Team Control Number

25830

For office use only		For office use only
T1 ________________		F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	C	F4 ________________

A Network-based Model for Measuring Author's Influence

Summary

Our model is based on this idea that how the whole network changes when one node is removed from it. Based on this idea, we firstly define the network's prestige by summing all the nodes' Katz-Bonacich centrality, and then calculate the change of the network's prestige when one node is removed. Accordingly, the influence of one node can be measured by the change of the network's prestige. To express this idea formally and mathematically, we establish an optimal model to find the most influential node and also rank the important nodes by their influences. As for the optimal model, we, by strict mathematical proofs, provide its solution in a concise form and also decrease the time complexity of solving it to $O(n^3)$ which is the greatest lower bound of our model because the calculation of $n \times n$ matrix's inverse needs the time complexity of $O(n^3)$. Besides, we compare our influence index with degree centrality, between centrality, and Bonacich centrality by exemplifying, and apply it into the coauthor network established from the database given by the contest.

In our model's framework, we prove a lemma to present how many influence a new comer will obtain when he or she connects an existed one in the network. As a result, we give the optimal strategy for the new comer to find the desirable coauthor in aim to increase the influence rapidly. Furthermore, we also make numerical experiments in the established coauthor network to demonstrate our result's correctness.

Note that two kinds of parameters $\alpha_i \ (i = 1, 2, \cdots, n)$ and β exist in our model, where the formers are used for reflecting the nodes' prior knowledge and the latter for reflecting the linkage's strength and controlling the model's robust especially considering the network's incompleteness. As for the given 16 network-science papers in this contest, we value α_i by each one's citation number which is a common standard for evaluating a paper's influence and compare the rank differences between highlighting the α_i and not highlighting it in the citation network formed by these papers. Also, we consider any network may be one part of the complete network existing in our world, thus we analyze our model's robustness by simulation test. In the simulation analysis, the nodes' influence rank in the sampled network (with different sampling rate and β values) is compared with their rank in the complete network. The main finding is the smaller β can better keep the nodes' rank in sampled network, but a small β may underestimate the network's linkage effect if considering the parameter's implication.

We also apply our model into a weighted network. The application demonstrates that our model can be applied into the network no matter it is weighted or unweighted, directed or undirected, since it has been applied to the coauthor network (unweighted and undirected network) and the citation network (directed network) in this paper. We hope our model can be proved useful in many potential applications in near future.

Contents

1 Introduction

1.1 Our Idea

Taking the significant influence of an element or a node into account is the pivotal task on using networks to measure influence and impact. However, the influence not only depends on itself. More often, we just have to consider all kinds of factors in the whole network as well. Thus, we think that, the influence of a node means the change of overall influence in the network after the node has been removed. So, we will consider the entire network instead of the node itself or its neighbors.

1.2 Model Design

- According to our idea, we establish our optimal model and prove it based on Katz-Bonacich centrality [Seeley, 1949; Katz, 1953; Bonacich, 1987; Ballester et al, s2006]. Then we optimize our algorithm so that the time complexity of the algorithm has been reduced to $O(n^3)$ from $O(n^4)$ when we calculate the influence.
- In Section 2.4.1, we got Erdos1 network statistics, while in Section 2.4.2 we apply this model to the erdos1 datasets obtained top10 influential writers.
- In Section 3, our design solves the problem of how to make a decision to improve the influence of the nodes, which contains new ones and existed ones in the network, on the basis of our model.
- With sensitivity analysis, we discover the rules about the factors and how they affect the results of our model based on plenty of data. In Section 4, by using 16 papers' data, we analyze the effect of α . And in Section 5, we analyze the effect of β with Kendall Tau statistics.
- In Section 6, we take the applications of different networks on different fields into consideration. Our model can be applied in all kinds of networks, whose complexity is from low to high, including undirected unweighted graph, directed unweighted graph and directed weighted graph. It almost covers various networks that have to calculate the influence of nodes.

2 An optimization-based model for measuring the key author

In this section, we extract data in Erdos1 and get the co-author network. We establish our optimal model and prove it based on Katz-Bonacich centrality. By applying this model, we have the influence ranking of the authors in the Erdos network.

2.1 The Optimal Model

We measure the significant influence of one author by examining how much the whole network's prestige changes when the targeted author is removed from the co-author network. Based on this idea, the more the whole network's prestige changes when removing the author, the more influential the targeted author is. Before presenting the optimal model formally, we should provide the definition of **the node's prestige** (here, the nodes represent the authors) and **the whole co-author network's prestige** firstly, and then explain the rationality of these definitions.

[**Definition 1**] (**The node's prestige**). We apply Katz-Bonacich centrality [Seeley, 1949; Katz, 1953; Bonacich, 1987; Ballester et al, 2006] to define the node's prestige. Consider a network **g** with adjacency matrix **G**, identity matrix **I** and a scalar *b* such that $\mathbf{B}(\mathbf{g},\beta)=[\mathbf{I}-\beta\mathbf{G}]^{-1}$ is well defined and nonnegative. The node i's prestige denoted p_i in network $\mathbf{g}$ is

$$p_i(\mathbf{g})=\alpha_i\cdot\sum\nolimits_{j=1}^{n}b_{ij}(\mathbf{g},\beta), \tag{1}$$

where, $b_{ij}(\mathbf{g},\beta)$ is the element of matrix $\mathbf{B}(\mathbf{g},\beta)$ at row *i* and column *j*. Here, β is a factor that captures how the value of being connected to someone decays with distance, while α_i captures the base value on each node. □

Note that the above mentioned base value can come from some priori knowledge of the node, for example if we have known some nodes are more important than the others as the question (2) hints, we can give a higher α_i to these known nodes. However, if we do not have any priori knowledge at all, all the α_i s should be given the same value; namely in this case, we don't need to highlight the α_i in formula (1).

[**Definition 2**] (**The whole co-author network's prestige**). The whole network's prestige can be obtained by summing all the nodes' prestige values. Let $p(\mathbf{g})$ denote the whole prestige in network $\mathbf{g}$, we have

$$p(\mathbf{g})=\sum\nolimits_{i=1}^{n}p_i(\mathbf{g}), \tag{2}$$

where, *n* is the number of nodes in network $\mathbf{g}$. □

The above definitions are based on Katz-Bonacich centrality [Seeley, 1949; Katz, 1953; Bonacich, 1987; Ballester et al, 2006] which is beyond these fairly direct measures of centrality such as input or output degree centrality, decay centrality, betweenness centrality, and so forth, because Katz-Bonacich centrality is more elegant both mathematically and in terms of the ideas that it captures [Jackson, 2010]. Katz-Bonacich centrality is based on the premise that a node's importance is determined by how important its neighbors are. That is, we might like to not only account for the fact that a node is connected or close to many other nodes, but that it is close to many other important nodes. This notion is central to citation rankings. Thus, such a centrality depends on how central its neighbors are, which depends on the centrality of their neighbors, and so forth. It can be found that the definition is supported by the recent researches as the literatures listed above mentioned. Besides, these definitions can meet the demands of the questions in the contest.

Next, we denote $\mathbf{g}_{-i}$ as the resulting network of removing node *i* from network $\mathbf{g}$. Then the most influential node can be mined by solving the following optimal question:

$$\max_i \{p(\mathbf{g}) - p(\mathbf{g}_{-i}) \mid i = 1, 2, \cdots, n\},$$
$$s.t. \quad p(\mathbf{g}) = \sum_{k=1}^{n} \alpha_k \cdot \sum_{j=1}^{n} b_{kj}(\mathbf{g}), \tag{3}$$
$$p(\mathbf{g}_{-i}) = \sum_{k \in n \setminus \{i\}} \alpha_k \cdot \sum_{j \in n \setminus \{i\}} b_{kj}(\mathbf{g}_{-i}).$$

Accordingly, **the value of $p(\mathbf{g}) - p(\mathbf{g}_{-i})$ can be the measure of the node *i*'s influence** from the viewpoint of the whole network rather than from the node's local information. Thus, our measure decides one node's influence by not only accounting for the fact that a node is connected or close to many other nodes, but that it is close to many other important nodes. We will exemplify the priority of our measure in section 2.3 and present the results by applying our new measure into the network established from the file Erdos1 provided by the contest.

2.2 The Optimal model's Solution

Recall the formula (3), $p(\mathbf{g}_{-i})$ is different with different *i*s, so it seems that the value of $p(\mathbf{g}_{-i})$ is needed to be calculated each time to get each node's influence value $p(\mathbf{g}) - p(\mathbf{g}_{-i})$ ($i = 1, 2, \cdots, n$). Note that calculating the $p(\mathbf{g}_{-i})$ needs to compute $\mathbf{B}(\mathbf{g}_{-i}, \beta) = [\mathbf{I} - \beta \mathbf{G}_{-i}]^{-1}$ with the time complexity $O(n^3)$, where *n* is the node number of the given network. Can we get the results of all $p(\mathbf{g}_{-i})$ ($i = 1, 2, \cdots, n$) by just calculating the inverse of matrix just once? It is the question we want to answer in this part. Of course, if we can get all the values of $p(\mathbf{g}_{-i})$s with *i* from 1 to *n* by just computing the $\mathbf{B}(\mathbf{g}, \beta) = [\mathbf{I} - \beta \mathbf{G}]^{-1}$ once, it will uplift our measure's applicability. First of all, we give **theorem 1** to lay the foundation for answering the question.

[**Theorem 1**] Given $\mathbf{B}(\mathbf{g}, \beta) = [\mathbf{I} - \beta \mathbf{G}]^{-1}$ as shown in our Definition 1, it holds that $b_{ij}(\mathbf{g}, \beta) \cdot b_{ik}(\mathbf{g}, \beta) = b_{ii}(\mathbf{g}, \beta) \cdot [b_{kj}(\mathbf{g}, \beta) - b_{kj}(\mathbf{g}_{-i}, \beta)]$ for all $k \neq i, j \neq i$.

Proof. Note that the matrix $\mathbf{B}(\mathbf{g}, \beta)$'s coefficients $b_{ij}(\mathbf{g}, \beta) = \sum_{l=0}^{+\infty} \beta^l g_{ij}^{[l]}$ count the number of paths in network **g** start at *i* and end at *j*, where paths of length *l* are weighted by β^l. Then it holds that

$$b_{ii}(\mathbf{g}, \beta) \cdot [b_{kj}(\mathbf{g}, \beta) - b_{kj}(\mathbf{g}_{-i}, \beta)] = \sum_{l=1}^{+\infty} \beta^l \sum_{r+s=l, r \geq 0, s \geq 1} g_{ii}^{[r]} (g_{kj}^{[s]} - g_{k(-i)j}^{[s]})$$
$$= \sum_{l=1}^{+\infty} \beta^l \sum_{r+s=l, r \geq 0, s \geq 2} g_{ii}^{[r]} g_{k(i)j}^{[s]} = \sum_{l=1}^{+\infty} \beta^l \sum_{r'+s'=l, r' \geq 1, s' \geq 1} g_{ki}^{[r']} g_{ij}^{[s']} = b_{ji}(\mathbf{g}, \beta) \cdot b_{ik}(\mathbf{g}, \beta).$$

Because of the co-author network **g** is symmetric here, it holds that $b_{ji}(\mathbf{g}, \beta) = b_{ij}(\mathbf{g}, \beta)$. Thus, the theorem holds. Note that $g_{k(-i)j}^{[s]}$ is the weight of *s* length paths from *k* to *j* that do not contain *i*, whereas $g_{k(i)j}^{[s]}$ is the one that contain *i*.

□

Based on the Theorem 1, we can deduce the following Corollary 1 directly by summing all the *k* and *j* when $k \neq i, j \neq i$.

[**Corollary 1**] The node's influence index (denoted by Inf_i) $p(\mathbf{g}) - p(\mathbf{g}_{-i})$ (namely, $Inf_i = p(\mathbf{g}) - p(\mathbf{g}_{-i})$) can be calculated by using

$$Inf_i = \frac{\sum_{j=1}^{n} \alpha_j b_{ij}(\mathbf{g}, \beta) \cdot \sum_{j=1}^{n} b_{ij}(\mathbf{g}, \beta)}{b_{ii}(\mathbf{g}, \beta)}. \tag{4}$$

Especially, when all the α_i s are equal (denote by α), we have

$$Inf_i = \alpha \frac{\left(\sum_{j=1}^{n} b_{ij}(\mathbf{g},\beta)\right)^2}{b_{ii}(\mathbf{g},\beta)}. \tag{5}$$

Proof. The formula $b_{ij}(\mathbf{g},\beta)\cdot b_{ik}(\mathbf{g},\beta) = b_{ii}(\mathbf{g},\beta)\cdot[b_{kj}(\mathbf{g},\beta) - b_{kj}(\mathbf{g}_{-i},\beta)]$ in Theorem 1 can be summed according to all the k and j when $k \neq i, j \neq i$ as follows:

$$\sum_{k\in n\backslash\{i\}} \alpha_k \sum_{j\in n\backslash\{i\}} \frac{b_{ij}(\mathbf{g},\beta)\cdot b_{ik}(\mathbf{g},\beta)}{b_{ii}(\mathbf{g},\beta)} = \sum_{k\in n\backslash\{i\}} \alpha_k \sum_{j\in n\backslash\{i\}} [b_{kj}(\mathbf{g},\beta) - b_{kj}(\mathbf{g}_{-i},\beta)],$$

Note that

$$\begin{aligned} p(\mathbf{g}) &= \sum_{k=1}^{n} \alpha_k \cdot \sum_{j=1}^{n} b_{kj}(\mathbf{g}) \\ &= \sum_{k\in n\backslash\{i\}} \alpha_k \cdot \sum_{j\in n\backslash\{i\}} b_{kj}(\mathbf{g}) + \alpha_i \cdot \sum_{j\in n\backslash\{i\}} b_{ij}(\mathbf{g}) + \sum_{k\in n\backslash\{i\}} \alpha_k b_{ki}(\mathbf{g}) + \alpha_i b_{ii}(\mathbf{g}), \end{aligned}$$

and then, we have

$$\begin{aligned} &\sum_{k\in n\backslash\{i\}} \alpha_k \sum_{j\in n\backslash\{i\}} \frac{b_{ij}(\mathbf{g},\beta)\cdot b_{ik}(\mathbf{g},\beta)}{b_{ii}(\mathbf{g},\beta)} + \alpha_i \cdot \sum_{j\in n\backslash\{i\}} b_{ij}(\mathbf{g}) + \sum_{k\in n\backslash\{i\}} \alpha_k b_{ki}(\mathbf{g}) + \alpha_i b_{ii}(\mathbf{g}) \\ &= p(\mathbf{g}) - p(\mathbf{g}_{-i}) = Inf_i. \end{aligned}$$

Accordingly, the formula (4) and (5) can be achieved at once. □

The Corollary demonstrates that all the nodes' influence Inf_i $(i = 1, 2, \cdots, n)$ can be obtained by just calculating the $\mathbf{B}(\mathbf{g},\beta) = [\mathbf{I} - \beta\mathbf{G}]^{-1}$ only once, because the formula (4) or (5) just needs the coefficients $b_{ij}(\mathbf{g},\beta)$ s of $\mathbf{B}(\mathbf{g},\beta)$ rather than any elements from $\mathbf{B}(\mathbf{g}_{-i},\beta)$ which changes with the different removing node *i*. As a result, the time complexity of our method keeps on $O(n^3)$, otherwise, it would be $O(n^4)$ without the Theorem 1 and Corollary 1.

2.3 Comparisons by Exemplifying

We have obtained the optimal model according to the series of proofs mentioned above. As a comparison, we also establish a simple network containing 9 nodes (**Figure 1**).

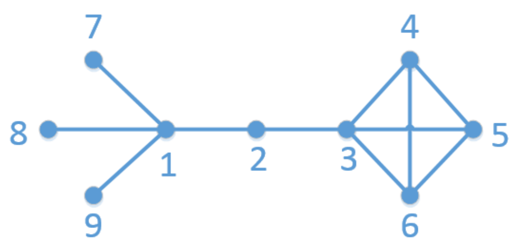

Figure 1 The network in our example

We collect the statistics of degree centrality, betweenness centrality, boncich centrality of each node in different β, and then get the value of each nodes' *Inf* according to formula (4) or formula (5). In **Table 1,** we examine the attributes of each node. We consider nodes 1 and 4, in which 4 along with 3, 5and 6, forms an alliance which influences each other, while 1 forms no such alliance with 7, 8 and 9. Thus, the more we emphasize this affecting connection, the more influential node 4 can make. In our model, the increase of β value means the enhancement of side connections, and the *Inf* value stresses more the abilities of the nodes influencing each other. So, according to the predictions of our model, when β increases, the *Inf* value of node 4 in this alliance will also increase. The actual data has proven our prediction. When β value equals 0.05,the *Inf* of node 1 is larger than that of node 4, while β value equals to or is larger than 0.1, the *Inf* of node 4 exceeds that of node 1. This proves that our model fully shows the ability of the nodes influencing each other in the network.

Table 1 Comparisons of different node centrality indexes

Node	degree	betweeness	bonacich	Inf(β=0.05)	Inf(β=0.1)	Inf(β=0.15)	Inf(β=0.2)
1	4	36	1.2152	**1.462**	**2.0873**	**2.9784**	**4.3803**
2	2	32	1.1224	1.2535	1.6649	2.3849	3.8397
3	4	30	1.2331	1.504	2.3298	3.8134	6.8916
4	3	0	1.1796	**1.3799**	**2.0108**	**3.1712**	**5.6755**
5	3	0	1.1796	1.3799	2.0108	3.1712	5.6755
6	3	0	1.1796	1.3799	2.0108	3.1712	5.6755
7	1	0	1.0608	1.1224	1.3031	1.5776	2.0269
8	1	0	1.0608	1.1224	1.3031	1.5776	2.0269
9	1	0	1.0608	1.1224	1.3031	1.5776	2.0269

2.4 Applications on the established network from the contest's database

2.4.1 Process of building the co-author network from Erdos1 file and its properties

With programming, we extract the authors whose Erdos number is 1, and establish an adjacency matrix based on the relationships between the authors. After that, we set up the co-author network, which is not entirely connected. The largest component of the established network is shown in **Figure 2**.

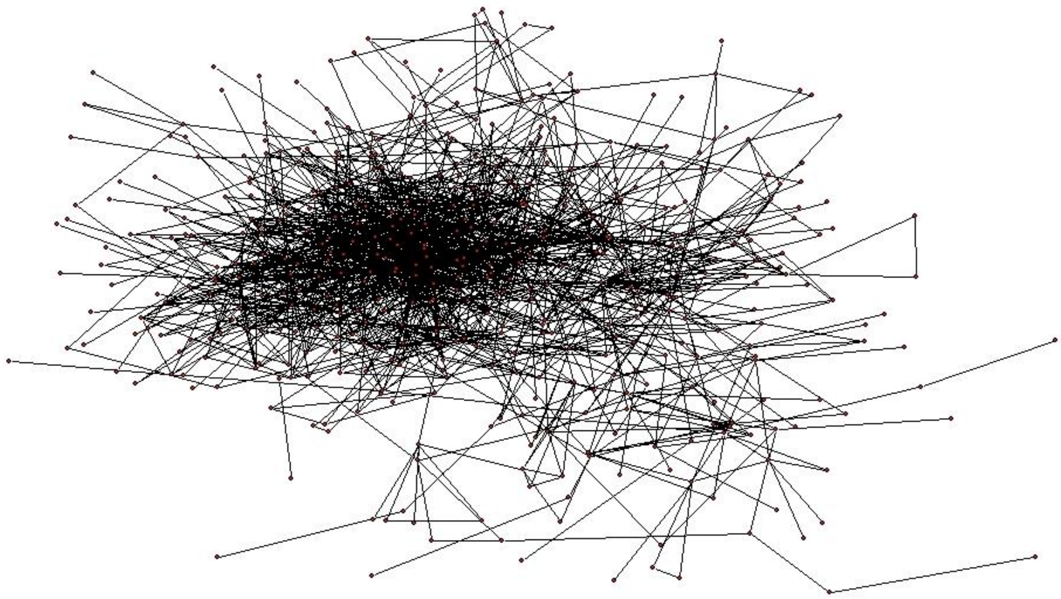

Figure 2 The largest component of the established network by the database

However, we find a mistake of the data in Erdos1 file. The author SARKOZY, ANDRAS and FUREDI, ZOLTAN should be co-authored with each other, so we supplemented this data. After the analysis of the properties of the network, we get some information displayed in **Table 2** and the distribution histogram of degree in **Figure 3**.

Table 2 The properties of this network

Index	Index value	Index	Index value
Number of nodes	511	Missing data	1
Number of edges	1640	Average clustering coefficient	0.343
Average degree	6.4188	The largest degree	52
Average path length	3.8246	Number of isolated node	37
Diameter of the largest component	10	Fraction of nodes in the largest component	91.19%

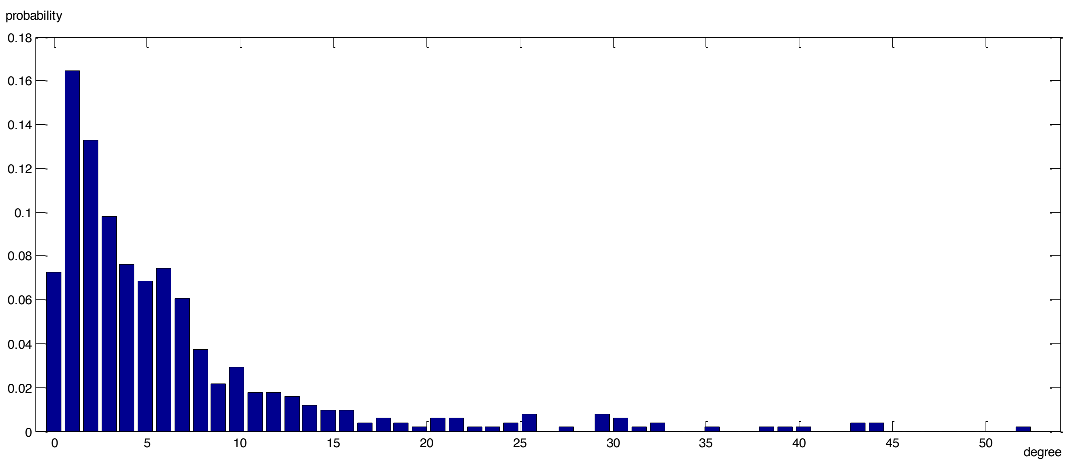

Figure 3 Degree distribution of the established network

2.4.2 The top 10 influential authors based on our model

Finally, we analyze the co-authored network to calculate the value of each author's Inf in different β according to the model we have designed in Section 2.2. The top 10 authors' information is displayed in **Table 3**. This is the application of undirected unweighted graph to our model, without considering the α value.

Table 3 Top 10 influential authors

Ranking	β=0.005	β=0.02
1	ANDERSON, JAMES MILNE	ANDERSON, JAMES MILNE
2	GRAHAM, SIDNEY WEST	ROGERS,CLAUDE AMBROSE*
3	ROGERS,CLAUDE AMBROSE*	GRAHAM, SIDNEY WEST
4	BONAR, DANIEL DONALD	BONAR, DANIEL DONALD
5	HARCOS, GERGELY	GAAL, STEVEN A. (GAL, ISTVAN SANDOR)
6	GAAL, STEVEN A. (GAL, ISTVAN SANDOR)	ULAM, STANISLAW MARCIN*
7	ULAM, STANISLAW MARCIN*	HARCOS, GERGELY
8	SPECKER, ERNST P.	SPIRO-SILVERMAN, CLAUDIA A.
9	SPIRO-SILVERMAN, CLAUDIA A.	SPECKER, ERNST P.
10	GYORI, ERVIN	GYORI, ERVIN

3 Who is the desirable co-author

We analyze how to make decision to improve the prestige of the new person and the existed one in the network optimally based on our model in this Section. In addition, we verify the accuracy of the decision with massive data.

3.1 The Optimal Strategy for a New Person

In order to boost one author's influence as rapidly as possible, he or she can make some strategies to find a co-author in the corresponding network. In the framework presented in Section 2, we have deduced out the formula (4) or (5) as the standard to compare these authors' influences. Here, we also use mathematical analysis to find out the desirable co-author who can boost one author's influence rapidly based on our established model, and we further compare the optimal result with the other results (namely to choose other persons as the co-author) based on the established network provided in the contest.

Intuitively, the author should find the influential authors to co-author some papers, but is the intuitive really correct? And can it be proved by strict mathematical analysis? In this part, we will give answers to these questions and provide the concise and correct standard. Note that a new person here means that he or she is a new node for the given co-author network, in which the new person has not co-authored anyone in the network.

We first prove the following Lemma 1. In this proof, please note that the node's prestige is different with its influence in our paper, which can be found in Section 2 and the examples in Section 2.3.

[**Lemma 1**] Consider the established network **g** with its adjacency matrix **G**. As for a new person i co-authoring with the author i' in the network, his or her influence index denoted by $Inf_{i\to i'}^{new}$ will be

$$Inf_{i\to i'}^{new}=\frac{\left(\alpha_i\cdot(1-\beta^2+\beta^2 b_{i'i'}(\mathbf{g}))+\beta\cdot\sum_{j=1}^{n}\alpha_j b_{i'j}(\mathbf{g})\right)\cdot\left(1-\beta^2+\beta^2 b_{i'i'}(\mathbf{g})+\beta\cdot\sum_{j=1}^{n}b_{i'j}(\mathbf{g})\right)}{1-\beta^2+\beta^2 b_{i'i'}(\mathbf{g})},\tag{6}$$

where, the parameters' meanings are the same with Section 2. Recall that $b_{i'j}(\mathbf{g},\beta)$ is the element of $\mathbf{B}(\mathbf{g},\beta)=[\mathbf{I}-\beta\mathbf{G}]^{-1}$ at row i' and column j.

Proof. In the first case, as for the new person i, his or her $\sum_{j\in n\cup\{i\}}b_{ij}(\mathbf{g}_{+i},\beta)$ and $\sum_{j\in n\cup\{i\}}\alpha_j b_{ij}(\mathbf{g}_{+i},\beta)$ for calculating its influence are (here, $\mathbf{g}_{+i}$ is the new network with adding the person i and its edge connected with the person i' into the original network **g**)

$$\begin{aligned}\sum_{j\in n\cup\{i\}}b_{ij}(\mathbf{g}_{+i},\beta)&=\sum_{j\in n\cup\{i\}}\sum_{l=0}^{+\infty}\beta^l g_{ij}^{[l]}(\mathbf{g}_{+i})=\sum_{j\in n}\sum_{l=0}^{+\infty}\beta^l g_{ij}^{[l]}(\mathbf{g})+\sum_{l=0}^{+\infty}\beta^l g_{ii}^{[l]}\\&=\sum_{l=0}^{+\infty}\beta^l g_{ii}^{[l]}+\sum_{j\in n}\sum_{l=1}^{+\infty}\beta^l g_{ij}^{[l]}(\mathbf{g})\\&=1+\frac{\beta^2}{1-\beta^2}b_{i'i'}(\mathbf{g})+\sum_{j\in n}\sum_{l=1}^{+\infty}\beta^l\sum_{r+s=l,r\ge1,s\ge0}g_{ii'}^{[r]}(\mathbf{g})\cdot g_{i'j}^{[s]}(\mathbf{g})\\&=1+\frac{\beta^2}{1-\beta^2}b_{i'i'}(\mathbf{g})+\frac{\beta}{1-\beta^2}\cdot\sum_{j=1}^{n}b_{i'j}(\mathbf{g})\end{aligned}$$

,similarly,

$$\sum_{j\in n\cup\{i\}}\alpha_j b_{ij}(\mathbf{g}_{+i},\beta)=\alpha_i+\frac{\alpha_i\cdot\beta^2}{1-\beta^2}b_{i'i'}(\mathbf{g})+\frac{\beta}{1-\beta^2}\cdot\sum_{j=1}^{n}\alpha_j b_{i'j}(\mathbf{g}).$$

Then, according to formula (4), its influence $Inf_{i\to i'}^{new}$ is

$$Inf_{i\to i'}^{new}=\frac{\left(\alpha_i\cdot(1-\beta^2+\beta^2 b_{i'i'}(\mathbf{g}))+\beta\cdot\sum_{j=1}^{n}\alpha_j b_{i'j}(\mathbf{g})\right)\cdot\left(1-\beta^2+\beta^2 b_{i'i'}(\mathbf{g})+\beta\cdot\sum_{j=1}^{n}b_{i'j}(\mathbf{g})\right)}{1-\beta^2+\beta^2 b_{i'i'}(\mathbf{g})}.$$

Especially, when all the α_j s are equal here, it becomes

$$Inf_{i\to i'}^{new}=\alpha\frac{\left(1-\beta^2+\beta^2 b_{i'i'}(\mathbf{g})+\beta\cdot\sum_{j=1}^{n}b_{i'j}(\mathbf{g})\right)^2}{1-\beta^2+\beta^2 b_{i'i'}(\mathbf{g})}.\tag{7}$$

□

Based on the formula (6) or (7) in the Lemma 1, the new person's influence can be calculated and be ranked. To calculate formula (6) or (7) just needs the known information from the $\mathbf{B}(\mathbf{g},\beta)=[\mathbf{I}-\beta\mathbf{G}]^{-1}$ which has been obtained in Section 2.

3.2 A satisfying Strategy for an Existed Person

In order to improve prestige, we will take the same strategy for an existed person as for a new person. The target person should co-author with the node whose value of bonacich centrality is the biggest. However, if he or she has been co-authored with the biggest one, we will choose the second biggest one. Repeat this process until you find a suitable node. Although we can't guarantee this case is the optimal solution, it must be the closest one. In the following example, based on our data, this strategy closely reaches the optimal solution and in most cases it is the optimal solution in the co-author network.

3.3 The numerical results based on the established co-author network

3.3.1 The new person

On the basis of the decision rules mentioned above, we will choose a co-author for the new person. As shown in **Table 4** and **Table 5**, it is distinct to see the growth in value of prestige and the new person's influence ranking after he co-authored with the top 10 authors in different β .

Table 4 Analysis of the new person where $\beta = 0.02$

Ranking	Name	Ranking after co-authored	The percent of improvement
1	ALON, NOGA M.	340	66.41%
2	RODL, VOJTECH	346	67.58%
3	GRAHAM, RONALD LEWIS	348	67.97%
4	BOLLOBAS, BELA	350	68.36%
5	FUREDI, ZOLTAN	355	69.34%
6	TUZA, ZSOLT	357	69.73%
7	HARARY, FRANK*	358	69.92%
8	SPENCER, JOEL HAROLD	373	72.85%
9	SOS, VERA TURAN	377	73.63%
10	GYARFAS, ANDRAS	387	75.59%

Table 5 Analysis of the new person where $\beta = 0.05$

Ranking	Name	Ranking after co-authored	The percent of improvement
1	ALON, NOGA M.	181	35.35%
2	RODL, VOJTECH	196	38.28%
3	BOLLOBAS, BELA	206	40.23%
4	GRAHAM, RONALD LEWIS	206	40.23%
5	FUREDI, ZOLTAN	208	40.63%
6	TUZA, ZSOLT	212	41.41%

7	SPENCER, JOEL HAROLD	213	41.60%
8	GYARFAS, ANDRAS	216	42.19%
9	SZEMEREDI, ENDRE	217	42.38%
10	FAUDREE, RALPH JASPER, JR.	221	43.16%

Intuitively, with enumeration, we will have the real optimal solution that co-authored with the author No.10. It follows that the model decision is the same as the optimal solution.

3.3.2 The existed person

We select the original reciprocal Top 5 writer, let them cooperate with theoretical collaborators selected by the model. The improvement of influence after cooperation shown in **Table 6** and **Table 7**:

Table 6 The improvement of influence where $\beta = 0.02$

Reciprocal ranking	Number of theoretical collaborators	The ranking after cooperation	The percent of improvement
ALAOGLU, LEONIDAS*	ALON, NOGA M.	341	33.27%
ANNING, NORMAN H.*	ALON, NOGA M.	340	33.27%
ASHBACHER, CHARLES D.	ALON, NOGA M.	340	33.07%
BENKOSKI, STANLEY J.	ALON, NOGA M.	339	33.07%
BUSOLINI, DONALD TERENCE	ALON, NOGA M.	335	33.46%

Table 7 The improvement of influence where $\beta = 0.005$

Reciprocal ranking	Number of theoretical collaborators	The ranking after cooperation	The percent of improvement
ALAOGLU, LEONIDAS*	ALON, NOGA M.	393	23.09%
ANNING, NORMAN H.*	ALON, NOGA M.	391	23.29%
ASHBACHER, CHARLES D.	ALON, NOGA M.	393	22.70%
BENKOSKI, STANLEY J.	ALON, NOGA M.	392	22.70%
BUSOLINI, DONALD TERENCE	ALON, NOGA M.	390	22.70%

With computer enumeration, the optimal solution is cooperated with ALON, NOGA M., which is the same as our decision.

4 Priori Knowledge—α

We establish a directed unweighted network composed of 16 papers. Applying our model to the network, we can get a ranking of references based on this network. For references, the number of citations is an important index of ranking, that is, the priori knowledge—α in our model mentioned in Section 2.

Table 8 shows the rankings of influence regardless of α (here, α means the numbers of citations) and that regarding α in different β.

Table 8 The ranking of 16 references

α value	The information of references	β=0.05		β=0.02	
		without α	with α	without α	with α
4531	On Random Graphs	8	5	8	5
13250	Statistical mechanics of complex networks	10	3	10	3
1956	Power and Centrality: A family of measures	3	9	3	7
18843	Emergence of scaling in random networks	1	1	1	2
25	Identifying sets of key players in a network	16	15	16	15
763	Models of core/periphery structures	7	11	7	11
16	On properties of a well-known graph, or, What is your Ramsey number?	15	16	15	16
1246	Navigation in a small world	2	9	2	9
1523	Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality	5	7	5	8
2748	The structure of scientific collaboration networks	4	6	4	6
10616	The structure and function of complex networks	11	4	11	4
680	Networks, influence, and public opinion formation	14	12	14	12
835	Identity and search in social networks	6	10	6	10
21688	Collective dynamics of `small-world' networks	9	2	9	1
33	Statistical models for social networks	13	14	13	14
498	Social network thresholds in the diffusion of innovations	12	13	12	13

We can reach the conclusion that, with the consideration of α value, the ranking of references has changed, and it fits more to the actual influence ranking of references. Therefore, α is of great accuracy and necessity in our model. When applying to networks of other fields, we can assign different α values to nodes according to the message of characteristic indices of this fields, to represent the characteristic indices of this node. The citation of α makes our model more accurate and more applicable.

Similarly, we can also collect universities, departments, or journals indicators and then use our model to calculate the influence ranking. The method is summarized as follows (for example: universities): First, we collect matrix G of academic partnerships between universities, the more collaboration between the two universities, the greater the value of the element of the position in the matrix. We regard total SCI papers published statistics for each university in recent years as the α value. Thus, we can obtain influential university rankings with different β. Similarly, the approach can be utilized in other areas.

5 Stability parameter—β

All of the networks in reality are not completed, they are all come from samples of the whole actual networks, just like the examples of the sixteen papers which are provided in the problem. For all samples of the whole actual networks, according to the different rate of samples, we can adjust the stability parameter β to improve rank preservation.

We have taken advantage of computers to create 500 directed emulational networks which own the number of n=200 nodes. Each of the emulational network will get an adjacent matrix. In condition of β=0.005 and β=0.02, every completed emulational network corresponding a seniority of effects. However, we extract 50%，60%，70%，80%，90%，95% side information of completed emulational networks one by one to constitute sampling network, to statistic the seniority of effects under these two different β.

The Kendall Tau [Yongli Li, 2013] in **Appendix 1**, shows the order that we obtained from sampling information and the rank-correlation coefficient from real ranking under this condition. All Kendall Tau information as shown below (**Figure 4**):

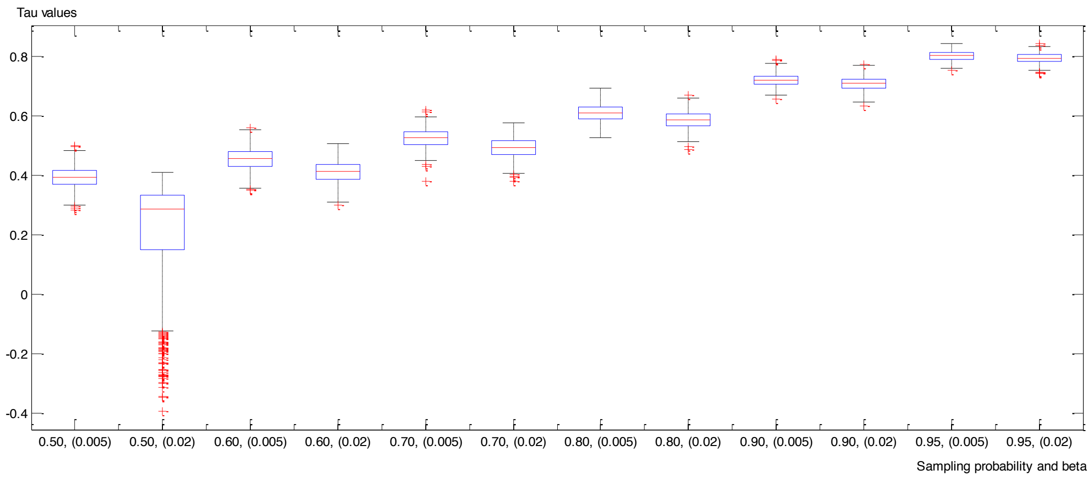

Figure 4 The statistics of Kendall Tau

Abscissa shows the sampling rate of sampling network edge information, and β are showed in parentheses.Ordinate shows all Kendall Tau statistics in this abscissa in 500 simulated networks.

We can see that with the increase of the sampling rate, Kendall Tau keep increasing. That means, for incomplete information network, our model can still get an influence ranking with high rank-correlation coefficient influence ranking. For the same sampling rate, the smaller β values, the higher the Kendall Tau. We can conclude that for incomplete information network, our model can further improve the ranking of the rank-correlation coefficient.

In our model, β value's increasing enhance the edge contact, then *Inf* values will be obtained. It emphasizes the points in the network's ability to effect each other. However, if we select a different β, it expresses a tendency to the influence evaluation, that is, whether the emphasis on network adjacency relationships.

The advantage of our model is to select different stability parameter β in various networks, then to improve the rank-correlation coefficient. However, this is also one of the shortcomings of our model. For the unknown network information, β is difficult to select, we need to go through a lot of historical data statistics to determine the appropriate β values. To select β in different fields, which requires further study.

6 Apply Our Model to Other Databases

We have found a database [D. E. Knuth , 1993], which is a network of coappearances of characters in Victor Hugo's novel "Les Miserables". Nodes represent characters as indicated by the labels and edges connect any pair of characters that appear in the same chapter of the book. The values on the edges are the number of such coappearances. We have established a network of characters, as shown in **Figure 5**. Apply our model to this network and the calculated result is shown in **Table 9**.

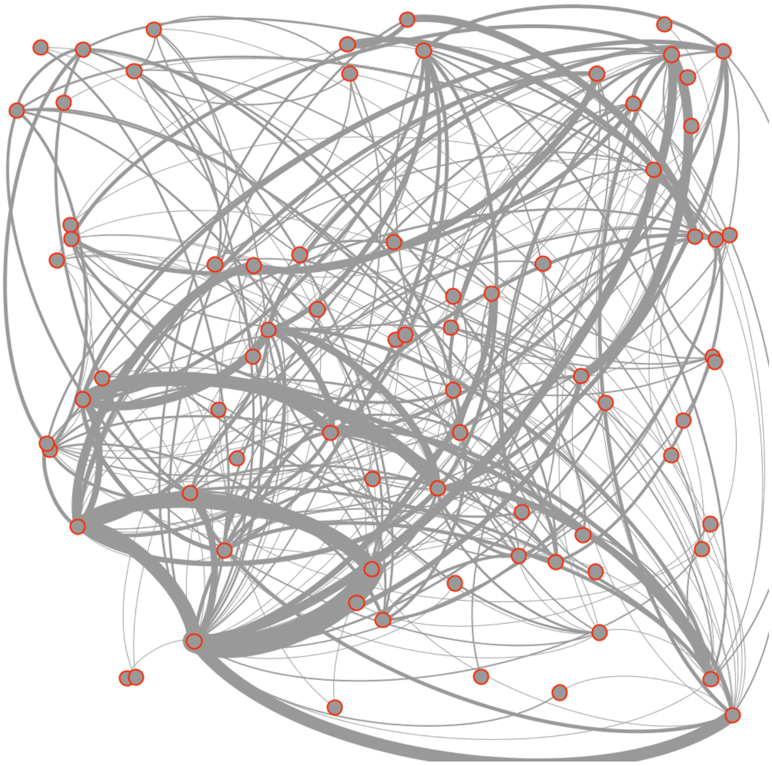

Figure 5 The network of characters in "Les Miserables"

Table 9 Top 10 characters

Ranking	β=0.005	β=0.01
1	Valjean	Valjean
2	Marius	Marius
3	Enjolras	Cosette
4	Courfeyrac	Enjolras
5	Cosette	Courfeyrac
6	Combeferre	Combeferre
7	Bossuet	Bossuet

8	Thenardier	Javert
9	Gavroche	Thenardier
10	Javert	Gavroche

This is a model applied in the undirected weighted graph. Our calculations can fully reflect the influence ranking of the role network in the "Les Miserables": Take the top10 people in different β, and they are all important roles in "Les Miserables", there is no denying that the hero Valjean is ranked first.

7 Conclusion

We started from the co-authored networks composed of Erdos1, and constructed our model. We then calculated node influences accordingly and analyzed by which method can we improve node influence in the networks to the largest extent. The main idea for establishing this model is that the influence of a node means the change of overall influence in the network after the node has been removed and we have considered the entire network. According to our design, for a node, building a connection with node of the biggest boncich centrality can maximize its influence in the network. Then, we analyzed the effects α and β values have on network influence models, while α value determined the base value of the nodes and the β value reflects the linkage's strength and controlling the model's robust especially considering the network's incompleteness. We also analyzed the applications of our model in different types of figures and in different fields, with a positive outcome.

7.1 Strengths

- Our model has considered the influence of the whole network to single node, which is more accurate than the ordinary analysis as we consider the changes of reputation after the node is deleted.
- We have taken into consideration the applications in different networks, different fields, and different data scales. By using α and β, we have made our model more widely applicable.

7.2 Weaknesses

- We need to consider β value of the nodes, which is the value of being connected to someone decays with distance. But in actual application, the values of β in different fields and networks usually require previous experiences, the acquisition of which needs massive data analysis of existing networks.
- As we have considered the entire network, when relatively large networks need calculating, a great deal of calculation is needed, and the acquisition of result will be time-consuming.

References

Ballester, C., Calvó - Armengol, A., & Zenou, Y. (2006). Who's who in networks. wanted: the key player. *Econometrica*, *74*(5), 1403-1417.

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, 1170-1182.

D. E. Knuth. The Stanford GraphBase: A Platform for Combinatorial Computing, Addison-Wesley, Reading, MA (1993).

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39-43.

Seeley, J. R. (1949). The net of reciprocal influence; a problem in treating sociometric data. *Canadian Journal of Psychology*, *3*(4), 234.

Yongli LI, Peng LUO and Chong WU. (2012). A new network node similarity measure method and its applications.

Yongli Li, Chong Wu, Gülşah Gülen and Jing WANG. (2013). Network Structure, Perception Level and Participants' Welfares.

Appendix 1 (The definition of Kendall Tau τ)

Given two vectors $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$, and then C_{ij} is defined as

$$C_{ij}=\begin{cases}1 & \text{If } (a_i<a_j \text{ and } b_i<b_j) \text{ or } (a_i>a_j \text{ and } b_i>b_j) \text{ or } (a_i=a_j \text{ and } b_i=b_j),\\ 0.5 & \text{If } (a_i=a_j \text{ and } b_i\neq b_j) \text{ or } (a_i\neq a_j \text{ and } b_i=b_j),\\ 0 & \text{Other cases.}\end{cases}$$

The number of concordant pairs is $C=\sum_{i=1}^n\sum_{j=i+1}^n C_{ij}$. Based on the above definition, Kendall's Tau rank-correlation coefficient is $\tau=(4C/(n(n-1)))-1$, with the following properties:

(1) If the two vectors imply the same ranking, then $\tau=1$. Otherwise, if the ranking implied by one vector is the reverse of the other, then $\tau=-1$.

(2) For all the cases, τ lies between −1 and 1, and the larger the value is, the more agreement between the rankings implied by the two vectors are.