Summary

In this paper, we propose the co-author network model and the complex citation network model by developing effective influence measure algorithms to respectively solve the problems within two different kinds of networks. Specifically, we totally construct three models focusing on and providing solutions to the following issues:

- Develop co-author network model to determine who have significant influence and publish important work within the network.
- Develop citation network to determine the influential papers, individuals of network science researchers and journals.

The co-author network can be divided into two sub-models, we build **Model I "Transfer and Feedback Iteration Model"** to analyze how the impact is transferred and determined through co-authorships between two researchers within the network. Firstly, we define the transfer factor and initial impact value of each co-author as the number of joint publications with Erdös and the transfer factor. And then, we use the model to calculate the final impact value after multiple times of iterations and the impact value is an index to determine the influential co-authors with the network.

Then, we build **Model II: "M-Nearest Neighborhoods Model"**, of which the foundation is generally similar to Model I. We change the way to describe the transferred impact between co-authors, that is, we replace the way of iteration to a simpler measure that the larger path length is, the smaller impact that will be transferred through co-authorships relations.

We introduce **Model III "Complex Citation Network Model"** to compare the significance of research papers in a **Complex Citation Network**. Three kinds of nodes (paper, researcher and journal) are included in the complex network which is more difficult to measure. The impact value can be only transferred from the paper to the cited paper. The number of citations is the key factor to judge whether a paper is influential in a certain academic field. Therefore we are also required to find available data of the papers, researchers and journals. Besides, we consider the randomness and non-linear property of the transfer function.

Furthermore, we applied Model III to the **Movie-Actor network** in real world and calculate the top 5 influential Chinese movie actors. Validation has been done for the result of our model, and it is proved to be reasonable and practical.

In addition, sensitivity analysis and solution of each model is provided. Changes of model parameters will impact the model results and we are supposed to analyze these parameters and how they affect the results, which can be very helpful to our in-depth model.

Finally we discuss the science, understanding and utility of modeling influence and impact within networks. And we introduce whether such a network model can be applied to our lives especially the aspect about the influence calculation. The model can help us to understand the reality of life and that how the elements in a certain network can influence each other.

Let's Network!

Kun Huang, Zuoyue Li, Binbin Chen
Zhejiang University, China

1. Introduction

Considering what we have learned from Erdös's academic life story, the impact of extensive connections through co-authorships can never be ignored. As one of the most famous mathematicians in 20th century, Paul Erdös had over 500 co-authors and published over 1400 technical research papers, thus, it has already existed an amazingly broad and robust co-author network after figuring out all of the relationships between Erdös and his co-authors. Besides, Erdös's role as a collaborator was so significant in the field of mathematics that the important influence within the network can represent researchers' academic capacity in the field. In order to measure such influence as Erdös produced, we are required to build network-based evaluation models that use co-author or citation data to determine impact factor of researchers, publication, and journals. To be more specific, our team's goal is to analyze influence and impact in research networks and then apply our models to other fields in the society. The structure of our essay is as follows:

Task 1: Co-author Network & Properties

- Set relationship adjacency matrix and plot the network
- Find and calculate properties

Task 2: Models of the Co-author Network

- Develop influence measure of the network
  - ✓ Model I: Transfer and Feedback Iteration Model (TFIM)
  - ✓ Model II: M-Nearest Neighborhoods Model (MNNM)
- Sensitivity analysis and solution

Task 3: Model of Complex Citation Network

- Search data of papers, authors and journals listed
- Develop influence measure of the network
  - ✓ Model III: Complex Citation Network Model
- Sensitivity analysis and solution

Task 4: Implement Algorithm on Different Field

- Determine the field (movie) and search data
- Apply the Model III to the field and give solution

Task 5: Discuss the Utility of Modeling Impact

- Apply the model to real world

2. Co-author Network

2.1 Build the Co-author Network

We build a co-author network of exact 511 researchers from the file Erdos1, and the links of each two nodes represent the co-authorships of researchers who coauthored a paper with Erdös, but do not include Erdös. Although we have extracted the raw data to a relatively smaller size, but the structure of the network is still so complex for the quantity of nodes and edges within the network are large that they are more likely to overlap in the network graph and it cannot be visible and meaningful.

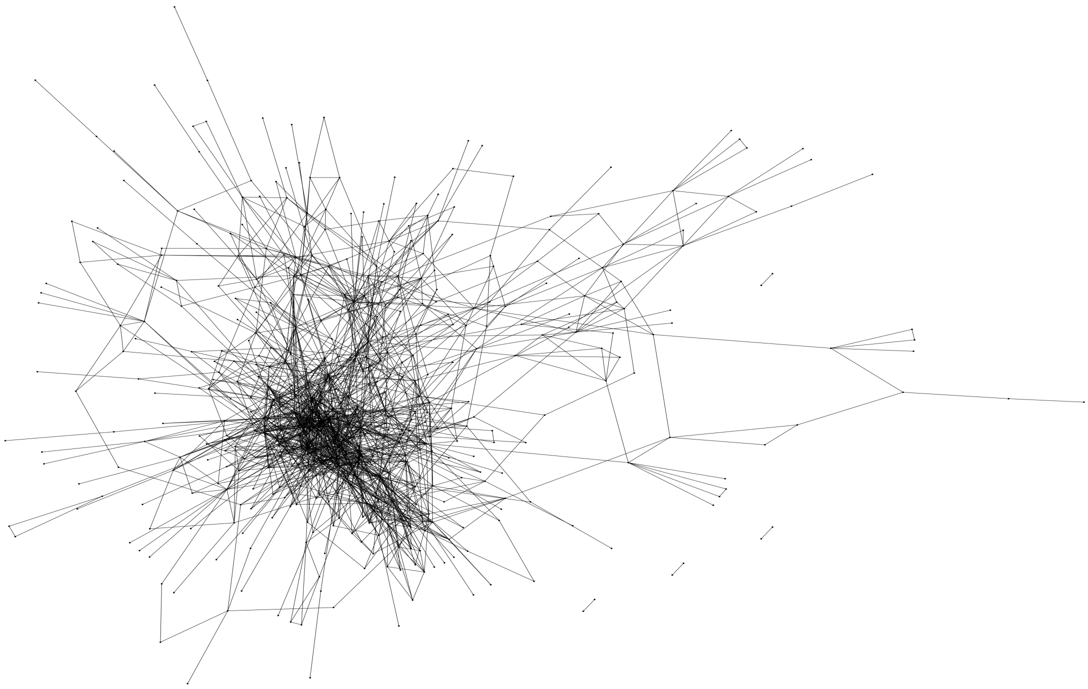


Figure 1. The Co-author Network

Considering the above situation we have discussed, it's necessary to limit the size of the network to analyze network's properties and to calibrate the influence measurement algorithm. We choose a set of nodes matching the requirement that number of joint publications with Erdös is larger than X_1, or the number of links to other co-authors is larger than X_2. When we set $X_1 = 20, X_2 = 25$, the number of nodes in the new set is 29. And the co-author network graph with the nodes in the new set is as below:

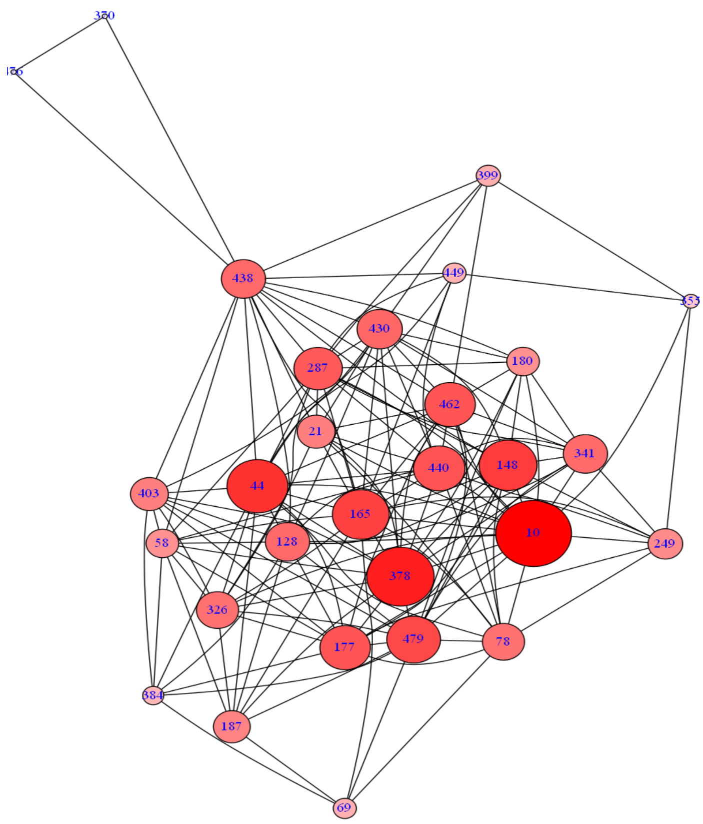


Figure 2. The Simplified Co-author Network

2.2 Main Properties of the Network

2.2.1 Symmetry

The first main property that we can obtained from the network graph is that almost of the relationships between each two nodes are symmetrical, that is, if a researcher is the co-author of another one then the relationship is also true in the opposite direction.

Due to the symmetry properties of the network, we simplify the directed graph to undirected graph in order to make the network graph more visible, and in addition, the algorithm based on the undirected graph is also much easier.

2.2.2 Density

The density D of a network is defined as a ratio of the number of edges E to the number of possible edges, given by the binomial coefficient $\binom{N}{2}$, giving $D = \frac{2E}{N(N-1)}$. The density of the network is 0.0126, indicating that it is a sparse network.

2.2.3 Clustering Coefficient

In graph theory, a clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together.

And in this part, we use the global version to give an overall indication of the clustering in the network. The transitivity ratio is a related concept. It is defined as:

$$C = \frac{3 \times number\ of\ triangles}{number\ of\ connected\ triples\ of\ vertices} = \frac{number\ of\ closed\ triplets}{number\ of\ connected\ triples\ of\ vertices}$$

The calculation result $C = 0.6676$.

2.2.4 Centrality

Eigenvector centrality is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

Another closely related centrality measure is Katz centrality. It is a generalization of degree centrality. Degree centrality measures the number of direct neighbors, and Katz centrality measures the number of all nodes that can be connected through a path, while the contributions of distant nodes are penalized.

The nodes with the top 5 Katz centrality and Eigenvector centrality are as follow.

Table 1. Top 5 Eigenvector Centrality and Katz Centrality

Rank	Node No.	Eigenvector centrality	Node No.	Katz centrality
1	10	0.2596	10	2931.0
2	378	0.2339	378	2573.8
3	44	0.2086	44	2393.0
4	165	0.2038	165	2387.1
5	148	0.2010	148	2280.5

3. Models of the Co-author Network

In this part, we will introduce our impact-measuring model of the co-author network aiming to find out who has the most significant influence within the network. Firstly, the impact value of each node is initialized as the number of joint publications with Erdös, because we consider the researchers who have ever collaborated more frequently with Erdös are much more likely to make significant influence within the network. If we do not consider Erdös 's academic impact to his co-authors in the network, in other word, we only care about the co-author's significant influence within the network; we initialize the impact value as a constant value, in our paper we give the constant value of 1. And then, three different algorithms are given to calculate the nodes' final impact value. Respectively, we put forward three models as below:

- Model I: Transfer and Feedback Iteration Model (TFIM)
- Model II: M-Nearest Neighborhoods Model (MNNM)

3.1 Model Assumption

- We do not consider the times of collaborations between Erdös' co-authors to their co-authors.
- The number of joint publications with Erdös can influence the co-authors' impact within the network.
- The closeness of co-authorships between each two researches is the same.

3.2 Symbols and Definitions

Notations below are used in either Model I or Model II.

Table 2. Notations

Variables	Description
R_{ij}	relation adjacency matrix (if node j is one of the co-authors of node i, then $R_{ij} = 1$, else $R_{ij} = 0$)
N	the number of nodes (co-authors of Paul Erdös)
I_i	The impact value of node i
Y_i	the first cooperation year with Paul Erdös of node i
D_i	the life-or-death situation of node i (if node i is known to be deceased, then $D_i = 1$, else $D_i = 0$)
T_i	the number of joint publications with Erdös of node i

3.3 Model I: Transfer and Feedback Iteration Model (TFIM)

3.3.1 The Foundation of Model I

As Erdös is the most influential author in the network, we have already used the number of joint publications with Erdös to show the initial significant influence of researchers. Besides, the number of their co-authors should also be considered as a key factor to determine the most influential co-author in the network, as the both sides of each co-authorship relationship can benefit from each other through the cooperation, and thus, we can conclude that researchers who have more co-authors would get a higher impact value. And we define a **"transfer factor"** as the influence from co-author to author. If one's co-author has greater impact value, then the transfer from the co-author to author becomes larger, that is, researchers will have a more important work if he has an influential co-author. Similarly, we define **"feedback factor"** as the influence from author to co-author and that means a significant author can bring extra impact value to his co-authors but the value is lower than "transfer factor".

There is another necessary factor **"damping factor"** that should be involved in the TFIM, because without damping factor, the impact value would be divergent. In order to get a convergent result, we try to apply the method used in the "Fourier Transform to Laplace Transform", and bring the **"damping factor"** to TFIM to decrease the two factors by exponential order after each time of iteration. In addition, "year factor" is also included in the model, which is the first time of the joint publication with Erdös. When the year interval is larger between two researchers, the contributions through co-authorships to both sides become smaller, that is, the value of year factor is smaller.

- Transfer factor, g, can be a nonlinear function depends on authors.
- Feedback factor, h, also can be a nonlinear function depends on co-authors.
- Damping factor, d, $d \in (1,+\infty)$, after each time of iteration, g and h will turn to $\frac{1}{d}$ of its former value (decreased).
- Year factor, y_{ij}, is a matrix, indicates the influence degree to mutual contribution due to the difference of year intervals. Each element within the matrix is calculated as below:

$$y_{ij} = \frac{1}{\left|\frac{Y_i - Y_j}{\max(Y) - \min(Y)}\right| + 1}$$

- Corrected relation adjacency matrix, P_{ij}, each element within the matrix is calculated as below:

$$P_{ij} = R_{ij} \cdot y_{ij}$$

Formula of the impact value after each time of iteration is as follow:

$$I_{k+1} = I_k + \frac{g}{d^k} \cdot P \times I_k + \frac{h}{d^k} \cdot P^T \times I_k, I_0 = T, k \in [0,+\infty)$$

where I_k means the column vector of impact value of all nodes after the k^{th} iteration (different from I_i in 4.2), R represents the relation adjacency matrix and R^T is its transposition, I_0 means the column vector of initial impact value that should be given before iteration. If we consider Erdös's impact to co-authors in the network, we initialize the impact value as T which is the column vector of the number of joint publications with Erdös of all nodes. Otherwise, if we only consider the structure of the network and ignore the impact of Erdös, T can be defined as a constant column vector with element value of 1.

We can obtain the results from a certain times of iteration, which is converged to the result when $k \to +\infty$.

3.3.2 Sensitivity Analysis

The results of Model I vary with the change of the parameters' values. Thus, we make the sensitivity analysis in this part to test the robustness of the results of the model in the presence of uncertainty and develop the measures when taking all of the factors into consideration, and then to determine who in the network has significant influence and who has published important works. Considering one of the main properties of the network is symmetry, as only the links between two certain nodes are unsymmetrical, the feedback factor h will almost not influence the results, so we define $h = 0$. In addition, to simplify the calculation, the nonlinear transfer factor g is set to a constant in our analysis. As a result, the directed graph is simplified to undirected graph.

Firstly, we use Matlab to randomly assign the parameters' value within the defined threshold. For example, we get random numbers from the domain $g \in (0,1)$ and $d \in (1,30)$. And then, we analyze the properties of the top three co-authors, and find out that the top three co-authors are always from the set of co-authors which is made up with the following eight co-authors. Thus, our sensitivity analysis is based on the changes of impact value of the eight co-authors when the parameters are changing.

Table 3. Top 8 influential co-authors

No.	Name
10	ALON, NOGA M.
44	BOLLOBAS, BELA
128	FAUDREE, RALPH JASPER, JR.
180	HAJNAL, ANDRAS
378	RODL, VOJTECH
399	SARKOZY, ANDRAS
430	SIMONOVITS, MIKLOS
438	SOS, VERA TURAN

We focus on analyzing the influence of uncertain inputs of transfer factor g and damping factor d to the outputs of the impact value I. Transfer factor reflects the contribution from co-author to author, when the value of transfer factor becomes larger, it means that the contributions from co-authors or the number of co-authors have a greater impact on the results, otherwise, the initial value of impact value will have a greater impact on the final result. The other input is damping factor, which will influence the total times of iteration. Larger damping factor results in a smaller number of times of iteration, which also indicates that the contributions from co-authors or the number of co-authors have a smaller impact, and the initial value of impact value is more important.

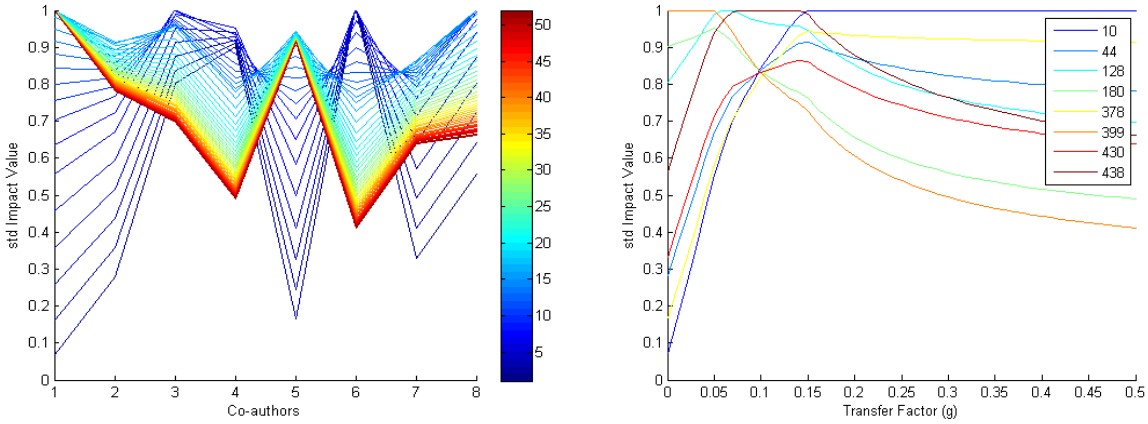
For the reason that when the parameter is different, the absolute value of impact value is not comparable to each other, we define $Istd$ (standard impact value) to compare the results under different conditions.

$$Istd = \frac{I - \min(I)}{\max(I) - \min(I)}$$

Input One: Transfer Factor

We vary transfer factor from 0 to 0.5 and the step is 0.01. Graphs on the left represent the standard impact value of each of the eight co-authors, while the graphs on the right represent the relationship between the transfer factor to impact value of each co-author.

The properties of graphs confirm our forecast that when transfer factor becomes larger, the contributions from co-authors or the number of co-authors have a greater impact on the results. The co-authors whose number is 10, 378 have similar academic experience that the number of joint publications is small but they have relatively more co-authors compared with other co-authors of Erdös. Thus, the final impact value of the co-authors who have more co-authors would be more sensitive to the changes of transfer factor.



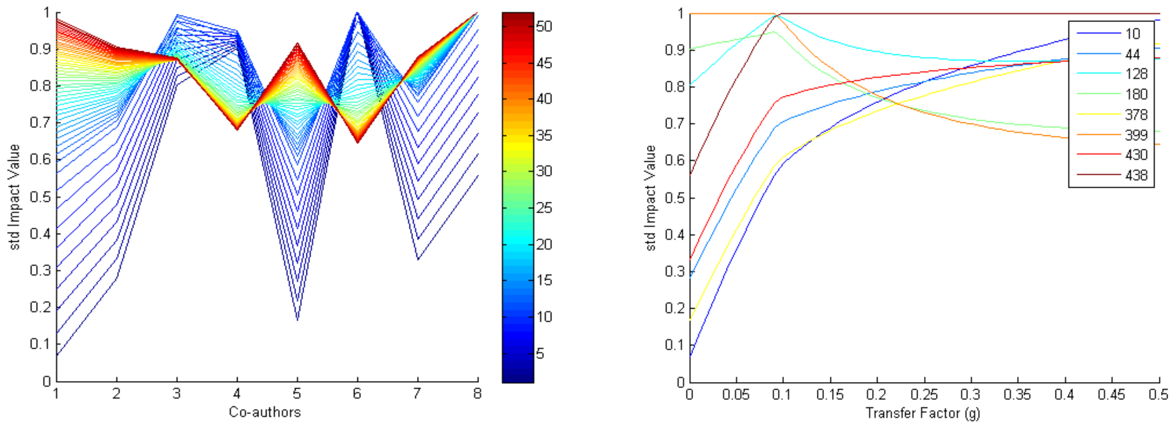
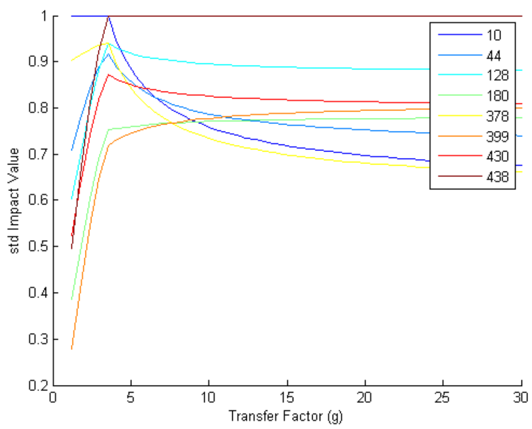


Figure 3. Sensitivity Analysis of g When $d = e$

Figure 4. Sensitivity Analysis of g When $d = 10$

Input Two: Damping Factor



We vary damping factor from 1.2 to 30.0 and the step is 0.06.

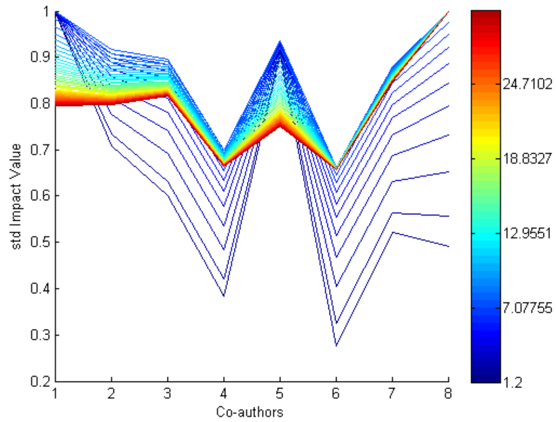


Figure 5. Sensitivity Analysis of d When $g = 0.2$

When the damping factor decreases, the impact value will also decrease, but, for the co-authors that have more co-authors, the degree of reduction will be smaller such as the co-authors number 10 and 378.

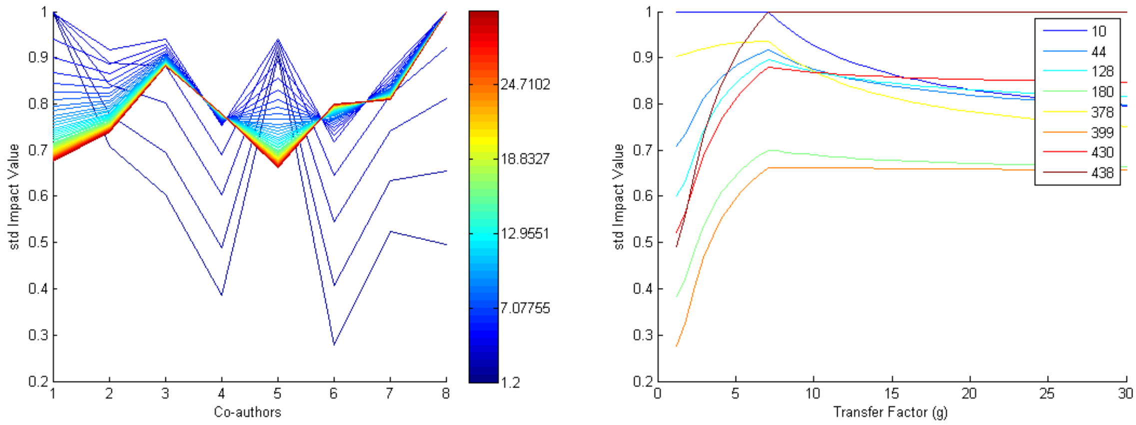


Figure 6. Sensitivity Analysis of d When $g = 0.4$

When the damping factor increases, the impact value will increase, and for the co-authors that have more number of joint publications with Erdös, the degree of increase will be larger such as the co-authors numbered 428,128 and 180, their times of collaboration with Erdös is 35, 50 and 56.

If we only consider the structure of the network and ignore the impact of Erdös, we find out that changes of parameters make few influence to the final results, we make this conclusion after the analysis of the table below:

Table 4: Sensitivity Analysis When Ignoring Initial Value

	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
	10	10	10	10	10	10	10	10	10	10
2	378	378	378	378	378	378	378	378	378	378
	165	165	165	165	165	165	165	165	165	165
	10	10	10	10	10	10	10	10	10	10
e	165	165	165	165	378	378	378	378	378	378
	378	378	378	378	165	165	165	165	165	165
5	10	10	10	10	10	10	10	10	10	10

	165	165	165	165	165	165	165	165	165	165
	378	378	378	378	378	378	378	378	378	378
	10	10	10	10	10	10	10	10	10	10
10	165	165	165	165	165	165	165	165	165	165
	378	378	378	378	378	378	378	378	378	378

The table lists three top co-authors who have significant influence within the network under the condition that the value of transfer factor is changing from 0.05 to 0.5 while the damping factor is set to the value of 2, e, 5 and 10. Besides, the results show that the co-authors ranked 4th to 6th is 148 FUREDI ZOLTAN, 187 HARARY FRANK and 479 TUZA ZSOLT.

3.3.3 Results and Discussions

In order to determine who in Erdo1 network has significant influence within the network, we do not consider the academic impact of Erdös. According to the result shown on Table 4 the top 3 researches are 10 ALON NOGA M, 165 GRAHAM RONALD LEWIS, 378 RODL VOJTECH.

In addition, we also give Eigenvector Centrality and Kats Centrality[1] when we consider the year factor. The results are as below.

Table 5. Comparison When Adding Year Factor

Model I: TFIM	Eigenvector centrality	Kats centrality
10 ALON NOGA M	10 ALON NOGA M	10 ALON NOGA M
165 GRAHAM RONALD L	378 RODL VOJTECH	378 RODL VOJTECH
378 RODL VOJTECH	148 FUREDI ZOLTAN	165 GRAHAM RONALD L
148 FUREDI ZOLTAN	165 GRAHAM RONALD L	148 FUREDI ZOLTAN
187 HARARY FRANK	479 TUZA ZSOLT	479 TUZA ZSOLT
479 TUZA ZSOLT	44 BOLLOBAS BELA	44 BOLLOBAS BELA

The results of the three measures are generally alike, and the existing two measures show almost the same results. Compared to the other two measures, the results of our model show the six co-authors in a different sort, thus, we conclude that our model still need further developments in the future.

If we consider Erdös's impact, then according to the analysis of the previous part in the paper, there are two groups of co-authors that have different characteristics but both have published important works and. The first group of co-authors keeps stable and large impact value when the transfer factor varies such as 128 FAUDREE RALPH JASPER JR and 180 HAJNAL ANDRAS while the other group show rather high impact when the transfer factor is in a certain range, but out of the range, their impact value will decrease suddenly, besides, they can keep stable and high impact when the damping factor varies such as 378 RODL VOJTECH and 10 ALON NOGA M.

3.4 Model II: M-Nearest Neighborhoods Model (MNNM)

3.4.1 The Foundation of Model II

Different from Model I: TFIM, Model II use an algorithm which has resemblance to Gaussian distribution that farther node contributes less impact to the central node while closer node delivers more impact. We define a **"maximum path length"** to limit the longest distance. We do not consider the nodes whose distance to the central is longer

[1] http://en.wikipedia.org/wiki/Network_science

than "maximum path length". Otherwise, the rest of node are ought to be given a weight, forming an M-Dimension column vector. Besides, we have to know adjacency matrixes under the condition of certain length of path, in which each of the elements indicate the length of the path between two nodes. And the adjacency matrix can be attained from the initial adjacency matrix.

- Maximum path length, M, regarding the network as a graph and the path length between two neighboring nodes equals one.
- Path length, l.
- M-Dimension column vector, W, $W_l \in [0,1]$, $W_{l+1} < W_l$.
- The k^{th} power of relation adjacency matrix, R^k, indicates the number of paths. And the paths must meet all the two requirements shown as below.
  - ✓ Path of two nodes, and the path length must be k .
  - ✓ All elements in the position of main diagonal of the adjacency matrix must be zero.

Formula of the impact value is as follow:

$$I = T + \sum_{l=1}^{M} W_l \cdot R^l \times T$$

where I and T has the same meaning as them in 3.2.

3.4.2 Sensitivity Analysis

Similar to the analysis of Model I, we make the sensitivity analysis of the weight vector W. We choose the vector W, of which the elements form a sequence that is a geometric progression with common ratio $q \in (0,1)$, and the value of first element should equal to q. When the value of q becomes larger, the connections to the points around the node are closer.

Using the same methods we mentioned in section 3.3.2, we elected 10 nodes this time, adding three nodes extra as below:

Table 6. Added Three Nodes

No.	Name
148	FUREDI ZOLTAN
165	GRAHAM RONALD LEWIS
403	SCHELP RICHARD H.

Input One: Transfer Factor

We vary transfer factor from 0.03 to 0.06 and the step is 0.002, and set $M = 10$.

Along with the change of transfer factor q in such a small range $q \in (0.03, 0.06)$, the impact value of each co-author is not stable.

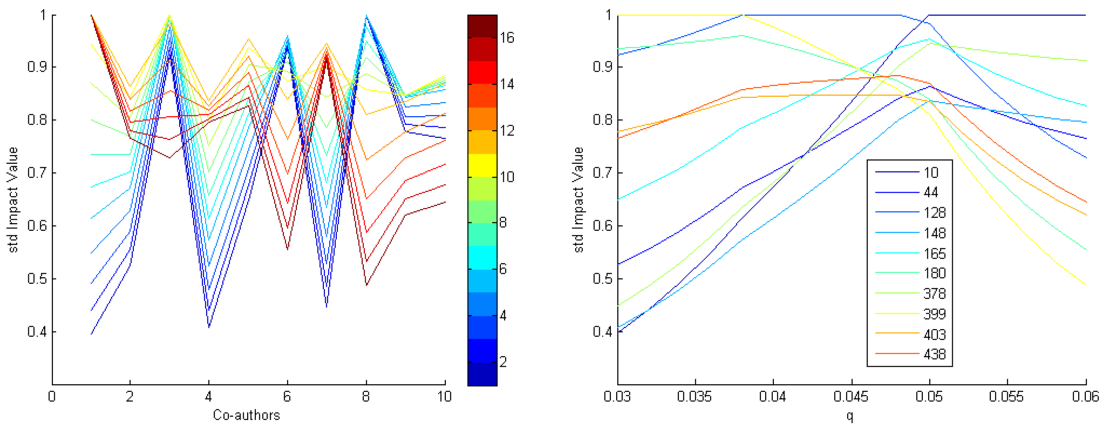


Figure 7. Sensitivity Analysis of q when $M = 10$

Again, when we only consider the structure of the network and ignore the impact of Erdös, the results keep stable.

3.4.3 Results and Discussions

When we ignore the impact of Erdös, the co-authors ranked one to five are as below. Comparing to the results of Eigenvector centrality and Katz centrality we have shown on the Table 5, the result of Model II is as the same as Katz centrality.

When we take the impact of Erdös into account, if $q \geq 0.07$, the co-authors ranked one to six are the same as we ignore the impact of Erdös. However, if $q \leq 0.03$, the co-authors ranked one to five are as below:

Table 7. Comparison of the Results in Three Different Conditions

Rank	consider the initial impact & $q \leq 0.03$	consider the initial impact & $q \geq 0.07$ **or** ignore the initial impact
1	399 ALON NOGA M.	10 SARKOZY ANDRAS
2	180 RODL VOJTECH	378 HAJNAL, ANDRAS
3	128 GRAHAM RONALD LEWIS	165 FAUDREE RALPH JASPER JR
4	403 FUREDI ZOLTAN	148 SCHELP, RICHARD H.
5	438 TUZA ZSOLT	479 SOS, VERA TURAN

3.5 Strengths and Weaknesses

Strengths

- Our model takes into account the nodes' connections and interactions including transfer factor and feedback factor.
- The influence of year factor to the iteration process is taken into consideration.
- We use several parameters to control the inputs and make our model more flexible to reach multiple objectives.
- Our model involves the impact of Erdös to the network.
- Damping factor shows that the influence of information is going to decrease after times of iteration.

Weaknesses

- We cannot set an effective domain for the parameters, which can limit the changes of results.
- Although we have considered the year factor, but without further research, we still cannot make sure the relationships between publication year and impact value.
- We simply initialize the impact value as the number of joint publications with Erdös, but we are uncertain whether or not it's efficient.
- We lack real data to testify our model so we can only give a result of simulation.
- The factors cannot precisely describe the real situation.

4. Complex Citation Network

In the context of network theory, a **complex network** is a graph (network) with non-trivial topological features—features that do not occur in simple networks such as lattices or random graphs but often occur in real graphs.[1] The citation network displays substantial non-trivial topological features, with patterns of connection between the elements including paper, author and journal are neither purely regular.

4.1 Building the Complex Citation Network

Our model consists in three groups of elements, paper, author and journal. As paper can only refer to the previous paper, the citation relationships between papers can form a directed graph, and there are no circles in the graph, that is, it is a directed acyclic graph (DAG), so, the topological ordering is possible in the graph.[2] Papers are the main elements in the citation network, and the graph also shows the relationships between papers to authors, and papers to journals, which form undirected sub-graphs.

Within the network, each paper can have multiple relationships with authors, but with only one journal. Besides, each author or journal can have multiple relationships with papers. The network graph using the example given in the problem is as below:

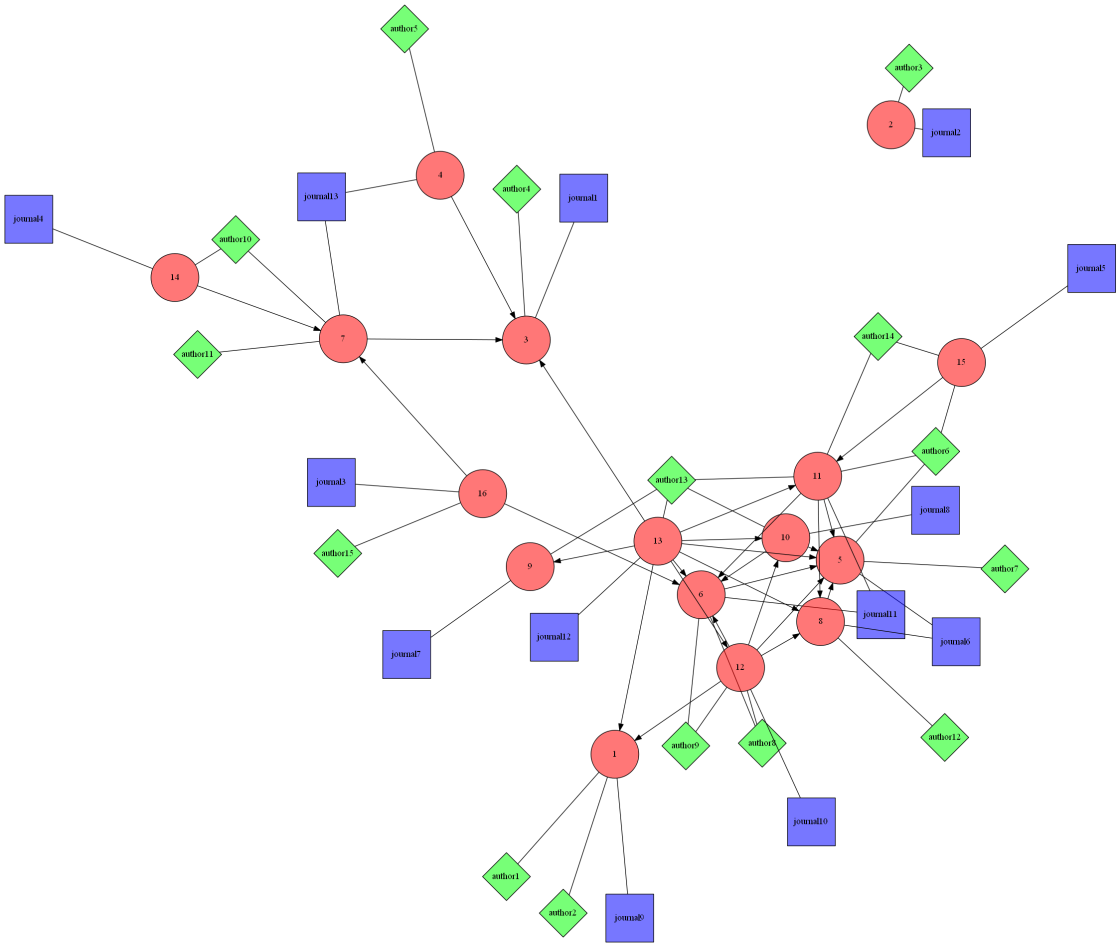


Figure 8. The Complex Citation Network

4.2 Main Properties of the Network

Most social, biological, and technological networks display substantial non-trivial topological features, with patterns of connection between their elements that are neither purely regular nor purely random. Such features include a heavy tail in the degree dis-

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

2 http://en.wikipedia.org/wiki/Topological_sorting

tribution, a high clustering coefficient.[1]

5. Model of Complex Citation Network

5.1 Model Assumption

- The number of total citations of the paper can be regarded as the initial impact value of papers.
- The influence of authors in the network is depending on the influence of papers.
- The influence of journals in the network is depending on the influence of papers.

5.2 Symbols and Definitions

Table 8. Notations

Variables	Description
P	the total number of papers
A	the total number of authors
J	the total number of journals
R_{ij}	paper i refers to paper j
S_{ij}	author i is the author of paper j
T_{ij}	paper j is published in the journal i
I_i, H_i, K_i	the impact value of paper i, author i and journal i in the whole academia
I'_i, H'_i, K'_i	the impact value of paper i, author I and journal i inside the network
C_i	citations of paper i (to present I_i)
D_i	the serial number of author of paper i
E_i	the serial number of journal of paper i
W	transfer weights between papers

5.3 The foundation of Model

The foundation of the complex citation model is similar to the co-author network model. But as a directed acyclic graph (DAG), the impact influence can be only transferred from the paper to the cited paper. And the transfer weights W, which has the same meaning as the transfer factor that is mentioned in Model I.

Firstly, we should initialize the impact value of I_0. And the impact value of a paper is mainly influenced by the following three factors.

- The number of citations is the key factor to judge whether a paper is influential in a certain academic field.
- The impact of the author is another way to tell the influence of paper.
- A more influential paper is usually published in a more authoritative journal.

Then, the impact value is the function of the influence of paper, author and journal.

$$I'_i = f(C_i, H_{D_i}, K_{E_i})$$

[1] http://en.wikipedia.org/wiki/Complex_network

where f can be either a nonlinear function or a linear function. The transfer path starts at the top of the hierarchy, which is the node of latest paper. The impact value of the paper cited would equal to its initial value plus the value transferred from other paper, which can be influenced by the transfer weight W and only if the impact value of nodes in previous layer have been calculated, can the nodes in the next layer start the same calculation process. As there is no iteration in this model, we do not consider the feedback factor and damping factor here, but, we still take the year factor into account. The influence relation between the impact value and year factor has changed, as in this model larger year interval of the two papers shows the long-time impact of the paper cited which can be a symbol of high impact of one influential paper.

The algorithm steps of our model are as follows:

- Step1: Set all the status of the nodes as "undecided".
- Step2: Find the undecided node i with the smallest Y_i, make its influence to the network, and then set the status of node i as "decided".
- Step3: if all the status of the nodes is "decided", we have completed the algorithm, else we should go back to step 2.

After we finished the calculation, we standardize the impact value to conform a standard. Besides to develop measures to determine the most influential paper in network science, this model can also be applied to determine the influence measure of an individual network researcher and the impact of a specific university, department, or a journal in network science. In our paper, we use our model to analyze the influence of a journal.

We collect the data of the citations C_i of each paper from Google Scholar and Microsoft Academic Search, and the influence of journal K_i that is evaluated by the impact factor from Impact Factor Search[1], and the influence of authors H_i using H-factor.

If a paper has more than one author, then we should use formulation as below to present the relation.

$$I_i' = f(C, \alpha H_a + \beta H_b, K)$$

where a and b are authors of paper i, α and β show their individual contribution to the impact value of paper i.

5.4 Sensitivity Analysis

In order to simplify the calculation, the algorithm are changed as below.

To calculate the influence of journals: We use average of the impact value of papers that are published in the journal, as we think the influence of papers published is more reasonable than the number of papers when we evaluate the influence of journals.

To calculate the influence of author: We use the sum of the impact value of the author's total publications to evaluate the author's influence and important work in network science.

We vary transfer weight from 0.1 to 0.9 and the step is 0.1.

[1] http://www.impactfactorsearch.com/

Table 9. The Results of the Most Influential Papers

Rank	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	5	5	5	5	5	5	5	5	5
2	6	6	6	6	6	6	6	6	6
3	12	12	12	12	8	8	8	8	8

5 - “Collective dynamics of 'small-world' networks”

6 – “Emergence of scaling in random networks”

8 – “Navigation in a small world”

12 – “Statistical mechanics of complex networks”

Table 10. The Results of the Most Influential Papers

Rank	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	8	8	6	6	6	6	6	6	6
2	9	9	8	7	7	7	7	7	7
3	6	6	9	8	8	8	8	8	8

6 - Watts D. 7 - Strogatz S. 8 - Barabási, A-L 9 - Albert R.

Table 11. The Results of the Most Influential Journals

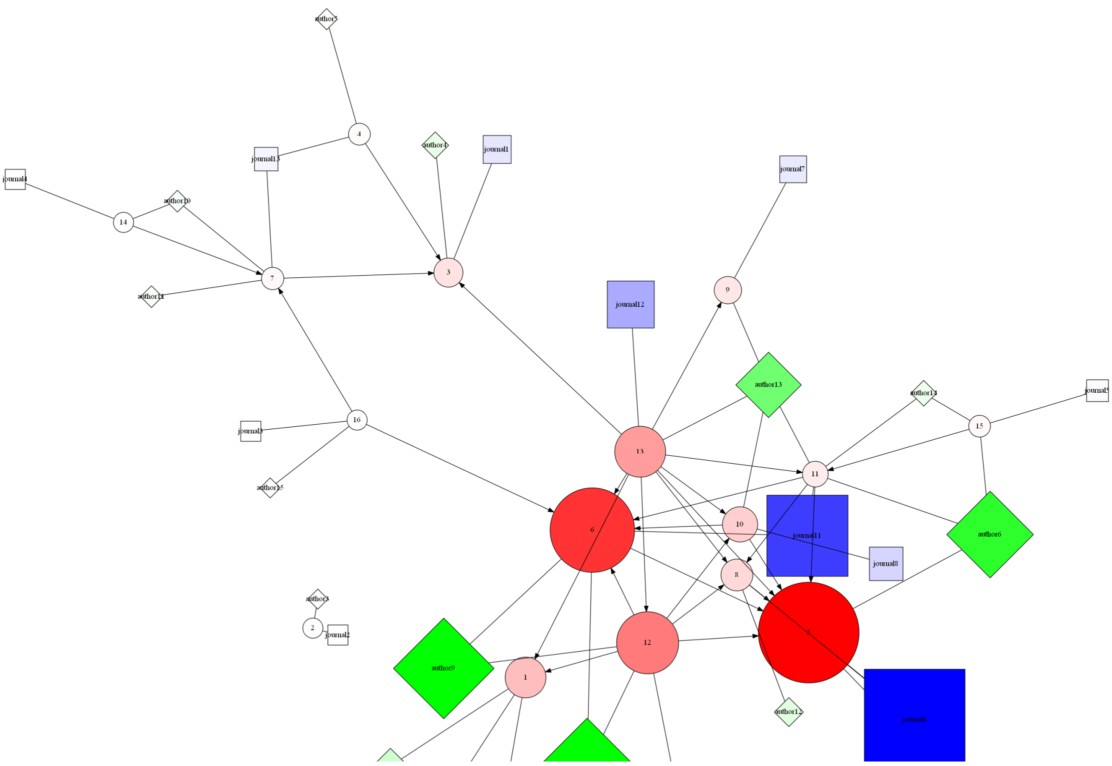
Rank	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	6	6	6	6	6	6	6	6	6
2	10	10	11	11	11	11	11	11	11
3	11	11	10	10	9	9	9	9	9

6 – Nature 9 - Publicationes Mathematicae

10 - Reviews of Modern Physics 11 - Science

5.5 Results and Discussions

According to the results of our model, we consider the paper **Collective dynamics of 'small-world' networks** is the most influential in network science, which is followed by



Emergence of scaling in random networks. An among the individual network science researchers, Watts D, Strogatz S, Barabási A-L, and Albert R have similar impact in the network science, as the results are not stable when the parameters change. Considering the measure role of a journal in network science，Nature is the most influential journal in the field, which is followed by Science and Reviews of Modern Physics.

Figure 9. Visible Result of the Network

5.6 Strengths and Weaknesses

Strengths

- Taking different kinds of nodes (paper, researcher, and journal) into a complex network to make it better to simulate the impact between them.
- We consider randomness of the influence transferred from authors to papers.

Weaknesses

- We cannot set an effective domain for the parameters, which can limit the changes of results.
- Although we have considered the year factor, but without further research, we still cannot make sure the relationships between publication year and impact value.
- There are still other factors that have not been considered in our model, which may have important influence on the results.

6. A Movie-Actor Network

6.1 Build the Movie-Actor Network

We have collected the data of movies and their main actors from internet.[1]

Then we build a movie-actor network of 37 movies and 80 actors. All the movies are Chinese movies whose box office rankings are within top 20 from 2010 to 2013. And our purpose is to find out the most influential actors, whose films can always successfully attract more audients. The difference of this network from the complex citation network is that there are no relationships between films, so the network only shows the direct links between actors and movies.

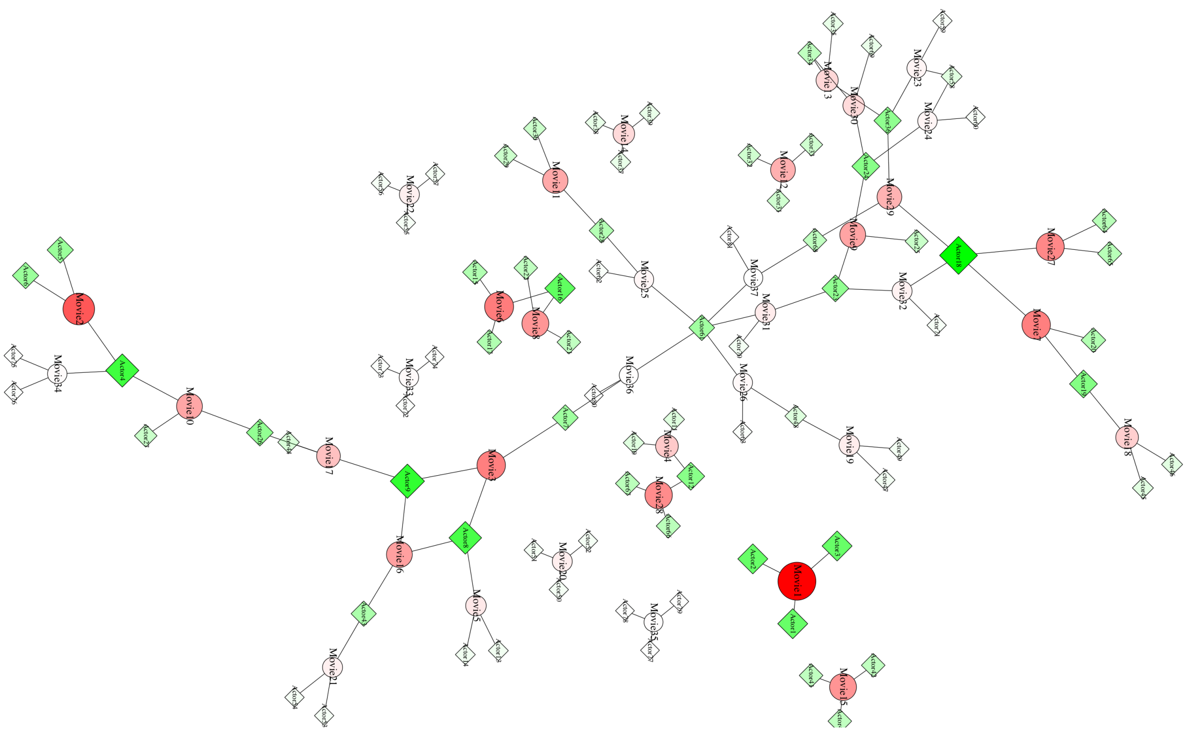


Figure 9. The Movie-Actor Network

6.2 Calculate the Influence of the Movie Actors

Model III can be applied to the movie-actor network directly. Even it is easier than the work in section 5 because there's no links between movies. The impact value of a movie within the network equals to the impact value in the whole movie industry, which equals to its box office.

The result of the network is as below.

Table 12. The Top 5 Actors

Actors	Standard Impact Value
Ge You	1.0000
Chen Kun	0.8063
Jackie Chan	0.7518
Zhou Xun	0.7119
Mark Zhao	0.6222

6.3 Validate of the Result

To validate the result of our model, we get the rank of the influence of movie actor from a professional box office website in China.[1] The top 5 actor is as follow:

Table 13. The Top 5 Actors on Internet

Actors	Standard Impact Value
Jackie Chan	9.52
Zhou Xun	9.45
Mark Zhao	9.42
Zhao Wei	9.38
Zhen Zidan	9.32

Compared to our result, Jackie Chan, Mark Zhao and Zhou Xun are in both top 5 list, indicating that our model have a certain accuracy, but some are not exactly the same due to the limit of the size and the domain of our collected data.

7. Further Modeling Influence

We live life in the network. When we wake up in the morning, we check our e-mail, make a quick phone call, walk outside (our movements captured by a high definition video camera), get on the bus (swiping our RFID mass transit cards) or drive (using a transponder to zip through the tolls). Each of these transactions leaves digital bread-crumbs which, when pulled together, offer increasingly comprehensive pictures of both individuals and groups, with the potential of transforming our understanding of our lives, organizations, and societies. So we can have a better understanding of the world if we can establish an influence and impact model of the elements existing in the network.

But the real-life network may have varieties of main differences from the simple network given in the problem, as it's more complex and more factors should be taken into consideration to get more scientific and meaningful results. Individuals and organizations such as companies often use the analysis results to make important decisions, thus, one of the key orientations of future development of our model is to enhance communication to decision makers and to adjust the model to making recommendations

[1] http://www.entgroup.cn/

more credible, understandable and compelling or persuasive.

Nowadays, people need a lot of information to make significant decisions in their lives. This kind of information can be the results of analysis of a certain network. When the situations change, we can just give the nodes in the network different meanings and establish a new network graph based on the requirements of the new problem. Using our model we have discussed in the previous sections, we could build a mathematical paper-researcher network, calculate the impact value of each node, and find the researcher with the greatest impact value to co-author with. Further, we could put ourselves into the network and using the mathematical model to simulate our own impact in some academic field and find the most influential field. The model would also help us to compute the impact of schools and advisors from the hybrid network in some field.

8. Reference

[1] Newman M E J. The structure and function of complex networks[J]. SIAM review, 2003, 45(2): 167-256.

[2] Borgatti S P, Mehra A, Brass D J, et al. Network analysis in the social sciences[J]. science, 2009, 323(5916): 892-895.

[3] Chung F. Graph theory in the information age[J]. Notices of the AMS, 2010, 57(6): 726-732.

[4] Garfield E. Science citation index: A new dimension in indexing[J]. Science, 1964, 144(3619): 649-654.

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

[6] Garfield E. The evolution of the science citation index[J]. International microbiology, 2010, 10(1): 65-69.

[7] Pinski G, Narin F. Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics[J]. Information Processing & Management, 1976, 12(5): 297-312.

[8] S. N. Dorogovtsev and J.F.F. Mendes (2002). "Evolution of Networks". Adv. Phys. 51: 1079.

[9] T. Wilhelm, J. Kim (2008). "What is a complex graph?". Physica A 387: 2637–2652.