International Journal of Multidisciplinary and Current Research

Research Article

ISSN: 2321-3124
Available at: http://ijmcr.com

Efficient Shortest Path Estimation in Social Network

Mirza Mahmood Baig and Samreen Zaman

Department of Mathematics, NED University of Engineering & Technology, Karachi, Pakistan

Accepted 01 Sept 2015, Available online 10 Sept 2015, **Vol.3** (**Sept/Oct 2015 issue)**

Abstract

By the passage of time social networking sites has become a vital part of every one's life. These sites have played an important role in connecting people from all around the world. These social connections can be studied as graphs. In these graphs the social actors(people) are the nodes and the relationships by which they are connected are the edges. The edges are given weights according to the intensity in their bonds. This research paper focuses on finding out the shortest path between social actors (people) by using Shimbel's algorithm. Previously researchers have applied Djikstra's algorithm to find the shortest path of social networks. Shimbel's algorithm is also used to examine the average distance between the social actors so that an estimate can be made that on average the people connected in social network are how many steps away from each other.

Keywords: *Social network; sociograms; shortest path; average distance.*

1. Introduction

Social Network is a social arrangement made up of group of social actors that are linked by family relations, work, friendships, interests and partners in crime. The social network's viewpoint is to endow with set of methods to explore the structure of social individuals and the diversity of theory illuminating the patterns observed in these structures (Wasserman and Faust, 1994).

Social Network Analysis (SNA) is the research of social networks that views social associations as networks, consisting of nodes (representing entity within the network) and ties (which shows interaction between the individuals for example organizations, friendship, and kinship etc.).SNA provides illustration as well as mathematical examination of individual relationships (D'Andrea, Alessia et al, 2009).

Analysts of social networks use pictorial representation that consists of nodes to denote actors and edges to exemplify relations or ties. The minute sociologists plagiarized this way of graphing relations from the mathematicians; they re-named their graphics by the word "socio-grams"(Robert A. Hanneman,2005).

In mathematics these kinds of graphs are known as "graphs", "signed graphs" or "directed graphs". A sociogram (graph representing social relations) is made up of actors (denoted by nodes) connected by relations (denoted by edges). A graph of social network can represent a single form of relations between the actors known as simplex, or multiplex relation which defines more than one kind of relation amongst the actors. Social ties can be directed graph (i.e. initiates from a source actor and scopes towards a target actor), or it can also be an undirected graph (i.e.a two sided relationship amongst the duos of actors). Arrows are used to represent directed relationships and line segments are used to symbolize undirected graphs. The strength of connections between actors in a social graph may be 'dualistic' symbolizing absence or presence of relationship or it may be 'signed' denoting negative, positive or no bond, it can also be 'ordinal' showing whether the connection is the strongest, weakest etc. (Robert A. Hanneman,2005).

wo forms of matrices i.e. "Adjacency matrix" and "Incidence matrix" denote graphs (Kousha Etessamiv,2014).

To exemplify a social networks graph an adjacency matrix is used acknowledged as "socio-matrix" (Wasserman and Fraust,1994).In a sociogram if two actors are directly linked then A (i,j)=1 and if not allied directly but linked by some intermediaries then A(i,j)=(leastnumeral of intermediaries)+1.

Stanley Milgram,Jeffrey Travers(1969)conducted a famous experiment known as the "Small World Problem".This experiment was designed to find out that any two persons picked arbitrarily from an enormous population(in this experiment the United States). If a direct connection between people doesn't exist than on average how many steps are they away from each other. In this experiment 296 peoplewere selected randomly in Nebraska and Boston (as preliminary persons) and a target person in Massachusetts, and the starting persons were asked to generate an acquaintance chain to the target person. Each starter forwarded a mail to a person he knows, and he thinks could know the target person or someone knowing the target person. The sender was insisted to choose the recipient in such a way as to advance the progress of the document toward the target. The result was that out of 296 folders 64 actually reached the target person i.e. 29% (Jeffrey Travers and Stanley Milgram , 1969).

The result of this experiment shows that the mean number of intermediaries is somewhat greater than 5. From this experiment the conclusion was drawn that in United States people are not more than 6 steps away. This theory is also known as "six degrees of separation" (Jeffrey Travers and Stanley Milgram , 1969).

Newman (2001) constructed a network of collaboration by using computer database of papers in biomedical research, physics and computer science. If two scientists have co-authored one or more than one papers they are considered to be connected in the networks (M.E.J Newman, May 2001).

Newman in addition to his research used this data again to find an evaluation of the strength of collaborative relations and used the weighted graph of collaborations between scientists to compute distance among scientists. He used Dijkstra's algorithm to calculate the shortest path between scientists (M.E.J. Newman, June 2001).

This research paper is based on finding the shortest path of a weighted undirected Social Network's graph. Shimbel's Method is used to find the optimal solution and then backward tracking is used to find the routes of these optimized distances. Shimbel's method is also applied to find out the average distance between the social actors (nodes) in order to find out that on average the people connected by this undirected weighted graph are how many steps away from each other.

2. Shortest path

2.1 Weighted undirected sociogram

If a sociogram is weighted, then the weight (strength) of the social connection may be its emotional intensity, frequency of connection or exchange of information.

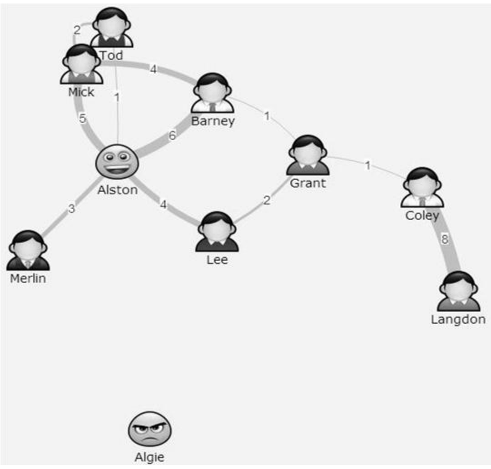

Figure 1: Social networks graph [2]

Data of social networks graph (figure 1) containing 10 actors (nodes) and their intensity of friendship represented by weighted edges.

1. Tod
2. Mick
3. Alston
4. Lee
5. Barney
6. Merlin
7. Grant
8. Coley
9. Langdon
10. Algie

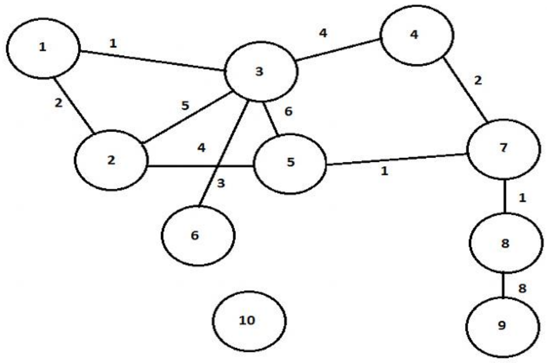

Figure 2: Social networks graph with nodes defined by numbers

Figure 1 has 10 nodes (actors) , node 10(Algie) is an isolated vertex .

2.2 Shimbel's Method

(Shimbel, 1955) introduced the following "min sum algebra":

Arithmetic:

For any arbitrary infinite or real numbers X and Y

X + Y ≡ min(X,Y)
And
XY ≡ the algebraic sum of X and Y

2.3 Structure matrix

A structural matrix is a square matrix. Used to represent both directed and undirected graphs, the construction for undirected graph is mathematically defined as:

$$SN = \begin{cases} \infty \;\; if\ there\ is\ \ no\ connected\ edges\ in\ the\ graph \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad if\ i = j \\ weights \qquad\qquad\qquad\qquad\qquad otherwise \end{cases}$$

(i.e. no loop at any node)

It gives weights when the edges are connected directly, if the edges are not directly connected then it gives infinity and elsewhere zero.(i.e. no loop on the node)

Using the definition of structural matrix, the matrix of figure 1 is represented as:

$$\begin{bmatrix} 0 & 2 & 1 & \infty & \infty & \infty & \infty & \infty & \infty \\ 2 & 0 & 5 & \infty & 4 & \infty & \infty & \infty & \infty \\ 1 & 5 & 0 & 4 & 6 & 3 & \infty & \infty & \infty \\ \infty & \infty & 4 & 0 & \infty & \infty & 2 & \infty & \infty \\ \infty & 4 & 6 & \infty & 0 & \infty & 1 & \infty & \infty \\ \infty & \infty & 3 & \infty & \infty & 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & 2 & 1 & \infty & 0 & 1 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & 1 & 0 & 8 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 8 & 0 \end{bmatrix}$$

Figure 3: Structural matrix of figure 2

But since the graph is of a social network, and in social networks graph the strength of the relationship is defined by its weight. So the weights of the edges will be inversed, the reason is that the more weight (interaction) is between two people the closer they are to each other. Applying this theory to social network in figure 2, and making its structural matrix (SN).Node 10 (Algie) is an isolated vertex so it is not included in the structure matrix

$$SN= \begin{bmatrix} 0 & \frac{1}{2} & 1 & \infty & \infty & \infty & \infty & \infty & \infty \\ \frac{1}{2} & 0 & \frac{1}{5} & \infty & \frac{1}{4} & \infty & \infty & \infty & \infty \\ 1 & \frac{1}{5} & 0 & \frac{1}{4} & \frac{1}{6} & \frac{1}{3} & \infty & \infty & \infty \\ \infty & \infty & \frac{1}{4} & 0 & \infty & \infty & \frac{1}{2} & \infty & \infty \\ \infty & \frac{1}{4} & \frac{1}{6} & \infty & 0 & \infty & 1 & \infty & \infty \\ \infty & \infty & \frac{1}{3} & \infty & \infty & 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & \frac{1}{2} & 1 & \infty & 0 & 1 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & 1 & 0 & \frac{1}{8} \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \frac{1}{8} & 0 \end{bmatrix}$$

Figure 4: Structural matrix (SN) of social network (in figure 2) after taking inverse weights

Applying Shimbel's method on matrix given in figure 4, to obtain the optimal matrix i.e.;Optimal matrix=initial matrix$^{(\text{number of nodes-1})}$.

In the given social network the total number of connected nodes is 9 so,the optimal matrix is obtained by finding out initial matrix to the power 8,the method has to be applied 3 times as the Shimbel's method follows geometric progression.

$$SN^2= \begin{bmatrix} 0 & 0.5000 & 0.7000 & 1.2500 & 0.7500 & 1.333 & \infty & \infty & \infty \\ 0.5000 & 0 & 0.2000 & 0.4500 & 0.2500 & 0.533 & 1.250 & \infty & \infty \\ 0.7000 & 0.2000 & 0 & 0.2500 & 0.1667 & 0.3333 & 0.7500 & \infty & \infty \\ 1.2500 & 0.4500 & 0.2500 & 0 & 0.4167 & 0.5833 & 0.5000 & 1.5000 & \infty \\ 0.7500 & 0.2500 & 0.1667 & 0.4167 & 0 & 0.5000 & 1.0000 & 2.0000 & \infty \\ 1.3333 & 0.5333 & 0.3333 & 0.5833 & 0.5000 & 0 & \infty & \infty & \infty \\ \infty & 1.2500 & 0.7500 & 0.5000 & 1.0000 & \infty & 0 & 1.0000 & 1.1250 \\ \infty & \infty & \infty & 1.5000 & 2.0000 & \infty & 1.0000 & 0 & 0.1250 \\ \infty & \infty & \infty & \infty & \infty & \infty & 1.1250 & 0.1250 & 0 \end{bmatrix}$$

Figure 5:SN^2 Matrix obtained by applying Shimbel's method on structural matrix (SN) in figure 4

As, the matrix obtained in figure 5, is not optimal so applying Shimbel's method again.

$$SN^4= \begin{bmatrix} 0 & 0.5000 & 0.7000 & 0.9500 & 0.7500 & 1.0333 & 1.4500 & 2.7500 & \infty \\ 0.5000 & 0 & 0.2000 & 0.4500 & 0.2500 & 0.5333 & 0.9500 & 1.9500 & 2.3750 \\ 0.7000 & 0.2000 & 0 & 0.2500 & 0.1667 & 0.3333 & 0.7500 & 1.7500 & 1.8750 \\ 0.9500 & 0.4500 & 0.2500 & 0 & 0.4167 & 0.5833 & 0.5000 & 1.5000 & 1.6250 \\ 0.7500 & 0.2500 & 0.1667 & 0.4167 & 0 & 0.5000 & 0.9167 & 1.9167 & 2.1250 \\ 1.0333 & 0.5333 & 0.3333 & 0.5833 & 0.5000 & 0 & 1.0833 & 2.0833 & \infty \\ 1.4500 & 0.9500 & 0.7500 & 0.5000 & 0.9167 & 1.0833 & 0 & 1.0000 & 1.1250 \\ 2.7500 & 1.9500 & 1.7500 & 1.5000 & 1.9167 & 2.0833 & 1.0000 & 0 & 0.1250 \\ \infty & 2.3750 & 1.8750 & 1.6250 & 2.1250 & \infty & 1.1250 & 0.1250 & 0 \end{bmatrix}$$

Figure 6: SN^4 Matrix obtained by applying Shimbel's method on matrix SN^2 in figure 5

In order to obtain the optimal matrix we again apply the Shimbel's method on matrix SN^4, so the matrix SN^8 would be the optimal matrix as the number of connected nodes is 9.

$$SN^8= \begin{bmatrix} 0 & 0.5000 & 0.7000 & 0.9500 & 0.7500 & 1.0333 & 1.4500 & 2.4500 & 2.5750 \\ 0.5000 & 0 & 0.2000 & 0.4500 & 0.2500 & 0.5333 & 0.9500 & 1.9500 & 2.0750 \\ 0.7000 & 0.2000 & 0 & 0.2500 & 0.1667 & 0.3333 & 0.7500 & 1.7500 & 1.8750 \\ 0.9500 & 0.4500 & 0.2500 & 0 & 0.4167 & 0.5833 & 0.5000 & 1.5000 & 1.6250 \\ 0.7500 & 0.2500 & 0.1667 & 0.4167 & 0 & 0.5000 & 0.9167 & 1.9167 & 2.0417 \\ 1.0333 & 0.5333 & 0.3333 & 0.5833 & 0.5000 & 0 & 1.0833 & 2.0833 & 2.2083 \\ 1.4500 & 0.9500 & 0.7500 & 0.5000 & 0.9167 & 1.0833 & 0 & 1.0000 & 1.1250 \\ 2.4500 & 1.9500 & 1.7500 & 1.5000 & 1.9167 & 2.0833 & 1.0000 & 0 & 0.1250 \\ 2.5750 & 2.0750 & 1.8750 & 1.6250 & 2.0417 & 2.2083 & 1.1250 & 0.1250 & 0 \end{bmatrix}$$

Figure 7: Optimal matrix (SN^8)

In figure 7 the optimal matrix shows the minimum distance between each and every nodes.

2.4 Backward Tracking

To apply backward tracking, any two vertices are selected. One is assigned as source and other as destination. From the weighted sociogram in figure 1 node 1(Tod) is selected as source and node 9(Langdon) is set as target.

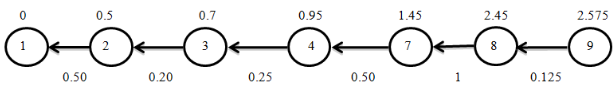

Figure 8: Backward tracking

The backward tracking in figure 8 defines that if for example Langdon(9) wants to approach Tod(1) for a piece of work who is not connected to him directly, then he have to approach Coley(8) and then Coley will contact Grant(7).Grant(7) is friends(connected) with Lee and Barney both but his interaction with Lee is more so he will ask Lee to pass on his message to Alston(3) who is friends connected with a lot of people in the network .Since , Alston(3) knows Tod(1) directly but they don't have a strong bond so Alston(3) would ask Merlin(2) to ask Tod(1) for the work Langdon(9) wants from him.

3. Average distance

The social network's graph defines human relationships and in the social network defined in figure (1) we have 10 people in which one is not in relation(connected) with anyone but all the other nodes are connected some are more socially active than the others. We will use Shimbel's method to find out that on average the people(nodes) in this network are how many "steps away". The term steps away shows that if one person wants to contact another person who is not directly connected to him than how many people would he need in between to pass on his message to the concerned person. The average distance in any network tells that the people in the network are how much closer (in contact) with each other.

The application of Shimbel's method was fine to use directly when we wanted to find out the shortest path between nodes (social actors).But if we want to find out average distance of the network in figure 1 then we have to normalize the weights by the average weight in the network.

Average weight in the network = sum of weights / no. of nodes = (1+2+5+4+2+4+6+1+1+3+8)/9
= 37/9 = 4.111

Now in order to find out the average distance among the nodes of the social network defined in figure 1, firstly the shortest distance among the nodes of the normalized distanced graph is calculated using Shimbel's method.

AD=

$$\begin{bmatrix} 0 & \frac{4.111}{2} & 4.111 & \infty & \infty & \infty & \infty & \infty & \infty \\ \frac{4.111}{2} & 0 & \frac{4.111}{5} & \infty & \frac{4.111}{4} & \infty & \infty & \infty & \infty \\ 1 & \frac{4.111}{5} & 0 & \frac{4.111}{4} & \frac{4.111}{6} & \frac{4.111}{3} & \infty & \infty & \infty \\ \infty & \infty & \frac{4.111}{4} & 0 & \infty & \infty & \frac{4.111}{2} & \infty & \infty \\ \infty & \frac{4.111}{4} & \frac{4.111}{6} & \infty & 0 & \infty & 4.111 & \infty & \infty \\ \infty & \infty & \frac{4.111}{3} & \infty & \infty & 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & \frac{4.111}{2} & 4.111 & \infty & 0 & 4.111 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & 4.111 & 0 & \frac{4.111}{8} \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \frac{4.111}{8} & 0 \end{bmatrix}$$

Figure 9: Structure matrix of normalized social network

Applying Shimble's method to get the optimal solution, Since the number of nodes are same so the optimal solution would be obtained after applying Shimbel's method thrice.

AD^8=

$$\begin{bmatrix} 0 & 2.0555 & 2.8777 & 3.9054 & 3.0832 & 4.2480 & 5.9609 & 10.072 & 10.586 \\ 2.0555 & 0 & 0.8222 & 1.8499 & 1.0277 & 2.1925 & 3.9054 & 8.0164 & 8.5303 \\ 2.8777 & 0.8222 & 0 & 1.0277 & 0.6852 & 1.3703 & 3.0832 & 7.1942 & 7.7018 \\ 3.9054 & 1.8499 & 1.0277 & 0 & 1.7129 & 2.3981 & 2.0555 & 6.1665 & 6.6804 \\ 3.0832 & 1.0277 & 0.6852 & 1.7129 & 0 & 2.0555 & 3.7684 & 7.8794 & 8.3933 \\ 4.2480 & 2.1925 & 1.3703 & 2.3981 & 2.0555 & 0 & 4.4536 & 8.5646 & 9.0785 \\ 5.9609 & 3.9054 & 3.0832 & 2.0555 & 3.7684 & 4.4536 & 0 & 4.1110 & 4.6249 \\ 10.072 & 8.0164 & 7.1942 & 6.1665 & 7.8794 & 8.5646 & 4.1110 & 0 & 0.5139 \\ 10.583 & 8.5303 & 7.7018 & 6.6804 & 8.3933 & 9.0785 & 4.6249 & 0.5139 & 0 \end{bmatrix}$$

Figure 10: Optimal matrix

Now, in order to find out the average distance between each node of the social network in figure 1, we will find out the average of the matrix in figure 7.(diagonal elements excluded)

Sum = 2.0555+2.8777+3.9054+3.0832+4.2480+5.9609+10.072+10.5858+0.8222+1.8499+1.0277+2.1925+ 3.9054+8.0164+8.5303+1.0277+0.6852+1.3703+3.0832+7.1942+7.7081+1.7129+2.3981+2.0555+6.1665+6.6804+2.0555+3.7684+7.8794+8.3933+4.4536+8.5646+9.0785+4.111+4.6249+0.5139

Sum = 162.6583
Mean=sum/36
=4.5182

The average of the matrix suggests that on-average nodes are 4.5182 steps away from each other; i.e. approximately 5 steps. This shows that each person in the network can reach the other person using an acquaintance chainof 5 intermediaries.

4. Comparison with other method

In this section Djikstra's algorithm is applied to find out the shortest path of thesocial networks graph defined in figure 2.In order to compare the results obtained by Dijkstra'smethod to the results obtained by Shimbel's method.

Dijkstra's algorithm is a single source shortest path algorithm so selecting node 1 as source, the following results are obtained.

Shortest distance between node 1 and node 9 is 2.5750 and the shortest path is 1-2-3-4-7-8-9
Shortest distance between node 1 and node 8 is 2.4500 and the shortest path is 1-2-3-4-7-8
Shortest distance between node 1 and node 7 is 1.4500 and the shortest path is 1-2-3-4-7
Shortest distance between node 1 and node 6 is 1.0333 and the shortest path is 1-2-3-6
Shortest distance between node 1 and node 5 is 0.7500 and the shortest path is 1-2-5
Shortest distance between node 1 and node 4 is 0.9500 and the shortest path is 1-2-3-4
Shortest distance between node 1 and node 3 is 0.7000 and the shortest path is 1-2-3
Shortest distance between node 1 and node 2 is 0.5000 and the shortest path is 1-2

Similarly other nodes are selected as source to find out the shortest distance and shortest path. After applying Dijkstra's algorithm on every node of figure 1,the following result is obtained:

$$\begin{bmatrix} 0 & 0.5000 & 0.7000 & 0.9500 & 0.7500 & 1.0333 & 1.4500 & 2.4500 & 2.5750 \\ 0.5000 & 0 & 0.2000 & 0.4500 & 0.2500 & 0.5333 & 0.9500 & 1.9500 & 2.0750 \\ 0.7000 & 0.2000 & 0 & 0.2500 & 0.1667 & 0.3333 & 0.7500 & 1.7500 & 1.8750 \\ 0.9500 & 0.4500 & 0.2500 & 0 & 0.4167 & 0.5833 & 0.5000 & 1.5000 & 1.6250 \\ 0.7500 & 0.2500 & 0.1667 & 0.4167 & 0 & 0.5000 & 0.9167 & 1.9167 & 2.0417 \\ 1.0333 & 0.5333 & 0.3333 & 0.5833 & 0.5000 & 0 & 1.0833 & 2.0833 & 2.2083 \\ 1.4500 & 0.9500 & 0.7500 & 0.5000 & 0.9167 & 1.0833 & 0 & 1.0000 & 1.1250 \\ 2.4500 & 1.9500 & 1.7500 & 1.5000 & 1.9167 & 2.0833 & 1.0000 & 0 & 0.1250 \\ 2.5750 & 2.0750 & 1.8750 & 1.6250 & 2.0417 & 2.2083 & 1.1250 & 0.1250 & 0 \end{bmatrix}$$

Figure 11: Result obtained by using Dijkstra's algorithm

The result obtained by applying Dijkstra's algorithm in figure 11 is same as the result obtained by Shimbel's method in figure 7.The benefit of Shimbel's method is that if at any point we change the source and the destination the same working would be carried on. In Djikstra's algorithm if the source and the destination is changed in between then the working has to be done from the beginning.

Conclusion

Social networking has turned out to be a vital part of our daily life, as it facilitates us to interconnect with people from all around the world. Social networking sites are made to provide help in online networking. The social networking sites are created to provision a common theme. The Shimbel's method is applied to measure the shortest path of the social networks graph and also applied to find out the average distance between the people (nodes).

We found that on average the nodes are not more than 5 steps away from each other, in other words we can say that in this social network's graph the degree of separation is "five". The result obtained by using Shimbel's method is also compared to the result obtained by using Dijkstra's method and both methods give the same results.

References

[1]. Wasserman, Stanley; Faust, Katherine (1994): Social Network Analysis: Methods and Applications, Cambridge University.

[2]. D'Andrea, Alessia et al. (2009) :An Overview of Methods for Virtual Social Network Analysis.

[3]. Robert A. Hanneman (2005): Introduction to social network methods ,Department of sociology ,University of California.

[4]. KoushaEtessamiv (2014): Discrete Mathematics and Mathematical reasoning, University of Edinburgh UK

[5]. Wasserman and Fraust(1994):Social network analysis,Methods and applications,Cambridge university press.

[6]. Jeffrey Travers, Stanley Milgram (dec.,1969) :An experimental study of the small world problem , sociometry ,volume 32 ,issue 4.

[7]. M.E.J. Newman (May 2001):Scientific collaboration networks 1 , Network construction and fundamental results,physical review E, volume 64, 016131

[8]. M.E.J. Newman (June 2001) : Scientific collaboration networks 2 ,shortest paths ,weighted networks and centrality.,physical review E, volume 64, 016132.

[9]. A. Shimbel (1955): Structure in communication nets, in: Proceedings of the Symposium on Information Networks (New York, 1954), Polytechnic Press of the Polytechnic Institute of Brooklyn, Brooklyn, New York, pp. 199–203.