Contents

1 Introduction **2**

2 Assumptions of Model **2**

3 Co-author Network Building **2**
3.1 Building Erdos1 Network . . . 2
3.2 Property Analysis of Erdos1 Network . . . 3
3.3 Graph of Erdos1 Network and Analysis . . . 3

4 Significant Influence within Erdos1 Network **4**
4.1 Analysis of the Task . . . 4
4.2 Degree Rank . . . 4
4.3 Closeness Rank . . . 4
4.4 Betweenness Centrality . . . 5
4.5 PageRank . . . 5
4.5.1 Implementation of PageRank . . . 5
4.5.2 Optimization of PageRank . . . 6
4.6 Result of Four Methods . . . 7
4.7 Analysis of Result . . . 7

5 Significant Influence within Scholar Network **8**
5.1 Analysis of the Task . . . 8
5.2 Pre-process of the Following Papers . . . 9
5.3 Modifications of PageRank . . . 9
5.3.1 Adjust Appropriate Initial Value . . . 9
5.3.2 Ignore Converged Nodes . . . 9
5.4 Measurement of Influence of Papers . . . 10
5.5 Measurement of Influence of Researchers . . . 10
5.6 Result of Influence Measurement of Papers . . . 10
5.7 Analysis of Result . . . 11
5.8 Extension of Influence Measurement . . . 11

6 Significant Influence within Facebook Network **12**
6.1 Analysis of the Task . . . 12
6.2 Apply Rank Algorithms on Facebook Data . . . 12
6.3 Result of Influence Measurement of Facebook Data . . . 12
6.4 Analysis of Result . . . 13

7 Application and Significance of Model **13**
7.1 Apply Model into Individuals . . . 13
7.2 Apply Model into Organizations . . . 13
7.3 Apply Model into Nations . . . 14
7.4 Apply Model into Human Society . . . 14

8 Test of Model **14**

9 Strength and Weakness of Model **14**

10 Conclusion **15**

1 Introduction

In this paper we will introduce a proper method to analyze certain networks and find the significant node. With some optimization and criterions for conversion, the modified method can be applied to a lots of fields computing the influence or impact of nodes within the complex network. Since we need to work out several different but related tasks, the parts below should be taken into consideration:

- Analyze properties of the co-author network and establish a model according to the built network.
- Find a method to evaluate significant influence within the network and optimize it with respect to kinds of requirements.
- Implement a general method that can be applied into lots of fields by converting a certain network given into related networks.
- Evaluate the application of the model and do the analysis.

On the basis of requirements above, what we have done is as follows:

- After building the Erdos1 network, we use four methods to compare the significance of different authors. And we choose PageRank algorithm at last with respect to its best performance.
- In order to improve its accuracy and efficiency, we modify the original algorithm a lot so that it can be applied into much more complex network.
- For the convenience of application in general fields, we propose some methods to convert a certain network into related networks using the specified data.
- Run the optimized algorithm on different data set to test the performance and find the merits and drawbacks.

2 Assumptions of Model

Overall, we use the PageRank algorithm to compute the ranking within networks. Here are some assumptions of optimization. [1]

1 Since the research paper with higher quality is more likely to be cited, the citation relationship between every two papers are not equal.

2 The citation relationship is much more rigorous than surfing on the Internet, an author hardly cites an article at random.

3 The isolated nodes is less important than connected nodes, except that there are lots of isolated nodes connecting to the same nodes.

3 Co-author Network Building

3.1 Building Erdos1 Network

The given file consists of 18000 lines of raw data, but we ought to remove most of them since we only need to build a co-author network of the 511 researchers from the file. Hence there is a lot to do with the data. All steps are as follows:

- Extract 511 researchers from the file and number them in ascending order from 1 to 511 so as to cope with researcher names easily.
- Construct a list where each item consists of all co-authors within 511 researchers for corresponding author.
- Construct a list where each item consists of one author and one co-author in order using the list constructed at the previous step.
- Remove items which repeat in the list constructed at the previous step since the co-author link is undirected. And the final list just represents the co-author network where the number stands for the corresponding author.

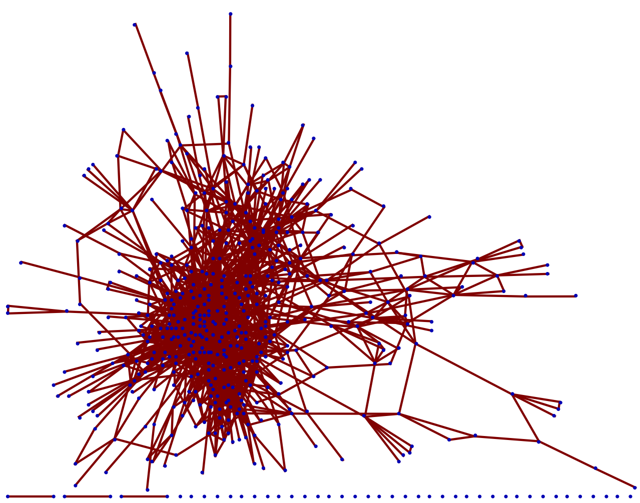
3.2 Property Analysis of Erdos1 Network

After building the co-author network of the Erdos1 authors, we find properties of the network from *Wikipedia* [2]. Here are the properties of this nework:

- Size: The size of a network can refer to the number of nodes $N = 511$.
- Density: The density D of a network is defined as a ratio of the number of edges $E(= 1641)$ to the number of possible edges(C_N^2), giving $D = 0.0126$.
- Average degree: The degree k of a node is the number of edges connected to it. Therefore, $\langle k \rangle = 6.42$.
- Average path length: Average path length is calculated by finding the shortest path between all pairs of nodes, adding them up, and then dividing by the total number of pairs. Average path length= 4.00.
- Diameter of a network: The diameter of a network is the longest of all the calculated shortest paths in a network. Diameter of a network= 14.

3.3 Graph of Erdos1 Network and Analysis

The graph of the Erdos1 network drawn by Mathmatica is shown as follow.



4 Significant Influence within Erdos1 Network

4.1 Analysis of the Task

To determine who has published important works or connects important researchers within Erdos1, we can use many methods to evaluate the importance of the nodes in the network using graph theory and graph-based data mining. At a glance, the amount of links of each author within the network is a major criterion, i.e. the degree of each node within the network. Apparently, one with the most co-authors is possible to have significant influence within the network. However, we can't use the degree of each node straightly, since the influence of each author is quite different. In addition to the degree of node, we may make best use of its adjacent nodes or even the second nearest nodes. In fact, each edge within the network should be assigned a different value on the basis of its linked nodes.

In graph theory and network analysis, centrality of a vertex measures its relative importance within a graph [4]. For Erdos1 network, the centrality of a vertex can reflect the influence of corresponding researcher in a manner. To dig into the information in the network, we use four different methods to measure influence and compare results of them in order to make our analysis of the network properties more comprehensive.

4.2 Degree Rank

Historically first and conceptually simplest is degree ranking, which is defined as the number of links incident upon a node (i.e., the number of ties that a node has). In a complex network, if few high-degree hubs control the whole network while most of the rest nodes have low degree, obviously degree ranking is a key factor to this kind of extreme heterogeneous network. Ranking the nodes in a complex network according to their degrees can help us tell the importances of different nodes in some ways.

The degree centrality of a vertex v, for a given graph $G := (V, E)$ with $|V|$ vertices and $|E|$ edges, is defined as:

$$C_D(v) = deg(v) \tag{1}$$

In Erdos1 network, $deg(v)$ is the number of co-authors for each v. Calculating degree centrality for all the nodes in a graph takes $\Theta(V^2)$ in a dense adjacency matrix representation of the graph, and for edges takes $\Theta(E)$ in a sparse matrix representation.

4.3 Closeness Rank

On the other hand, the nodes of high importance are relatively located in the center of the network, so closeness ranking should also be considered to analyze this Erdos1 network. The farness of a node is defined as the sum of its distances(shortest paths) to all other nodes, and its closeness is defined as the inverse of the farness.
The closeness centrality of a vertex v, for a given graph $G := (V, E)$ with $|V|$ vertices and $|E|$ edges, is defined as:

$$C_C(v) = \frac{1}{\sum_{y \neq x} d(y, x)} \tag{2}$$

where $d(y, x)$ is the shortest path between x and y.

Thus, the more central a node is, the lower its total distance to all other nodes. Therefore, closeness can be regarded as a measure of how long it will take to spread information from a node to all others sequentially. However, closeness ranking gives high

scores to those nodes which are near the center, but neglects the position differences of the nodes. In fact, it works well for the star network but is not suitable for other types of network, for instance, the one with small-world property.
In calculation aspect, we can use Brandes' algorithm to easily sum up the shortest paths for each nodes since every edge has the same weight.

4.4 Betweenness Centrality

The nodes with high betweenness score are important to transfer more information. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes [5]. It was introduced as a measure for quantifying the control of a human on the communication between other humans in a social network by Linton Freeman [6]. In his conception, vertices that have a high probability to occur on a randomly chosen shortest path between two randomly chosen vertices have a high betweenness.

The betweenness of a vertex v in a graph $G := (V, E)$ with V vertices is computed as follows:

1 For each pair of vertices (s, t), compute the shortest paths between them.

2 For each pair of vertices (s, t), determine the fraction of shortest paths that pass through the vertex in question (here, vertex v).

3 Sum this fraction over all pairs of vertices (s, t).

More compactly the betweenness can be represented as:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3}$$

where σ_{st} is total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.
From a calculation aspect, both betweenness and closeness centralities of all vertices in a graph involve calculating the shortest paths between all pairs of vertices on a graph, which requires$\Theta(V^3)$ time with the FloydWarshall algorithm. However, in the case of unweighted graphs just like Erdos1 network, the calculations can be done with Brandes' algorithm which takes $O(VE)$ time. Normally, these algorithms assume that graphs are undirected and connected with the allowance of loops and multiple edges. When specifically dealing with network graphs, oftentimes graphs are without loops or multiple edges to maintain simple relationships (where edges represent connections between two people or vertices). In this case, using Brandes' algorithm will divide final centrality scores by 2 to account for each shortest path being counted twice.

4.5 PageRank

4.5.1 Implementation of PageRank

Since the Erdos1 network is also a link graph in nature, it's a good idea to analyze this link problem using PageRank algorithm [3]. PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set.

For Erdos1 network, PageRank is a probability distribution used to represent the likelihood that an author will try to co-author with someone.

In the general case, the PageRank value for any author A can be expressed as:

$$PR(A) = \frac{1-d}{N} + d \sum_{a \in S_A} \frac{PR(a)}{L(a)} \tag{4}$$

i.e. the PageRank value for an author A is the sum of two terms. The first term is dependent on the PageRank values for each author a contained in the set S_A (the set containing all authors co-author with A), divided by the number $L(a)$ of links from author a. The other term is dependent on the reciprocal of the number of authors. And d in the formula represents the probability that the author will continue to co-author with someone.

With the basic PageRank equation, now we can use the complete PageRank algorithm.

For ease of calculation, PageRank values of all authors are represented as a dominant eigenvector. The program needs to calculate the approximate or exact value of this eigenvector. The program will compute the final result iteratively though PageRank can be computed either iteratively or algebraically.

During each iteration, the program will calculate the probability distribution of all authors. The computation is as follow:

$$PR(A_i; t+1) = \frac{1-d}{N} + d \sum_{a \in S_{A_i}} \frac{PR(a;t)}{L(a)} \tag{5}$$

Especially, at $t = 0$, an initial probability distribution is assumed as:

$$PR(A_i; t=0) = \frac{1}{N} \tag{6}$$

After several iterations, the collection to adjust approximate PageRank values will be more closely to reflect the theoretical true value. The author with the largest PageRank is most likely to the one who has significant influence within the network.

4.5.2 Optimization of PageRank

Although PageRank algorithm performs well while ranking websites in search engine results, it has the limitation of treating all adjacency nodes equally. Using uniform parameters in probability transfer matrix may not be fair in this network. Therefore, we use a modified method for evaluating nodes importance in complex networks, which are based on contributing degree of nodes similarity after studying the shortcomings of PageRank algorithm [7]. This method suggests using the contributing degree of nodes similarity to reconstruct the modified probability transfer matrix among all adjacency nodes, and replacing uniform parameter in PageRank algorithm by normed nodes closeness values, to reappraise nodes importance in complex networks. We believe this modified method can find important nodes in complex networks more precisely and effectively.

The steps of the optimized PageRank algorithm are as follows:

1 Calculate the max-eigenvalue λ_1 of adjacency matrix A. Choose parameter $\alpha \in (0.9, 0.99)$.

2 Calculate nodes similarity matrix S, $S = [E - \varphi A]^{-1}$, where $\varphi = \alpha/\lambda_1$. [8]

3 Get probability transfer matrix M from nodes similarity matrix S: $m_{ij} = \frac{s_{ij}}{\sum_{j=1}^{n} s_{ij}}$

4 Calculate the closeness value of each node, modify the parameters in PageRank algorithm and get the value of PageRank:

$$PageRank = \lambda M \times PageRank + (1 - \lambda) \times \frac{Closeness}{Sum(Closeness)} \tag{7}$$

where λ is an adjustment coefficient.

4.6 Result of Four Methods

Table 1: Co-author Numbers of Some Top Researchers

Author	Co-author Number
ALON, NOGA M.	52
HARARY, FRANK	44
BOLLOBAS, BELA	43
GRAHAM, RONALD LEWIS	44
SOS, VERA TURAN	38

Table 2: Order of Top 10 Authors for Degree Rank and Closeness Rank

Order	Degree Rank	Closeness Rank
1	ALON, NOGA M.	GUY, RICHARD KENNETH
2	GRAHAM, RONALD LEWIS	ALAVI, YOUSEF
3	HARARY, FRANK	AJTAI, MIKLOS
4	BOLLOBAS, BELA	ALAOGLU, LEONIDAS
5	RODL, VOJTECH	ALON, NOGA M.
6	TUZA, ZSOLT	GYARFAS, ANDRAS
7	FUREDI, ZOLTAN	SARKOZY, ANDRAS
8	SOS, VERA TURAN	TROTTER, WILLIAM THOMAS, JR.
9	SPENCER, JOEL HAROLD	GOULD, RONALD J.
10	GYARFAS, ANDRAS	ALLADI, KRISHNASWAMI

4.7 Analysis of Result

Using above four different methods to evaluate nodes importance, table 1 and table 2 show the top 10 vital nodes of each methods.

From table 1, we can see that degree ranking, betweenness ranking and PageRank algorithm have very similar results. Their top 10 important nodes are almost the same on the whole considering that the node size is 511. However, result of closeness ranking has a very big difference comparing to other three methods. This is because closeness ranking has its own problem, which prefers the nodes near the center, but neglects the position differences of the nodes. Actually it works well for the star network but should

Table 3: Order of Top 10 Authors for Betweenness and PageRank

Order	Betweenness Centrality	PageRank
1	ALON, NOGA M.	HARARY, FRANK
2	HARARY, FRANK	SOS, VERA TURAN
3	BOLLOBAS, BELA	POMERANCE, CARL BERNARD
4	CHUNG, FAN RONG KING (GRAHAM)	ALON, NOGA M.
5	GRAHAM, RONALD LEWIS	BOLLOBAS, BELA
6	GYARFAS, ANDRAS	GRAHAM, RONALD LEWIS
7	SOS, VERA TURAN	STRAUS, ERNST GABOR
8	FUREDI, ZOLTAN	TUZA, ZSOLT
9	ALAVI, YOUSEF	HAJNAL, ANDRAS
10	RODL, VOJTECH	RODL, VOJTECH

not be applied to other types of network, like the one with small-world property etc. There are also some hortcomings for the others. For instance, degree ranking can't tell the importance of the nodes with the same degree and betweenness ranking is not able to evaluate the importance of leaves (which are contained in no shortest paths) and they would have a betweenness of 0.

From table 1 and table 2, we can see that ALON, NOGA M. comes first in degree ranking and betweenness ranking. However, he is only NO.4 in PageRank list. This explains the point mentioned earlier in this section that we can't simply evaluate a node according to its degree but need to consider more about the adjacency nodes importances. For this reason, we think that ALON, NOGA M. is less important than HARARY, FRANK who is ranked first in PageRank list. He also has high scores in degree ranking and betweenness ranking. Thus, we believe the optimized PageRank performs best in this Erdos1 network because it not only can overcome the limitations of degree ranking, closeness ranking and betweenness ranking, but also can work out the limitation of treating all adjacency nodes equally in PageRank algorithm.

In conclusion, the node representing HARARY, FRANK seems to be the most important node in this Erdos1 network. Therefore, we say HARARY, FRANK has significant influence within this network.

5 Significant Influence within Scholar Network

5.1 Analysis of the Task

Just like the last task, we can construct a network built from these 16 research papers and measure influence using PageRank as well. Nonetheless, the network is too simple to get enough information. Therefore, it needs us to use a web crawler to find enough papers following from these 16 papers. Now that we have enough papers to build a pretty complex but useful network, we are able to compare the significance of all papers using PageRank. The paper with the largest PageRank is most likely to be the one with significant influence.

5.2 Pre-process of the Following Papers

To get enough works that follow from these 16 papers, we meant to crawl following papers from Google Scholar, however, the number is so large that it's impossible to get all information in just 4 days. Therefore, we turn to Microsoft Academic Search which contains less papers of which the quality is approximately $\frac{1}{4}$ of that from Google Scholar.

Firstly, we search for all papers that follow each within 16 research papers, i.e., all papers that have cited these 16 research papers. However, we don't know the relationship of these crawled papers, so we can't build a network to describe them. Then we continue to search for all papers that cite these crawled papers. Now we are able to build a network including 16 research papers and all crawled papers.

As the last step, we cope with these data in a similar way to Task 1. Firstly, we number all papers in the first level which cite the 16 research papers in ascending order. Secondly, we construct a list where each item consists of a paper and a cited paper after removing the repeated papers.

Finally, we get a similar network to Task1 which can be ranked using the ranking algorithm.

5.3 Modifications of PageRank

For this task, we only use PageRank to rank these papers. There is no fundamental difference between task2 and task3, but the huge number of nodes may increase time and space needed using the same algorithm. Thus it's absolutely essential to improve the efficiency of PageRank algorithm in order to cope with complex network.

5.3.1 Adjust Appropriate Initial Value

When computing the PageRank values iteratively, the collection to adjust approximate PageRank values will be more closely to reflect the theoretical true value. Originally we set all initial values as $\frac{1}{N}$, which is independent with the significance of each node. So a appropriate initial value which is close to true value will decrease much time to compute iteratively.

K-Shell Decomposition [9] is just an efficient way to rank the nodes approximately in a network. The k-shell decomposition algorithm is a well-established method for detecting the core and the hierarchical structure of a given network. The assumption is that , if these nodes are authors in a paper network, the authors in the higher k-shell levels are more influential in the network than authors in lower k-shell levels.

In general, the k-shell decomposition algorithm searches for nodes in each shell level from level 0 to higher level. At last, the algorithm will classify these papers into many shells. Now we can assign appropriate initial values to all papers which can reduce the number of iterations.

5.3.2 Ignore Converged Nodes

In practice, original PageRank algorithm is not efficient to cope with complex network. Nodes with low PageRank converge faster while the one with high rank spend more time to converge. During the later iterations, the PageRank values of most of the nodes have been decided, which could be passed in this iteration in order to reduce operations. Hence when the PageRank value of certain node is converged already, we could ignore this node in later iterations.

5.4 Measurement of Influence of Papers

Now that we have built the paper network and modified the PageRank algorithm, it's possible to measure the influence of all papers including 16 research papers.

Firstly, run PageRank algorithm on this huge network to rank all papers. Then what we need is just the ranking of 16 research papers, so extract these 16 papers with PageRank values and reorder these papers. Now the higher PageRank value of the paper is, the more important the paper is.

5.5 Measurement of Influence of Researchers

Although we have already built a huge network of all papers, it doesn't make sense as for influence of researchers. What we should do now is to get the relationship among all researchers.

In the beginning, we meant to rank all researchers according to their own research papers. It's an intuitive and rational idea to rank researchers according to their published papers, but it will waste much information within the paper network. So we want to build a network of researchers, then get the influence of researchers from the network directly.Now the fundamental problem is how to convert a paper network into a researcher network without loss of information.

For a researcher network, each node represents a different researcher, therefore, the relationship between any two nodes should be converted from the relationship in paper network. In paper network, a link with direction stands for the citation from one paper to another paper. If paper A written by authors a and b cites paper B written by authors c and d, then there will be links with direction from a to c and d, from b to c and d in researcher network. However, there doesn't exist a link between a and b or c and d, because links in network are used to compute PageRank values and the link between two co-authors doesn't make a difference to their significance. It is worthwhile to note that there could be more than one links between two nodes, and the more links there are, the stronger the connection between two noes is. Now the remaining task is to simplify this network by removing many repeated links. After removing these repeated links, each edge in the simplified network should be assigned a value proportional to the number of links before removing operation.

Here are the steps of conversion from the paper network into researcher network:

1. Extract all researchers included in these papers.
2. Add a link from one author to another author if the previous one's paper cites the latter one's pear.
3. Removing all repeated edges, and assign a value proportional to the number of corresponding links to each edge.

Now that we have built the network of all researchers, we can use PageRank algorithm to compute the PageRank values of all researchers. After that, we could determine the role or influence measure of an individual network researcher according to PageRank values.

5.6 Result of Influence Measurement of Papers

Table 4 contains the research papers and corresponding cited number in order.

Table 4: Order of Influence of 16 papers

Order	Research Paper	Cited by
1	Collective dynamics of ‘small-world’ networks	5190
2	Emergence of scaling in random networks	4499
3	Statistical mechanics of complex networks	3536
4	On Random Graphs	1112
5	The structure and function of complex networks	3081
6	Power and Centrality: A family of measures	510
7	Scientific collaboration networks: II...	284
8	Navigation in a small world	311
9	Identity and search in social networks	208
10	The structure of scientific collaboration networks	165
11	Social network thresholds in the diffusion of innovations	154
12	Models of core/periphery structures	193
13	Networks, influence, and public opinion formation	100
14	Identifying sets of key players in a network	49
15	Statistical models for social networks	8
16	On properties of a well-known graph...	4

5.7 Analysis of Result

From Table 4, we find that *Collective dynamics of small-world networks* is the most influential in network science. The major criterion is the PageRank value of each paper. Observing the cited number, we can find an intuitive relationship between the significance and cited number. However, there are two special cases including *On Random Graphs* and *Navigation in a small world*.

In conclusion, the PageRank value of each paper is proportional to the cited number roughly. However, these special cases mean that the cited number is not the only decisive factor. The quality of papers that cite the same paper is also an essential factor. When running PageRank algorithm, the paper which has been cited many times will increase the influence of papers it cited. As a result, some of papers with less but higher quality papers can get much larger PageRank values.

However, we didn’t crawl all citing articles, so there is still some errors in the computation. But with more data, the PageRank values of papers will be more accurate.

5.8 Extension of Influence Measurement

With enough research papers in network science, then we can measure the role, influence, or impact of a specific university, department, or a journal in network science. Similar to ranking authors, we should first convert the page network into university network, department network or department network. Then we can use PageRank algorithm to evaluate the PageRank value of each node which represents the significance of the node.

In general, the PageRank algorithm can be applied to all kinds of networks. What we need to do is just the conversion from the page network into the corresponding network. Therefore, there is one requirement for the methodology. The data collected should contains enough information related to relevant fields. For example, the collected papers should include authors, published journals, universities or departments. Without related information, we have no way to convert links in the original network into links of new types. With appropriate data, firstly extract the new type of information as new nodes, then draw the links on the basis of the original links in the network, and simplify the new network and assign values to each edge accordingly. The last step is run the PageRank algorithm on the new network and get the PageRank values of all nodes as a criterion to measure the influence.

6 Significant Influence within Facebook Network

6.1 Analysis of the Task

In this part, we need to implement our algorithm on a completely different set of network influence data. The main job for us is to search for a network data which is different from the previous Erdos1 network or paper network. Social network such as Facebook or Tweeters may be a good choice. As for algorithm, we mainly use optimized PageRank to analyze the data, and three other methods will also help us get some interesting conclusions.

6.2 Apply Rank Algorithms on Facebook Data

Here we choose a dataset of the famous social network Facebook from http://snap.stanford.edu/data/egonets-Facebook.html. In social networks, "nodes" of the network are people and the "links" are the relationships between people. This data has been anonymized by replacing the Facebook-internal ids for each user with numbers from 1 to n. Each line in this data including two numbers represents that these two numbered people are friends on Facebook. The original network has a size of 4039 nodes and 88234 edges and we restrict it to 1000 nodes and 9890 edges for convenience of calculation and analysis. We still use four different evaluating methods including optimized PageRank algorithm to analyze this data.

6.3 Result of Influence Measurement of Facebook Data

Table 5 shows the top 10 vital nodes of each method.

Table 5: Top 10 Vital Nodes of Four Methods

PageRank	Degree Rank	Betweenness Rank	Closeness Rank
1	1	1	687
108	349	349	829
349	687	35	714
415	415	108	699
687	108	415	806
699	377	687	720
377	476	699	746
26	413	377	748
484	484	199	824
120	374	174	831

6.4 Analysis of Result

Table 5 shows the top 10 vital nodes of each method. Analysis from table 5, we can see that the result of closeness ranking is still quite different from the other three, which means that this Facebook network is not a typical star network. Node 1 is ranked first through three methods, which reveals its high importance in this social network. Node 108 and 349 are similar, they are also the popular stars in this social network. Node 687 is No.5 in PageRank list, and in the first position using closeness ranking, which suggest that Node 687 is closer to the center of network than anyone else.

The result proves a third time that we can not only focus on the isolated properties of each node, but also should consider more about the neighborhood. Take this social network for instance, a large number of friends can't suggest a person be the super star of this network because maybe most of his friends are those silent people who has very few friends. For example, node 108 has fifth large number of friends but can't make it into top 10 in our algorithm rank.

7 Application and Significance of Model

Complex networks, which are composed of a number of nodes that are interconnected by a set of edges, are widely observed in a vast range of natural and artificial systems in recent years, ranging from the brain through the Internet to human society. Identifying the most important nodes and to define the importance of the nodes become more and more essential. Apart from the fields mentioned above, nodes importance evaluation methods in complex network also play an important role in other fields.

7.1 Apply Model into Individuals

First of all, at the individual level, for instance, if one wants to boost influence as soon as possible, he can choose someone to co-author with as follows:

1 Construct a new node to represent himself.

2 Add a new node into the original author network by linking with one author in this network. Run the algorithm and get a ranking position of himself.

3 Repeat the second step for every author in the network.

4 Observe the result and clearly we should choose the one who make us get maximum ranking score.

Likewise,when one needs to expand his social network, it is a shortcut that he makes friends who are the key node in the social network.

7.2 Apply Model into Organizations

As for organizations, we take the school and the company for example. For schools, we regard all kinds of majors in school, as well as other institutions including schools, research laboratory and so on, as nodes, then repeat the steps in 5.1, we can figure out which is the focus of the future academic work. In addition, when mentioning the company, we think of apartments and sections of other corporations as nodes, which are linked by trades and cooperations, then repeat the steps in 5.1, we can decide the future direction of enterprise.

7.3 Apply Model into Nations

Modeling influence or impact within networks has shown its great practical value and realistic significance in national politics. Though peace and development have become the two major topics in today's world, an undercurrent of terrorism is a serious threat to human life and safety and the further development of human society, which is called the political plague of the 21st century and regarded as one of the ten major public nuisances in the 21st century. The subway bombings in the Russian capital this March shocked the world; $7 \cdot 7$ London bombing, $9 \cdot 11$ American case and $3 \cdot 14$ Lhasa riots feared the people after the events. Countries around the world have agreed to work jointly to combat terrorism. To these organizations such as terrorist organizations and criminal organizations, the police can quickly locate and arrest the terrorist leaders and key people by using nodes importance evaluation methods in complex network to destroy the organization, maintain the stability of human society, economy and politic and accelerate social development.

7.4 Apply Model into Human Society

Influence methodology also has wide application in the field of human society. Infectious diseases, such as AIDS (acquired immune deficiency syndrome) and H1N1 (influenza A) pose a great threat to human beings. AIDS caused a panic in human beings. SARS (atypical pneumonia) breaking out in 2003 has negative effects on the world economy and the human health. Since March last year, H7N9 has become the focus of the world, threatening the human health and crashing the warming world economy. So, how can we control these infectious diseases to decrease the losses? Maybe we can seek answers from complex networks, especially the nodes importance evaluation methods in complex network, which can prevent further spread of the diseases and minimize the losses by isolating and treating the critical patients.

8 Test of Model

Sensitivity of PageRank.We show the sensitivity of PageRank algorithm according to [10]. The PageRank vector π is perfectly conditioned with regard to changes in the personalization vector v. The sensitivity of the PageRank vector π to changes in the damping factor α depends on a condition number that is bounded by $2/(1-\alpha)$. Adding an inlink to a node increases its PageRank. Especially, if a link is added from node i to node $j \neq i$ and if node i does not have a link to itself then the PageRank of node j increases. Adding an outlink to a node decreases the PageRank.

9 Strength and Weakness of Model

Strength:

- The optimized algorithm can evaluate the PageRank values efficiently and accurately.
- The optimized algorithm can be applied into many fields related to complex network.

Weakness:

- The process of the network construction is tedious , whats more, the previous work cant lessen the task when adding a new node to the old node group.
- The modle ignores the factor of time. Since the Pagerank is a static algorithm, the model will be out of work when coming aross the situation where it needs to analyze the trend or potential of something.

10 Conclusion

Mining vital nodes is an important step of network science. In this paper, we build a Model using PageRank algorithm. In addition, we also optimize the algorithm and summarize a general method which can be used to measure influence in networks.

During the calculation and validation, we find out that our model shows a quite good accuracy and efficiency to the network though there are still some weaknesses to be remedied. We hope that this optimized PageRank algorithm can provide a convincing node importance rank in many complex network environments.

References

[1] http://www.webpronews.com/addressing-assumptions-of-the-original-pagerank-2008-01

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

[3] http://en.wikipedia.org/wiki/PageRank

[4] http://en.wikipedia.org/wiki/Centrality

[5] M.E.J. Newman, *A measure of betweenness centrality based on random walks*, Social Networks, 27: 3954, (2005)

[6] HE Nan, Gan Wen-yan, LI De-yi, *Evaluate Nodes Importance in the Network using Data Field Theory*,(2007)

[7] He Jian-Jun, *Node Importance Evaluation in Complex Network*, (2010)

[8] Leich E A, Petter Holme, Newman M EJ., *Vertex similarity in networks*, Physical Review E, 73(2):1-10, (2006)

[9] Philip E. Brown, Junlan Feng, *Measuring User Influence on Twitter Using Modified K-Shell Decomposition*, Fifth International AAAI Conference on Weblogs and Social Media, (2011)

[10] ICF Ipsen, RS Wills, *Mathematical Properties and Analysis of Googles PageRank*, Bol. Soc. Esp. Mat., (2006)