Research on Stock Network Modeling and Risk propagation based on Complex Network Analysis

Xin Jin Ying Wu Chu Jin
Central University of Finance and Economics, BeiJing, PR China, 100081
xin_jin@163.com

Abstract

The price fluctuation in the stock market is often accompanied by risks, the spread of which affects the entire financial system and even the stable operation of the social economy as a whole. Considering the relatively mature research for the network characteristics of the stock market, this paper attempts to employ the complex networks analysis method to study the risk of the stock market. We firstly analyze the determinants of the stock influence (SI), put forward the measurement method of SI and thus establish the risk propagation model. On that basis, we propose GASI, the influence maximization algorithm for stock network, and dig out the key nodes that maximize SI. Experimental results show that price fluctuations of a small number of stocks may largely impact the other stocks, even the entire stock market. These influential stocks distribute in almost all industries, but are mainly in finance, manufacturing and mining. In addition, compared to the PageRank algorithm, the GASI algorithm can influence more nodes and widen the range of risk propagation, and effectively identify key stock nodes with greater risk influence from the stock network.

1. Introduction

Since the 21st century, the trend of economic globalization is increasingly apparent, the flows of international capital are accelerating and the financial market is increasingly international. Stock market, bond market, banks and other financial institutions link through the circulation and exchange of the asset value, and thus constitute a complex and real-time changing financial system. Financial system is affected by many factors, making the overall behavior of the financial system difficult to predict and resulting many potential risks in national economic activities. Stock market has a vital position in the financial system because not only volatility in stock market reflects the operation of the overall social economic, but also modern finance treats it as a risk measure. Abnormal price fluctuations of stocks, partial or individual, may cause Domino-like contagion effects, resulting in the systematic risk in the stock market. With the advent of the era of big data, mass data of finance and news can be achieved through the Internet. And how to use these data effectively to assist the financial risk analyze becomes a meaningful cross research direction between technology and economics research.

Long time has passed since the domestic and foreign scholars studied the risk of the stock market. Scholars in many fields, such as finance, econometrics, statistics, mathematics, etc. constantly put forward various theories, models and methods of the reasons of market risk, risk measurement, risk propagation, etc. Complex networks use graph structure for complex system modeling and use graph theory and heuristics to find meaningful kinetics in the complex system. With the continuous improvement of complex network theory and graph theory, more and more scholars try to use the complex network analysis method and heuristic strategy to study the stock market recently.

For this paper, research motivation is: whether we can build a complex network structure of the stock market system by the collected financial data, and use quantitative method to evaluate the influence power of the price fluctuation of nodes on that basis. Through analyzing the transmission of fluctuation, the transmission rule of the financial risk is detected and the key nodes in the process of risk propagation are dug out.

The main contributions of this paper are:

(1) To define the SI, analyze determinants of price fluctuations in stock market and thus put forward the measurement method of SI.

(2) To put forward the risk propagation model of stock network based on the SI model and linear threshold model, and then simulate the risk transmission rule caused by price fluctuations in the network.

(3) To put forward an improved greedy algorithm to find the key node that maximizes the influence of risk propagation based on the proposed model.

1530-1605/16 $31.00 © 2016 IEEE
DOI 10.1109/HICSS.2016.150

2. Relative Works

2.1. Research on the stock network

Plenty of stocks are in the stock market. The price changes, trading volume volatility and new shares issuing create a complex relationship, which constitute the stock network after abstraction. In the study of the stock network, network construction methods can be roughly divided into two categories: minimal spanning tree method (MTS) and the threshold method. Mantegna [1] defines the distance between nodes based on correlations coefficient between stocks and builds financial network by MTS the first time. Garas A et al. [2] investigate the topological property of the stock network through removing the weak and strong links. They find that the weak links contribute to the overall connectivity of network significantly more than the strong links, but the strong ones allow the nodes to form stable communities. Junghoon Lee et al. [3] establish stock network by using intraday volume data in Korean stock market and study the relationship between market volatility and network properties. They find a strong negative correlation between the market volatility and the network density.

All studies above are based on the assumption that the time series of closing price represent the behavior of the each stock. But the assumption ignores the behavior information contained by volumes and other data. Therefore, Brida and Risso [4] put forward the use of symbolic time series analysis (STSA) to construct minimal spanning tree that each stock is represented by binary time series of volumes and closing prices. However, the method applies only to bivariate time series, rather than multidimensional ones. Gan Siew Lee and Maman A Djauhari [5] propose the multidimensional minimal spanning tree method (Multidimensional MST, MMST) to build stock network. They use signature sequences of opening, closing, high and low price to represent stocks and then calculate the distances according to the RV coefficient of the stock matrix.

The threshold method is widely applied to the establishment of stock network and is easier than MTS. Zhihui Yang and Hanmei Jia et al. [6] establish stock correlation network based on the correlations coefficient and find that the network has small-world property and non-scaling property. Also, they consider that the small amount of dominant stocks influence and reflect the entire market. Lingyan Wu et al. [7] construct the Shanghai stock market network model on the basis of correlation coefficient and optimum threshold value. She then verifies the stability of the topological structure of the constructed network from the perspectives of network node, community structure and structural motif.

2.2. Research on risk propagation

Financial risk is the probability of deviation of the financial act from its expected goal and this uncertainty brings both benefit and loss. Affected by various factors, financial risk will spread in the financial market, trigger large-scale price fluctuation and even destroy the stability of the financial system. Financial contagion refers to the spread of price fluctuation from one or more markets to the other, making a co-movements process in asset prices in other markets.

It is generally considered that risk propagation is caused by two factors: fundamentals and non-fundamental factors. Fundamental connections between markets, such as common economic conflicts and trade links, make dependency between markets. Non-fundamental factors refer to the factor that is unrelated to the observed macroeconomic and fundamental factors and is mainly by liquidity problems, information asymmetry, risk preference, portfolio changes and the other shocks. This kind of risk propagation is called a true contagion [8].

The existence testing methods mainly consist of the correlation coefficient method, the co-integration test, the causality test and the volatility spillover analysis. Correlation coefficient can mainly analyze correlation coefficient changes before and after the occurrence of market impact [9], but cannot prove the causality of the risk propagation. Co-integration only focuses on the long-term market equilibrium relationship of risk propagation, but cannot reflect the short-term relationship. Mean volatility spillover analysis commonly uses VAR model to test the price volatility transmission between markets. Typically, the VAR model and causality analysis are combined to analyze risk propagation effect and if one party is the other one's Granger cause, the risk spreads from the former to the latter. Volatility spillover analysis mainly applies to GARCH models to study the transmission of price volatility. The latter two methods overcome the disadvantages of the first two and are more widely used. Yi Fang et al. [10] analyze the risk propagation in domestic and international metals futures markets based on the BEKK model. They find that not only reactions of the domestic metal futures market to macroeconomic shocks are asymmetrical, but also the macroeconomic factor is not the major internal reason of the volatility spillover.

Current studies on financial risk propagation are all about the contagion effects among different countries or markets, using varieties of econometric methods and models to explore the existence and direction of the risk propagation. These studies, however, only include that risk transmit from one market to the other one in the macroscopic level and lack studies on transmission rule of the internal market.

2.3. Research on maximizing of complex networks

With the emergence and development of complex network analysis models, applications and research of influence of complex network are more interested in recent years. Influence maximization problem [11], in particular social networks (such as Facebook, Twitter, Microblog, etc.), is one of the hot-button issues in the complex networks research. Maximizing impact of complex network is to find the set of most influential k nodes that can disseminate information to the rest of the network in the most extent. The rule that information transmits in complex networks is called an information diffusion model. The most basic models are linear threshold model, independent cascade model and weighted cascade model.

Kempe and Kleinberg et al. [12] prove that the influence maximization problem is NP-complete, and put forward a hill-climbing greedy algorithm. Although the algorithm guarantees the optimal solution within 1-1/e, it is not practical according to the complexity on analyzing large-scale social network. Peiyun Zhang et al. [13] expand the basic linear threshold model and put forward the impact-maximizing algorithm based on the probability transfer matrix. The probability of influence between nodes at a certain time can be obtained by matrix multiplication method. Also, the algorithm has low time complexity and suitable for large-scale complex network. Kai Wu et al. [14] put forward measurement of user influence WIR algorithm by summarizing the key factors that determine the influence of Microblog users. They use greedy algorithm to dig out the Top-K nodes.

The node-connected edges in social network represent specific relation, such as mutual concern relations in Microblog, and influence power is measured generally by degrees. However, the node-connected edges in stock market do not exactly mean real links, such as assets and liabilities, mutual guarantee and other relations, and more mean a "up and down together" phenomenon of the stock price. Thus, we cannot use node degrees solely to measure the impact of stock but need to consider other features. Therefore, this paper analyzes and quantifies the key factors of SI in order to put forward the measure method, and then study the risk influence maximization problem in stocks networks.

3. Establish the stock Network Model

For a limited stock network, we abstract it to an undirected weighted network model G = <V, E>. The node set V is the set of all nodes in the network share (stock), the edge set E is the mutual influence relationship of stocks. The correlation coefficient of any two stock nodes i and j is:

$$\rho(i,j) = \frac{E(R_iR_j)-E(R_i)E(R_j)}{\sqrt{Var(R_i)Var(R_j)}} \quad (1)$$

Where $R_i=(r_{i1}, r_{i2}, \cdots, r_{in})$ is the return vector of stock i, $r_{ik}=\ln P_i(k)-\ln P_i(k-1)$ is the first k logarithm return of stock i, $P_i(k)$ is the closing price of the stock i at day k. By definition, the interval of the correlation coefficient $\rho(i,j)$ is [-1,1].

After getting the correlation coefficient matrix of the stock network, screening the appropriate correlation coefficients to determine the edge of the network at a given threshold θ. When $|\rho(i,j)| \geq \theta$, there is an edge between nodes i and j and thus get the edges set of stock E of stock network. The threshold is inversely proportional to the number of edges in the network. Within a certain interval, the number of edges shows a significant reduction trend with the increase of threshold. While without this interval, the impact of thresholds is small because the stock distribution of stock network is either over sparse or dense and thus is not suitable for further study. Therefore, the optimal threshold must lies within this interval. In this paper, the optimal threshold is a value on which basis we can build a stable network topology that has large clustering coefficient, small-world property, non-scaling property and more nodes of maximum connected subgraphs.

4. Research on Risk propagation in the Stock Network

4.1 Risk propagation Model

Risks in stock network diffuse with the spread of price fluctuation. In fact, price fluctuations propagate from one node to the other along the edge in the form of information. If information could impact the node, the price of the node will fluctuate and risk will emerge. Therefore, we will study the risk propagation model based on the information diffusion model of social network.

This paper uses independent cascade model on the basis of SI to simulate risk propagation process. At initial, all nodes in stock network are in a stable state, but a sudden fluctuation of a node may affect its neighbor nodes from stable to unstable. And the probability of success is linked to the influence power of nodes, the greater the power the more likely the activation. For a stable node with several neighbor nodes, the impact of these neighbor nodes is in a random order and is assumed not to be addictive.

Essentially, SI is the extent of effect to the other stocks. The more extent of effect a stock has, the larger the SI is. Social networks influence refers to the information transmission ability and a node has larger influence if it can transmit the information to more nodes. Measured by the degree of the node generally, the influence is not fully applicable to stock network. This paper defines the SI in stock network as follows:

Definition1: *Stock Influence (SI)* in stock network refers to the potential of stock to trigger the return fluctuation of the other stock when its price fluctuates.

According to the definition above, this paper proposes key factors that determine SI are:

1) The importance of stocks. Stock position in the market and industry reflects the importance of the stock. The larger the circulation value, the more important the stock is, and thus its price fluctuations is more likely to result in price fluctuations of other stocks.

For the stock network G = <V, E>, the importance of node i is defined as:

$$\mathrm{Imp}(i) = \frac{P_i}{2} * \left(\frac{CS_i}{\sum_{k \in A} P_k * CS_k} + \frac{CS_i}{\sum_{j \in V} P_j * CS_j}\right) \quad (2)$$

Where CS_i is the outstanding capital stock of stock i, P_i is the latest closing price of the stock i, A is the set of all stock in i's industry.

2) The correlation between industries. The more closely industries where the two stocks belong are, the more likely the interplay of price fluctuation exists.

For example, the stock price fluctuations of a listed company will often affect investors' evaluation and forecast of other companies in the same industry, and thus affect the stock price of other company.

According to the expression of correlation coefficient (1), the correlation coefficient between industry A and B is defined as follows:

$$\mathrm{R}_{AB} = \rho\left(\frac{\sum_{i \in A} R_i}{\mathrm{num}(A)}, \frac{\sum_{j \in B} R_j}{\mathrm{num}(B)}\right) \quad (3)$$

3) The correlation between stocks. The correlation between stocks is the degree of long-term consistency on two stocks' return volatility and the correlation is determined by the integration of many factors such as cooperation partnership.

The degree of correlation between stocks is represented by the correlation coefficient $\rho\,(i,j)$ mentioned above.

Thus, the $\mathrm{SI}(i,j)$ indicates the degree of influence on the stock j by i and the specific expression is as follows:

$$\mathrm{SI}(i,j) = \mathrm{Imp}(i) * \mathrm{R}_{AB} * d + \rho_{ij} * (1 - d) \quad (4)$$

Where d is the adjustment factor that can be chosen in the interval (0,1). Stock i belongs to the industry A , while stock j belongs to the industry B.

4.2 Risk influence maximization algorithm based on the SI

In stock networks, the key node refers to the node that causes the price fluctuations among the most nodes when the price changes itself. In another word, it has the greatest risk diffusion range. In general, the degree of a node can be used to measure its position in the network. However, achieving the maximum influence range does not mean selecting the maximum degree as the initial node directly. Because of the industry characteristics of stocks, nodes will converge into several groups and will more likely to interconnect and gather in one group if they belong to the identical industry. Thus, selecting the large degree nodes only will ignore some key nodes with small degree that link different groups in the network. As a result, we need to consider the special character of the stock network to design the appropriate influence maximization algorithm.

The SI measure method reflects the characters of stock market and makes the process of information diffusion in stock network more realistic, while the greedy algorithm is a typical algorithm that solves the influence maximization problem in the complex network. For the characteristics of the stock network, this paper proposes Greedy Algorithm on SI (GASI) to solve the influence maximization issue.

The core processes of GASI are:

1) To establish the SI matrix by calculating he influence of SI (i, j) for each node based on equation (4).

2) To establish the information diffusion model of the stock network by the SI matrix of nodes.

3) On the basis of independent cascade model of stock network, selecting the node that has the largest increment influence scope step by step by the method of the greedy algorithm, and getting the set of seed nodes eventually.

The GASI algorithm pseudo-code is follows as:

Algorithm: GASI algorithm

Input: Stock network G = <V, E>, the number of seed node K
Output: set of seed node S

```
BEGIN
  Initialize S=∅
  For each edge(i, j) In E Do
    Calculate SI(i, j)
  End for
  For i = 1 To K Do
    For each vertex v In V/S Do
      s_v = 0  # the influence scope of node set S∪{v}
      For j= 1 To R Do #R is the number of nodes in V
        s_v += Calculate_influence(S∪{v})
      End for
      s_v /= R
    End for
    S = S ∪ {max_v {s_v} }  # select the node v that has the largest increment influence scope as a new seed node
  End for
END
```

5. Simulation experiment

To establish stock network, the paper selects 280 constituent stocks in CIS 300(ShangHai ShenZhen 300 Index) as network nodes. We collect closing prices from June 2012 to June 2014 and gather the outstanding capital and industries of these stocks as sample data. We remove data in every Saturday, Sunday and holiday when trading does not occur. In addition, individual stocks may have shorter stop trading period, so we view the daily returns of these stock during this period to be zero. Experimental data is collected from the Shenzhen Stock Exchange http://www.szse.cn/ and RESSET financial research database http://www.resset.cn/cn/.

5.1 Construction of Stock Networks

Relations between thresholds and the number of edges in stock network are shown in Figure 1. It shows that the optimal threshold should be within the interval [0.2,0.55]. To further simplify the calculation, the optimal threshold interval is defined in the interval [0.44,0.5], in which the number of edges changes within the range of 3290~ 6239, and thus the size of stock network is relative moderate. Then, we draw Stock networks under different thresholds by Pajek software, compute the number of nodes and clustering coefficient for each maximum connected subgraphs and plot the line graph in Figure 2. As shown in Figure 2, when the threshold is between 0.45 and 0.46, the changes of the number of nodes and clustering coefficient are relatively flat. Also, within this interval, the degree distribution of nodes follows the power-law distribution. Thus, the optimum threshold is confirmed to be 0.46. The constructed stock network is shown in Figure 3.

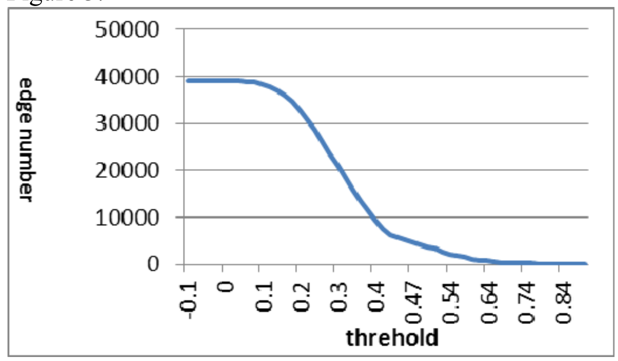

Figure 1. Line graph of the edge number and thresholds

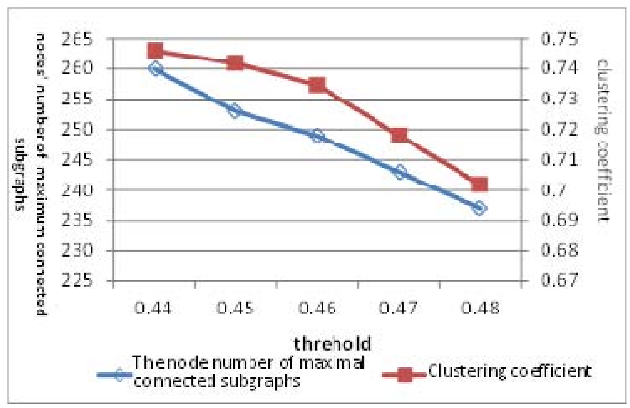

Figure 2. Line graphs of thresholds, the node number of maximal connected subgraphs and the clustering coefficient

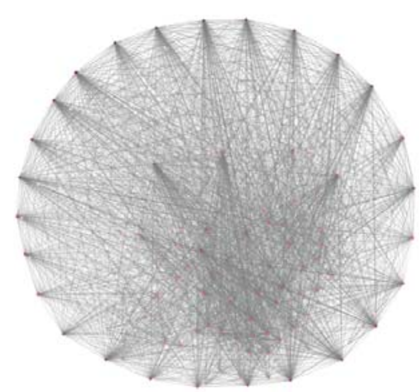

Figure 3. Stock Network (threshold is 0.46)

5.2 Results

When we use GASI to conduct influence maximization analysis for the constructed networks above, we find that the first 20 seed nodes cover a total of 10 industries, including finance, manufacturing, services, mining, construction industry, transportation, press publishing, real estate, hydropower industry and comprehension industry. Among them, the number of stocks that belong to finance, manufacturing, mining, construction industry accounts for 65% of the seeds set, and their total market capitalization has a great proportion in the CSI 300 constituent stocks. This suggests that in the Shanghai and Shenzhen stock market, a small number of stocks can impact largely on others and can even bring price fluctuations to the entire market. These influential stocks spread almost all industries, but mainly distribute in finance, manufacturing and mining, which is fairly consistent with the reality.

Table 1. The Seed Node List (K = 20)

Industry	Securities in short
Finance	BOC, CTB, China Life, CITIC Securities, Haitong Securities.
Manufacturing	TCL, Qingdao Haier, Foton, Tong Ren Tang, Wuliangye
Mining	CNPC, Pingdingshan Tianan Coal Mining , Zijin Mining
Real Estate	Poly Real Estate
Construction	China Building Construction
Hydropower Industry	Yangtze Power
Transportation	Hainan Airlines
Services	NARI
Press Publishing	China South Publishing & Media Group
Comprehensive Industry	Zhangjiang Hi-Tech Park Development

In the analysis of influence maximization algorithm in complex networks, we generally assess the scope of effectiveness and availability based on the influence scope in the information diffusion model of the ultimately selected seed sets. To further verify the effectiveness of the GASI algorithms in solving the influence maximization problem, this paper compares the transmission ranges between GASI algorithms and PageRank algorithm. PageRank algorithm is a common influence measurement algorithm that calculates the PageRank value of each node through link relations between nodes. The influence power of one node to the others is the equally distributed degree of the node. When the number of selected nodes seeds increase gradually, calculating the influence scope of GASI algorithm and PageRank algorithm respectively. The result is shown in Figure 4.

From Figure 4, the seed nodes selected by GASI algorithm have wider influence scopes than by PageRank algorithm. And the number of nodes selected by two algorithms has a difference of about 9%. It suggests that GASI algorithm, which combines the SI, simulates the transmission process of price fluctuations effectively and identifies more risk influential stock nodes. In addition, the fact that GASI outperforms the PageRank algorithm suggests that the SI, which is determined by many factors together, cannot be simply measured by degrees. Thus, introducing the SI solves the risk propagation problem better.

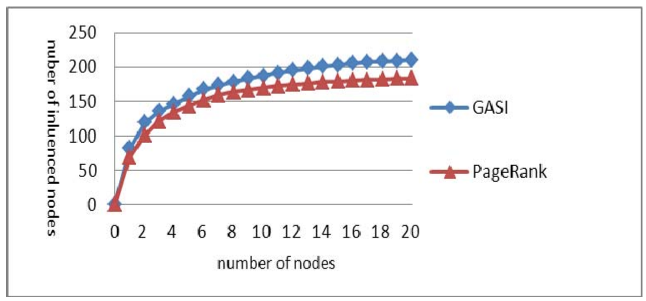

Figure 4. The Influence scope of GASI algorithm and PageRank algorithm

6. Conclusion

In this paper, we establish a network of Shanghai and Shenzhen stock markets by using the correlation coefficient and the threshold, define the SI and provide the measure method. We then establish a risk propagation model on basis of the SI and propose GASI algorithm for the risk influence maximization problem. Experimental result shows that, seed nodes selected by the GASI algorithm cover most of industries, especially in finance, manufacturing and mining, indicating that the three sectors occupy the important positions in the stock risk propagation process. We further find that when compared to the PageRank algorithm, the GASI algorithm can effectively identify stock nodes with more transmission influence and has a wider influence scope. Financial system is a complex and dynamic system with a complex relationship between stocks and their agencies. Therefore, through constructing a risk propagation model of stock network to analyze the risk diffusion mode of nodes, it is beneficial for assessing the scale and influence of the potential financial risk and it is helpful for analyzing patterns and rules of financial risk propagation in the financial market. Also, it has a certain practical significance of grasping the dynamic changes of the financial network.

However, the risk propagation model and algorithms digging out the key nodes proposed in the paper still need further improvement and research. How to improve the linear threshold model to simulate the risk propagation in stock network is our further work.

7. References

[1] R N Mantegna. Hierarchical structure in financial markets[J]. The European Physical Journal B, 1999, 11(1): 193-197.

[2] Garas A, Argyrakis P, Havlin S. The structural role of weak and strong links in a financial market network[J].European Physical Journal B,2008,63(2):265-271.

[3]Junghoon Lee, Janghyuk Youn, Woojin Chang. Intraday volatility and network topological properties in the Korean stock market [J]. Physica A,2012,4(391): 1354-1360.

[4]Brida, Juan Gabriel & Risso, Wiston Adrián, 2008. "Multidimensional minimal spanning tree: The Dow Jones case," Physica A: Statistical Mechanics and its Applications, 387(21) : 5205-5210

[5] Gan Siew Lee, Maman A Djauhari. Multidimensional stock network analysis: An Escoufier's RV coefficient approach[J]. AIP Conference Proceedings,2013,1 (1557):550-555.

[6] Zhihui Yang, Hanmei Jia. Research on Network structure of the correlation between stock returns[C]// Proc of Chinese Association of automation control theory professional committee, Chinese Association of automation control theory Committee B volume. Beijing: Science Press, 2011: 532-5736.

[7] Lingyan Wu, Hua Han, Ningning Song. The construction of stock network model based on correlation coefficient and optimal threshold[J]. Complex system and complexity science, 10, 2013 (4): 49-55.

[8]Kaminsky, G. & C. Reinhart. On crises, contagion, and confusion[J]. Journal of International Economics, 2000, 51(1):145-168.

[9] Carvo, Sarah and Reinhart. Capital flows to Latin America: Is there evidence of contagion effects? [C] .The World Bank policy research working paper, 1996.

[10] Yi Fang, Yishan Zhang. Empirical research on the risk propagation of domestic and international metal futures markets [J]. Financial research, 2007(5): 133 一 145.

[11]Wang Yue, Huang WeiJing, Zhong Lang. Influence maximization with limit cost in social network[J]. Science China, 2013,56(7):1-14

[12]Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network[C]//KDD 03:Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York:ACM, 2003:137-146.

[13] Peiyun Zhang, Xiuwen Gong. Influence maximization algorithm in social network based on probability transfer matrix [J]. Computer Engineering, 2013, 39 (11): 41-45.

[14] Wu Kai, Xinsheng Ji, Jinshi Guo, Caixia Liu. Influence maximization algorithm on Microblog network [J]. Journal of computer applications, 2013, 33(8) : 20191-2094.