note that Erdos network shares some features with text book network prototypes, but also differs from them. Its clustering coefficient is high while average length path is low, so

the structure is like a small world network, in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of steps. Therefore, even a change of one shortcut can make a great difference in Erdos network. However, since the degrees distribution follows a power law, it also has the property of free-scale networks that a few nodes have many neighbors and they are followed by nodes with smaller degree, which means that there are a few active co-authors who have cooperated with many other co-authors in this network.

Table 2. Properties of network: diagonal elements characterize small-world network

Properties / Networks	Clustering Coefficient	Average Path Length	Degree Distribution
Regular	*High*	High	Equal and fixed In/Out degrees to each node
Random	Low	*Low*	Exponential
Scale Free	Low	Variable	*Power-law distribution*

Influence Measure for undirected network: Shapley value

We calculate each node's contribution to the cohesion of the whole network through Shapley approach as the measurement of its influence. And the results are compared with betweeness, which is the most commonly used traditional measurement.

Conventional Approaches

Centrality is traditionally used to measure the impact of players in a network, which has three aspects: degree, closeness, and betweenness [Freeman, 1979]. Degree is defined as the number of edges incident to a node, which is the simplest concept of centrality. Closeness measures a node's path length to all other nodes. Both of degree and closeness ignore the weight of edges, and can't be used for

weighted networks. Betweenness centrality is the most commonly used among the three, which quantifies the number of times a node acts as a bridge. However, the optimality of betweenness in identifying the node whose removal from the network would most reduce cohesion is not guaranteed under typical definitions of cohesion. [Stephen P. Borgatti, 2006]

Shapley Value

Because of the defects of conventional approaches in measuring a node's contribution to cohesion of the whole network, Stephen P. Borgatti (2006) introduced a measurement of network's cohesion. We firstly adjust it to make it applicable for both weighted and unweighted network. In addition, we combine it with Shapley approach to obtain the Shapley value of each node.

The measurement of a network's cohesion is shown as follow:

$$\text{DF} = 1 - \frac{2\sum_{i>j}\frac{1}{d_{ij}}}{n(n-1)} \tag{5}$$

Where n denotes the number of nodes in the network, and d_{ij} represents the distance between node i and j. If there is no link between node i and j, d_{ij} is infinite. In order to take the weight of edges into consideration, we adjust the measurement by defining d_{ij} as:

$$d_{ij} = \begin{cases} 1 + \dfrac{1}{p_{ij}}, & \text{i and j are linked} \\ \infty, & \text{i and j are not linked} \end{cases} \tag{6}$$

Where p_{ij} denotes the weight of the edge between node i and j. Particularly in an unweithed network, p_{ij} is infinite, which can be shown as:

$$p_{ij} = \begin{cases} weight\ of\ edge, & \text{weithed netwok} \\ \infty, & \text{unweithed network} \end{cases} \tag{7}$$

Stephen P. Borgatti (2006) gets each node's contribution simply by subtracting DFs of the network with and without the node, which ignores the order of the nodes. This problem can be solved by applying Shapley approach [Shapley, 1953]. Shapley value is typically used in game theory, and has been developed recently into various fields, including influence measurement in networks. The Shapley value can be showed in the following way:

$$\Omega_i(v) = \frac{1}{N!} \sum_{O \in \pi_N} \left(v\big(Pre^i(O) \cup i\big) - v\big(Pre^i(O)\big) \right) \tag{8}$$

Where O is defined as a permutation of players, $v\left(Pre^i(O)\right)$ is the value obtained in permutation O by the players preceding player i and $v\left(Pre^i(O)\cup i\right)$ is the value obtained in the same permutation when including player i. That is, $\Omega_i(v)$ gives the average marginal contribution of player i over all permutations.

To avoid too much calculation due to the rapid increase of number of permutations with the number of players, the Shapley value can be approximated by the average contribution nodes to network's cohesion over k randomly sampled permutations, which can be shown as:

$$\Omega_i(v) \approx \widehat{\Omega}_i(v) = \frac{1}{k}\sum_{O\in\pi_k}\left(v\left(Pre^i(O)\cup i\right) - v\left(Pre^i(O)\right)\right) \qquad (9)$$

It can be easily proved that the sample mean $\widehat{\Omega}_i(v)$ is an unbiased estimator of the population mean $\Omega_i(v)$[Bluhm M, Faia E, Krahnen J P., 2013].

In this case, we combine the adjusted cohesion measurement with the Shapley approach by defining the value used in Shapley equation as the cohesion measurement DF:

$$v = \text{DF} \qquad (10)$$

As a smaller DF indicates a more cohesive network, it is obvious that a negative Shapley value shows a node's positive contribution to the cohesion of the whole network. And the smaller a node's Shapley value is, the more contribution it makes to the network's cohesion, thus it is more influential in the network.

Application in Edros Network

We firstly apply this methodology to the Edros Network built above, which is an unweighted network, where   , We find the Top 10 influential co-authors, and compare the results with that ranked by betweenness, as is shown in **Table 3**.

Table 3. Top 10 influential co-authors ranked by Shapley value and Betweenness

Rank	Name(Shapley value)	Name(Betweenness)
1	ALON, NOGA M. (-0.001865084)	HARARY, FRANK* (9587.729986)
2	GYARFAS, ANDRAS (-0.001762582)	SOS, VERA TURAN (8912.230744)
3	SOS, VERA TURAN (-0.001612491)	RUBEL, LEE ALBERT* (8573.363779)
4	KLEITMAN, DANIEL J. (-0.001596847)	STRAUS, ERNST GABOR* (8300.743649)

5	TUZA, ZSOLT (-0.00151064)	POMERANCE, CARL BERNARD (7434.669391)
6	GRAHAM, RONALD LEWIS (-0.001494127)	FUREDI, ZOLTAN (7410.069488)
7	BOLLOBAS, BELA (-0.00149078)	ALON, NOGA M. (6871.898202)
8	HARARY, FRANK* (-0.001472003)	GRAHAM, RONALD LEWIS (6817.155314)
9	FAUDREE, RALPH JASPER, JR. (-0.0013609)	BOLLOBAS, BELA (6699.546721)
10	RODL, VOJTECH (-0.001300204)	PACH, JANOS (6073.643237)

We can see from **Table 3** that ALON, NOGA M. makes the most contribution to the Edros Network's cohesion, while HARARY, FRANK* bridges the most pairs in the network. Moreover, there are five same co-authors in Top 10 ranked by the two different measurements, who are ALON, NOGA M., SOS, VERA TURAN, GRAHAM, RONALD LEWIS, BOLLOBAS, BELA, and HARARY, FRANK*. They are supposed to be the most influential co-authors in the network.

Application in Movie Star Network

To implement our algorithm on a completely different set of network influence data, we dig the cooperative relations between 20 movie stars who have all cooperated with Jude Law. The same method is used to build the movie star network and find the Top 5 influential stars in the network. Note that the movie star network is a weighted network, and the weight of each edge depends on the times the stars have cooperated.
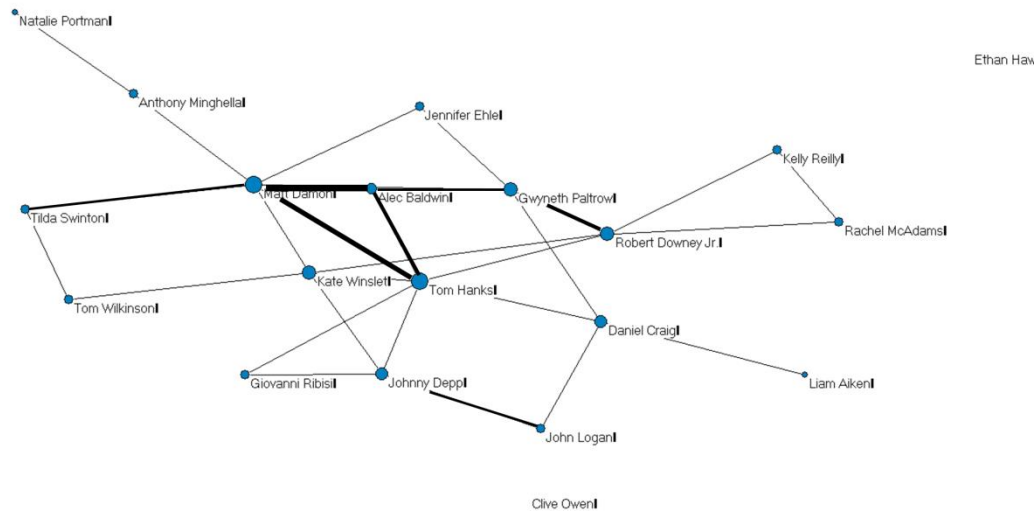
Figure 4. the Movie Star Network (The width of the edges reflects the times the two stars have cooperated, and the size of the nodes reflects their degrees)

Table 4. Top 5 influential movie stars ranked by Shapley value and Betweenness

Rank	Name(Shapley)	Name(Betweenness)
1	Matt Damon (-0.045389808)	Matt Damon (29.1436926)
2	Tom Hanks (-0.041607503)	Tom Hanks (24.3010304)
3	Robert Downey Jr. (-0.018930241)	Robert Downey Jr. (18.4906711)
4	Daniel Craig (-0.01560871)	Daniel Craig (13.6563631)
5	Gwyneth Paltrow (-0.014915669)	Kate Winslet (13.5588972)

The Top 5 influential movie stars ranked by Shapley value and betweenness are almost the same, which is due to the relatively small size of the network. The difference in the fifth most influential star indicates that the removal of Gwyneth Paltrow has more impact on the cohesion of the network than that of Kate Winslet, although the latter bridges more pairs in the network.

Strengths

● The Shapley value is superior to degree and closeness centrality in that it can be applied to both unweighted and weighted networks.

● The Shapley value is more appropriate than betweenness in the scenario where

a node has higher betweenness, but the removal of it has less impact on the network's cohesion.

- The Shapley value considers the order when calculate the contribution of each node to the cohesion of the whole network, so the result is more accurate than that of simply subtracting the DFs of the network with and without the node.

Weaknesses

- The calculation of Shapley value is more complex than that of conventional centrality.
- The Shapley value can't be applied to directed networks, since the definition of the cohesion measurement DF does not reflect the directions of edges. So another method needs to be introduced to measure the influence of nodes in directed networks.

Utility

The Shapley value is applicable to most undirected networks.

Individual level

- Improve social influence. Real social network is a typical undirected network, where one can always use the Shapley value to measure others' impact on the network. And choosing the people with lower Shapley value to make acquaintance with can improve one's social influence efficiently.
- Make career decisions. As is illustrated above, cooperation networks like Edros Network and Movie Star Network are also undirected networks. Cooperating with colleagues with lower Shapley value may boost your career status rapidly.

Organizational level

- Choose business collaborator. Business cooperation network is similar to Edros Network and Movie Star Network, so the same strategy can be applied in order to boost influence.

National level

- Improve international relationship. Similarly, countries that want to be more active in international events can choose to establish diplomatic relations with those with lower Shapley value.

Influence measure for directed network: Optimized Pagerank

The undirected network discussed above will produce a symmetric adjacency

matrix. In contrast, directed networks, such as the web page network and the paper citation network, are useful to study relative asymmetries and imbalances in link formation. To analyze the influence ranking within a directed network, we take the paper citation network as an example to study the methodology called Pagerank. And we use the list of all 16 papers provided by the 2014 ICM problem file as our data. In addition, the Pagerank measurement is optimized to apply to the real citation network better according to the total numbers of the references of a certain paper under consideration and the outside papers citing the target paper.

The basic Pagerank model

Pagerank is an algorithm originally used to measure the value of a website page according to the number and quality of links to it. Because both of the website network and the citation network can be regarded as directed graphs when papers or webs are nodes and reference relationships or web links are arcs, we can conduct Pagerank method to evaluate the contribution of papers. The main idea of this method is that the relative importance of a certain node in a directed network hinges on the importance of nodes attaching outbound links to it. Specifically, in the citation network, the distribution of a target paper depends on the number and influence of papers citing it. So the Pagerank value of a paper can be calculated as the following equation:

$$\mathrm{PR}(p_i) = d \sum_{p_j \in \Pi(p_i)} \frac{\mathrm{PR}(p_i)}{\mathrm{L}(p_j)} + \frac{1-d}{N} \qquad (11)$$

Where $p_1, p_2, p_3, \ldots, p_N$ present the papers in the network as the subscript means their corresponding number based on the initial order, $\mathrm{PR}(p_i)$ means the Pagerank value of the target paper p_i , $\Pi(p_i)$ is the set of papers that have cited p_i , $\mathrm{L}(p_j)$ is the number of outbound links from p_j, and N is the total number of the papers in the network, which is 16 in this case .

The denominator $\mathrm{L}(p_j)$ means that every paper distributes equal link weight to each of its reference in the network. And it's worth noting that there would always be at least one paper that does not cite other papers in the network, such as the paper first published. So here a damping factor $d,$ is introduced into the model to deal with the papers with no outbound links by assuming every paper to link out to all other papers in the collection evenly with a probability of $1 - d$. The empirical value of d is 0.85, which means a paper's Pagerank is 0.85/N even no paper links to it.

The iterative method

According to the definition equation of Pagerank value, we use MATLAB to

operate the algorithm. Again, the adjacency matrix ,W, extracted from the references of each paper with equal link weight is used to describe the relations and links among the total 16 papers, where $r_{ij} = 1$ means that p_i cites p_j .In matrix form, the Pagerank matrix can be manifested by the following equation.

$$PR = d * (W^*)'PR + (1 - d)\left(\frac{E * E'}{n}\right) \qquad (12)$$

Where W^* means the modified adjacency matrix after dividing every element by the row sum.

Then we use the iterative method to get the most accurate value results under the assumption that the distribution is evenly divided among all papers at the beginning of the computational process. According to the Pagerank value of every element, we can get the most influential paper with the highest Pagerank value (see results in column 2 of **Table 5.**).

The weakness of the basic model

Since the basic Pagerank method is derived from websites rank, such an algorithm has weakness and cannot be simply applied to the paper influence rank due to the following distinguishes between them.

- Once built, the citation link or network is static while the website links are dynamically adjustable.
- The citations must be unidirectional whereas it is generally seen that two web pages link to each other.
- The noise of the citation network is much less than the website network. In general, every citation linked to a certain paper has relevance with it in a sense, but some page links in the website network contains useless information, such as advertisements, site navigations and friendly links.
- The lack of enough citations makes the Pagerank value meaningless in a small network.

So it is necessary to improve the basic model in order to satisfy the properties of the citation network.

The optimized Pagerank model

By taking other papers as a whole into account, we consider that other papers outside the network can also influence the contribution of the target paper. To optimize the Pagerank method, we assume that there is a virtual N+1 node, which presents all of the other papers beyond the 16 papers (see **Figure 5**). These papers
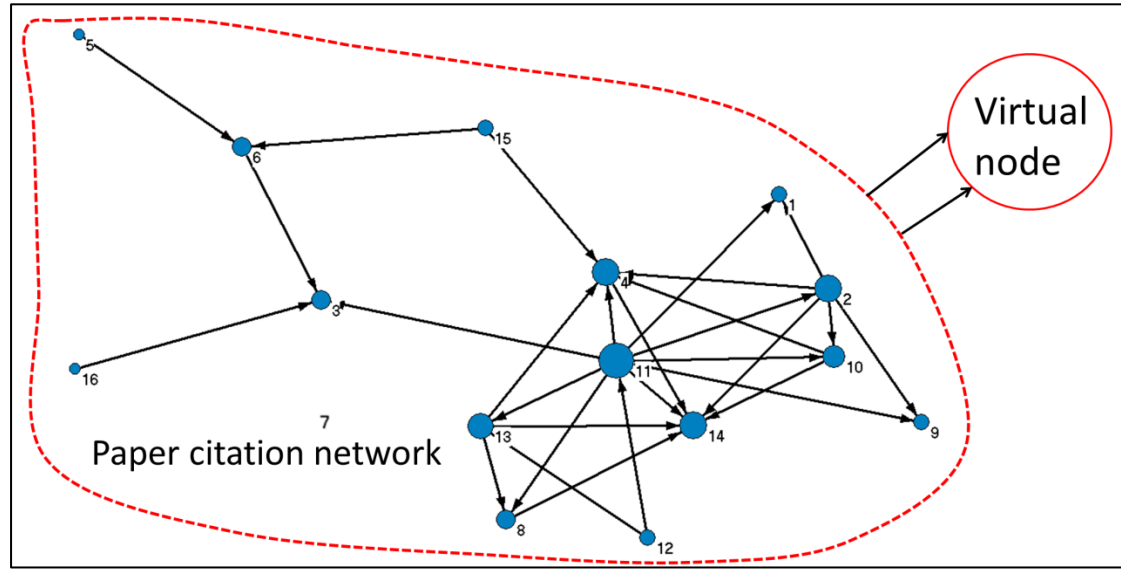
may also cite or be cited by the target paper.

Figure 5. Extensive citation network. The size of a node corresponds to the all degree (in and out) of a paper. Arrows mean one paper cites another. Imaginary line separates the initial citation network and the outside virtual node.

After adding the virtual node, we get a new modified adjacency matrix(N+1×N+1) of the extensive citation network.

$$
\begin{bmatrix}
r_{11} & r_{12} & \cdots & r_{1j} & \cdots & r_{1N} & r_{1(N+1)} \\
r_{21} & r_{22} & \cdots & r_{2j} & \cdots & r_{2N} & r_{2(N+1)} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
r_{i1} & r_{i2} & \cdots & r_{ij} & \cdots & r_{iN} & r_{i(N+1)} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
r_{N1} & r_{N2} & \cdots & r_{Nj} & \cdots & r_{NN} & r_{N(N+1)} \\
r_{(N+1)1} & r_{(N+1)2} & \cdots & r_{(N+1)j} & \cdots & r_{(N+1)N} & r_{(N+1)(N+1)}
\end{bmatrix}
$$

Where
$$
r_{ij} = \begin{cases} 1/C\text{, when } p_i \text{ cites } p_j \text{ and C is the number of all reference papers of } p_i \\ 0\text{, when } p_i \text{ does not cite } p_j \end{cases}
$$

$$
r_{i(N+1)} = 1 - \sum_{j=1}^{N} r_{ij}
$$

$$
r_{(N=1)j} = 1/N
$$

$$r_{(N+1)(N+1)} = 0$$

As the improved matrix shows, every reference paper in the original network get equal citation link weight from the source paper while the new N+1 node(the other related papers) get the rest of the reference weight. Also, the N+1 node distribute its link weight to N papers evenly. By this way, we consider the whole list of references instead of the limited 16 objects in the network and adjust the weight allocated to the cited paper according to the total number of the references. Similarly, we can use the iterative method to calculate the Pagerank value results (see column 3 of Table X)in the optimized model.

Involving Times Cited

The model discussed above assumes that the outside papers attach equal weight to the paper in the network as the element of the N+1 row is the same. Furthermore, we think that outside papers endows the inside papers with different weight on the basis of the target paper's Times Cited since its publication. So the elements of last row are changed into:

$$r_{(N+1)j} = T_j / \sum_{j=1}^{N} T_j \qquad (13)$$

Where T_j is the Times Cited data of p_j from *Google Scholar.*

After considering the real Times Cited, the third method involves both the papers having outbound links and inbound links to the initial network and can reflex the real integrated impact of the target paper. And the new outcome can be seen in the column 4 of **Table 5.**

Results

We can get three groups of results from the basic model, optimized model and further optimized model (involving Times Cited) respectively. In addition, we select the top 5 from 16 papers in accordance with their Pagerank values to compare the effectiveness of these three methods in the Pagerank model. The importance ranking results can be seen in **Table 5**, where the value under the paper number is its Pagerank value.

Table 5. Pagerank results of three methods in the Pagerank model

Rank	Basic Pagerank Model	Optimized Pagerank Model	
		No Times Cited	With Times Cited
	Paper number	Paper number	Paper number
1	14 (0.7681)	14 (0.3339)	14 (1.0312)
2	3 (0.3846)	4 (0.295)	4 (0.8562)
3	4 (0.2309)	6 (0.2871)	2 (0.6146)
4	6 (0.1515)	8 (0.2842)	11 (0.5071)
5	8 (0.1104)	3 (0.2806)	1 (0.2618)

As shown in table above, the most important paper remains the same under the three methods which is No.14, *Collective dynamics of `small-world' networks*. It is because No.14 is always contributing most and cited by other important papers more frequently (its Times Cited on *Google Scholar* is 21688) no matter in the network with limited size or that involving other papers.

Additionally, the papers which rank 2nd to 5th are changed after the model is optimized. Specifically, the ranking of No.4 is elevated from 3rd to 2nd after extending the network to outside papers because No.4 gets larger link weight relatively. Another obvious change is that No.3 is absent from the Top list mainly since its Times Cited (1956) is relatively small. On the contrary, No.2 and No.11 appear on the Top list of optimized Pagerank Model with Times Cited due to their large Times Cited (13250 and 10616 respectively).

According to the results, the optimized Pagerank method is not confined to the target papers and can be better applied to distinguish the most important paper from others by evaluating synthetically the Pagerank values of papers within an extensive network that considers both outbound and inbound links.

Strengths

● High practicability. The Pagerank model is suitable to analyze the importance ranking in the complex network, especially in directed graphs cause it can distinguish outbound links from inbound links.

- Optimized model. By involving outside citation links into the initial network, the optimized Pagerank model can break through the size limit of the citation network so as to evaluate the integrated influence of a paper.

- Wide application. The model can be applied universally to the entity importance ranking, such as the key words, scientific institutions and countries, as well as the websites, papers and journals.

Weaknesses

- Data limitations. The effective practice of the Pagerank model needs a large deal of information about the reference relationship or the accuracy of the influence results changes along with the different network size and data abundance.

- Improved application. Using Pagerank to assess papers or institutions is a new field, so it still remains several problems of applications, such as the dynamic noise of the network, and needs further practice to improve the application methods.

Utility

Our optimized Pagerank model is applicable to many networks especially directed networks.

Individual level:

- Choose the one to follow on social network society. Social relationships on the internet can be considered as a directed network because the links can be one-way. For example, if you follow a friend on twitter while this friend does not follow you in return, then there is a link from you to your friend, but the reversed link does not exist. In order to get as much information as possible in a field in a short time, we can build a network to figure out those who have the highest Pagerank values and those who play the most important roles in this network.

- Make academic decisions. With the data of citation relationships among papers written by different thesis advisors, we can figure out the Pagerank value of each thesis advisor. Thesis advisors with higher Pagerank values are likely to have stronger scientific research ability. In a similar way, by studying the papers written by all the teachers and students at the school, we can choose a satisfying school to go to.

National level:

- Control systemic risk. Interbank borrowing market is a typical directed network which represents the debtor-creditor relationships between different financial institutions. Financial institutions with high Pagerank value are too interconnected

to fail because they play important roles in financial contagion and have a great influence on other institutions. According to the Pagerank model, the supervisor is supposed to pay special attention to these financial institutions to control the systemic risk of the whole country.

Sensitivity analysis

Shapley value's sensitivity to parameter *k*

One fatal flaw of the Shapley value is the huge computation. Although the complexity has been mitigated from Equation (8) to (9) by replacing the total number of permutations N! with the sample size of *k*, it still needs a lot of time and memory space to run the program. So we explore the sensitivity of the results to the sample size k by tracing 5 co-authors' rank under different *k*, as is shown in **Figure 6.**
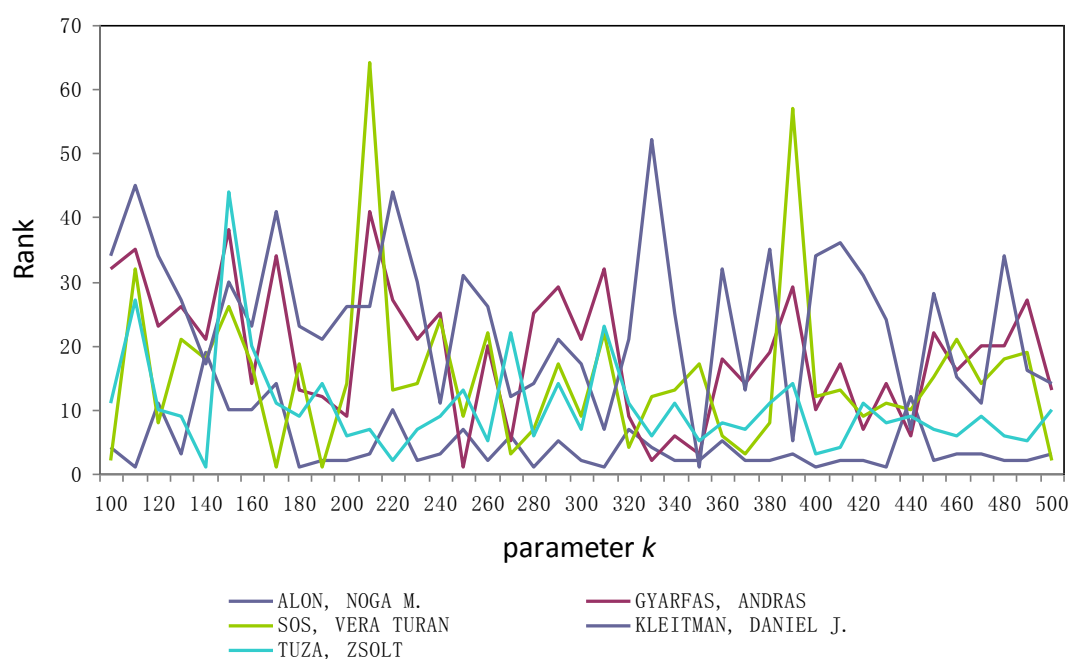

Figure 6. Sensitivity analysis results of Shapley value to parameter *k*. The 5 different lines present 5 sets of ranks of 5 co-authors.

As is shown above, the co-authors' ranks differ remarkably under different value of parameter *k*. And ranks tend to be stable with the increase of sample size *k*, which is consistent with the basic principle of statistics. So we suggest selecting *k* that are large enough to guarantee the accuracy of the results, but not too large to save time and space. In this case, a figure over 400 can provide reasonable accuracy.

Pagerank value's sensitivity to parameter *d*

Considering the principle and computational process of Pagerank algorithm, it is easy to find that the results of this methodology may be influenced by the damping factor d which is an empirical constant 0.85 in the model above. Since the parameter d can change the probability of directing to a random node in the network, we can conduct the sensitivity analysis of Pagerank value rank to parameter d. Here we use the optimized Pagerank method with Times Cited .By changing the value of d from 0.05 to 1 with the interval of 0.05, we observe the corresponding ranks of the original Top 5 papers (in the column 4 of **Table 5**). The outcome of sensitivity analysis can be seen in **Figure 7**.
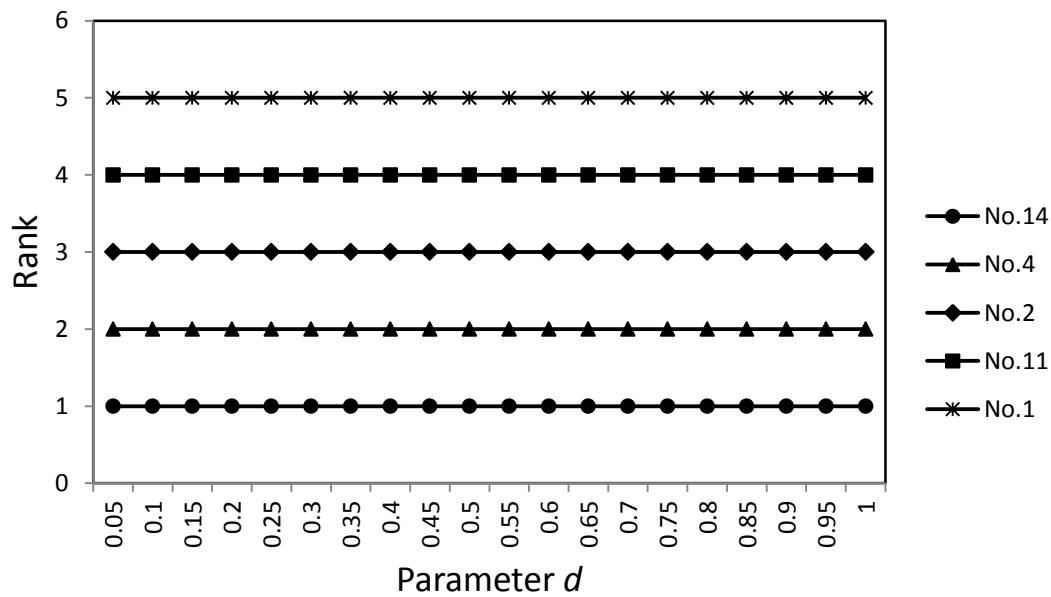

Figure 7. Sensitivity analysis results of Pagerank values to parameter d. The 5 different lines present 5 sets of ranks of the 5 certain paper.

Obviously, the results reveal that the rank of each paper is not changing along with the change of parameter d, which means that the Pagerank method result is not sensitive to d.Although d can change the probability of random links, its action principle is to add a constant weight to every element in the network to avoid the misconvergence of Pagerank value but does not change the original importance of a certain paper and the link weight distribution. As a result, the parameter d has nothing to do with the Pagerank value.

Reference

Albert,R. and Barabasi,A-L.(2002).Statistical machanics of complex networks. *Reviews of modern Physics*,74:47-97.

Markose, S., Giansante, S., Gatkowski, M. and Shaghaghi, A. R.(2009).Too interconnected to fail: Financial contagion and systemic risk in network model of cds and other credit enhancement obligations of us banks. *analysis*.

Freeman L C.(1979).Centrality in social networks conceptual clarification[J]. *Social networks*, 1(3): 215-239.

Borgatti, S. (2006).Identifying sets of key players in a network. *Computational and Mathematical Organization Theory*, 12: 21-34.

Shapley L S.(1952).A value for n-person games[R]. *RAND CORP SANTA MONICA CA.*

Bluhm M, Faia E, Krahnen J P. (2013).Endogenous Banks' Networks, Cascades and Systemic Risk[R]. *Mimeo, Goethe University*.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, *393*(6684), 440-442.

Borgatti, S. P., Jones, C., & Everett, M. G. (1998). Network measures of social capital. *Connections*, *21*(2), 27-36.

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404-409.

Friedkin, N. E. (1991). Theoretical foundations for centrality measures. *American journal of Sociology*, 1478-1504.

Knoke, D., & Yang, S. (2008). *Social network analysis* (Vol. 154). Sage.

Watts, D. J. (1999). *Small worlds: the dynamics of networks between order and randomness*. Princeton university press.

Moody, J., & White, D. R. (2003). Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review*, 103-127.

Kleinberg, J. M. (2000). Navigation in a small world. *Nature, 406*(6798), 845-845.

Everett, M. G., & Borgatti, S. P. (2000). Peripheries of cohesive subsets. *Social Networks*, *21*(4), 397-407.

Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social networks*, *21*(4), 375-395.

Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, *13*(5), 533-549.

Seidman, S. B. (1983). Network structure and minimum degree. *Social networks*, *5*(3),

269-287.

http://baike.baidu.com/link?url=WckfiBvPGGGMxZd5DISq5UKMIcfibz2Nn-HwEOym

TU0QbYQf2hmZy17u4oa9_Kz2ZquoU91EHOIjlq6T8DL-va

http://baike.baidu.com/link?url=aOCz9onz7l-zRZrgHOqsCHpjgd_A2Q_Wvz55QJ8mX

uT1znQ54NPTdCHFxoIVTAfpFlyk4dv2ryrKI3eSsitu1K

http://en.wikipedia.org/wiki/Complex_network