

Co-Movement in Stock Markets Based on Directed Complex Network

Zhen Du¹, Pujiang Chen¹, Na Luo¹ & Yingjie Tang¹

¹ School of Jinan University, Guangdong, China

Correspondence: Pujiang Chen, School of Jinan University, Guangdong, China. Tel: 86-137-5007-5014. E-mail: 316592902@qq.com

Received: December 5, 2015

Accepted: December 22, 2015

Online Published: January 25, 2016

doi:10.5539/ijef.v8n2p179

URL: <http://dx.doi.org/10.5539/ijef.v8n2p179>

Abstract

In this paper, directed complex network is applied to the study of A shares in SSE (Shanghai Stock Exchange). In order to discuss the intrinsic attributes and regularities in stock market, we set up a directed complex network, selecting 450 stocks as nodes between 2012 and 2014 and stock yield correlation connected as edges. By discussing out-degree and in-degree distribution, we find essential nodes in stock network, which represent the leading stock. Moreover, we analyze directed average path length and clustering coefficient in the condition of different threshold, which shows that the network doesn't have a small-world effect. Furthermore, we see that when threshold is between 0.08 and 0.15, the network follows the power-law distribution and behaves scale-free.

Keywords: directed complex network, stock market, small-world effect, scale-free property

1. Introduction

Due to the effect of guiding fund and risk transfer and so on, financial market plays a crucial role in national economy. For a developing economy, a healthy and stable financial system will give it soaring wings, and therefore it is significant to study the intrinsic regularities of financial market. In recent years, the applications of complex network to financial market have grown to be scholastic hotspots, and the orientation of its research has gradually developed from undirected to directed way. Given that the financial market shares many similarities with the complex system, including dynamics, nonlinearity, self-similarity, aperiodicity, self-organization, sensitivity and so forth, the complex system can be applied to the studying of financial market suitably. Various applications of complex network have been so for quite some time. Mantegna RN (1999) firstly made a linkage between stock market and complex network. By using the stock price data, the model established a stock correlation network, with individual stock as node and correlation coefficient as edge, which made clustering analysis of the S&P 500. Boginski (2005) and others established a complex network model with 6546 stocks in ASE (American Stock Market). It turned out that the network shows the characteristic of no scale. Moreover, Onnela J P and others (2006), built dynamic asset trees to study the correlation of 477 stocks in NYSE (New York Stock Exchange). A Scale-free weighted complex network is constructed and applied to the S&P 500 stocks (see Kim & others, 2007 for more information). This network model calculated the financial influence value of individual stock by summing all weight values of individual node's edges, and confirmed its absolute value has scale-free property.

For the present time, few of the domestic studies of economy were able to focus on the financial market compared to foreign countries, yet still some of the researchers managed to study the Securities Market with the complex network theory. Zhuang Xintian and others (2007) use the complex network, to study SSE (Shanghai stock exchange) and came to a conclusion that it has scale-free property as well as small-world effect. Complex network models combined with coarse graining analysis has emerged as a natural and convenient approach to the study on Hong Kong Stock Market (see Fang Weidong, 2008). The network model, researching the correlation between Hang Seng Index and Trading Volume, found the crucial nodes in stock market by calculating the betweenness centrality and inverse participation ratio. Gao Yachun (2013), who established a complex network with financial dealers and products as nodes, studied the static topology and dynamic evolution of financial market. Furthermore, complex network was also applied to Interbank market and futures market to judge whether it is scale-free and has small-world effect (see, for example, Liu Chao & Zhang Ding). According to the result, the former market has features of both and the latter one is equipped with small-world effect and assortativity except scale-free property.

In the past, previous researches about complex network in financial market are basically limited to undirected field. Chen Hua (2012), who built directed complex network successfully with A share in SSE, selected high frequency data after March, 2011. In this paper, by investigating daily stock data from 2012 to 2014, we firstly established directed complex network to study the market with the help of MATLAB. Then we mainly discussed structural property of the network, including the out-degree, in-degree, degree distribution, average path length, and clustering coefficient. In particular, we modified the statistic method of Chen about the structural property we mentioned above. Finally, according to related methods of undirected complex network, we judged whether it has scale-free property and small-world effect.

2. The Construction of Directed Complex Network in Stock Market

2.1 Stock Yield Rate

The logarithmic yield of stock a can be calculated by the formula below:

$$R_{\Delta t}^a(t) = \ln \frac{P_a(t)}{P_a(t-1)}$$

Where, $R_{\Delta t}^a(t)$ is the logarithmic yield rate of stock a from period $t-1$ to period t , $P_a(t)$ is the close price of stock a at time t , and Δt represents the time interval of logarithmic yield rate of stock.

2.2 Stock Correlation Coefficient

2.2.1 The Formula of Correlation Coefficient

Here a and b are two randomly chosen stocks, and Δt is the time interval. Since we are going to carry out our analysis based in time interval, we define stock correlation coefficient as:

$$Q_{\Delta t}^{ab}(\tau) = \frac{E(R_{\Delta t}^a(t)R_{\Delta t}^b(t+\tau)) - E(R_{\Delta t}^a(t))E(R_{\Delta t}^b(t+\tau))}{\sqrt{\text{var}(R_{\Delta t}^a(t))\text{var}(R_{\Delta t}^b(t+\tau))}}$$

Where $Q_{\Delta t}^{ab}(\tau)$ represents the correlation coefficient which measures the influence of price of stock b arising from stock a after interval τ ; function $E(x)$ and $\text{var}(x)$ stand for the mean values and variance of variable x respectively; τ is the interval when calculating correlation coefficient. (attention should be paid to distinguish τ and Δt : Δt is the interval when calculating rate of return).

Through calculation we can get correlation coefficients of every two stocks and then transform them into numerical matrix.

2.2.2 The Implication of Correlation Coefficient

Firstly, the values of correlation coefficient are between -1 and 1, i.e. $-1 < Q_{\Delta t}^{ab}(\tau) < 1$.

If $0 < Q_{\Delta t}^{ab}(\tau) < 1$, it means that there is a positive correlation between stock a and stock b , namely, the fluctuation of price a will make the price of b fluctuate in the same direction; similarly, if $-1 < Q_{\Delta t}^{ab}(\tau) < 0$, it means that there is a negative correlation between stock a and stock b , and the fluctuation of price a will make the price of b fluctuate in the opposite direction; moreover, $Q_{\Delta t}^{ab}(\tau) = 0$ implies that there is no correlation between a and b , and they won't interact each other.

The interaction relationship between stock a and stock b has direction as its characteristics. In the numerical matrix of correlation coefficient, if $|Q_{\Delta t}^{ab}(\tau)| < |Q_{\Delta t}^{ba}(\tau)|$ implies the direction flow is stock b to stock a , there is a directed edge from b to a in the directed network graph; if $|Q_{\Delta t}^{ab}(\tau)| > |Q_{\Delta t}^{ba}(\tau)|$, it means the direction flow is from the stock a to stock b , and similarly in the directed network graph the directed edge is from a to b ; if $|Q_{\Delta t}^{ab}(\tau)| = |Q_{\Delta t}^{ba}(\tau)|$, it means that there is no direction flow between stock a and stock b , therefore there is no directed edge in directed network graph between the two stocks. From the analysis above we can set up a complete directed complex network in stock market.

2.3 The Threshold

In normal cases, threshold should be specified. In fact, when the threshold value exceeds the absolute value of

correlation coefficient of stock a and stock b , we believe that their interaction is so weak that can't make clear influence. We define θ as the threshold value, and suppose that there is no directed edge from a to b in the complex stock network when $|Q_{\Delta}^{ab}(\tau)| < \theta$.

3. The Structural Property of Directed Complex Network in Stock Market

It should be noted that as a complex system, it possesses many structural properties, and therefore, only by using different parameters we can mine the data as much as possible and find the inwardness as well as the law, which can guide us to practice. However, structural properties in directed complex network are different from undirected ones. For undirected complex network and its structural properties, there are a greater number of researches. So in this paper we only focus on directed complex network. Refer to Chen, in what follows we define several indexes to study directed complex network.

3.1 Out-Degree and In-Degree

In directed complex network, the out-degree of node a of a graph is the number of edges that start from node a to the other nodes. Similarly, the in-degree of node a of a graph is the number of edges that start from the other nodes to node a .

3.2 The Mean Value of Degree

In the following part we define the mean value of out-degree of all nodes as the mean out-degree; similarly, the mean in-degree is the mean value of in-degree of all nodes in the graph.

3.3 Interval Degree Distribution

In the study of graphs and networks, the degree k of a node in a network is the number of connections it has with other nodes and the degree distribution is the probability distribution of these degrees over the whole network. Note that there are 450 nodes in our study, and the number of different degrees has reached 225. What's more, the value of each degree is so small that it is not significant enough in statistics. Throughout this paper we divide the degrees into groups and suppose that the interval of every two groups is 10 per unit degree. After that, we will calculate the number, and the probability of every group, and carry out related analysis on the distribution of the nodes, In the following part we will explain it on details.

3.4 Average Directed Path Length

Suppose that there are n nodes in the undirected complex network, and the maximum number of undirected edges is $n(n-1)/2$. Differently, in the directed complex network, there are $n(n-1)$ edges at most. Let d_{ij} (the distance between node i and node j) define the number of directed edge on the shortest path which connects 2 nodes. L represents the average directed path length. In other words, L is the mean value of the distance between two randomly chosen nodes, and therefore we can define it as:

$$L = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)}$$

Additionally, L describes the separation degree in complex network. The larger L becomes, the smaller separation degree is. Otherwise, if L is small, it means the separation degree in the network is large.

3.5 Directed Accumulation Coefficient

Assume that there are k_i nodes around node i , and therefore there are $k_i(k_i-1)$ directed edges at most of these k_i nodes. Let B_i defines the number of the directed edges among those k_i nodes, C_i denotes the accumulation coefficient of node i , and C represents the accumulation coefficient of the whole network; that is;

$$C = E\left(\frac{B_i}{k_i(k_i-1)}\right)$$

Where function $E(x)$ is to compute the mean of variable x . C describes the aggregation of nodes in the complex network, namely, the tightness of network. The rising C means the larger aggregation degree of the network. On the contrary, the aggregation degree is smaller.

3.6 Small-World Effect

When a large-scale complex network has small average path length and large aggregation coefficient at the same time, we regard such kind of network possesses small-world effect.

3.7 Scale-Free Property

A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. Generally, a testing way whether the network has scale-free property is to compute the degree of the node and take its logarithmic value of the degree as abscissa against the logarithmic value of the probability of the degree as ordinate. We then do regression analysis, and if the data can be fitted to a linear, the network can be regarded as a scale-free network.

4. Empirical Study and Results

Here we consider a practical application—a sample network of Shanghai Stock market. In this section, we selected closing price data of A shares in SSE (Shanghai Stock Exchange), from 1st January, 2012 to 31st December, 2014, to carry out our empirical study. After removing some delisted stocks from original 920 stocks, there are 450 stocks left and 726 data points in each individual stock (the financial data from CSMAR). With the help of MATLAB 2012, we set up a complex network on computer and calculate some available indexes to study the market.

4.1 Data Complementation

After the screening of the remaining 450 stocks, there are still some lack of data in 1 to 3 days among some of these stocks. In order to supplement the data, we use a random disturbance term. Specifically, to supplement the missing data is to create a random number between $[0, 1]$, and add it to the closing price the day before. Here we use this algorithm and take Great Wall Automobile (601633) for example:

Table 1. The example of data complementation

time	12 th January, 2012	13th January, 2012	16th January, 2012	17th January, 2012	18th January, 2012
deletion	12.65	12.26	deletion	12.99	12.77
Supplement	12.65	12.26	12.67	12.99	12.77

4.2 The Construction of Directed Network

Intuitively, we draw the image of network(using MATLAB2012). Taking a network of which the threshold is $\theta=0.04$, the interval is $\tau=1$ and $\Delta t=1$ For example, stock network image is displayed in figure 1 and figure 2(the number of stocks in the network is 40 in figure 1 and 100 in figure 2 respectively).

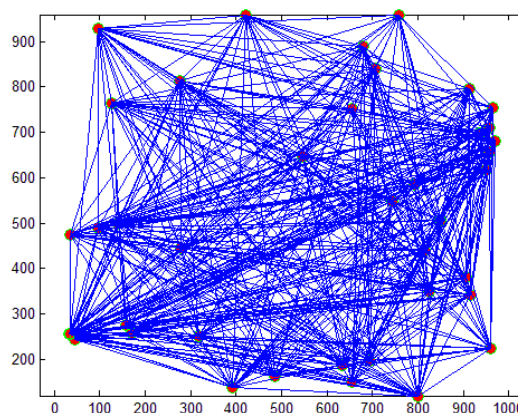


Figure 1. Network topology graph of 40 stocks

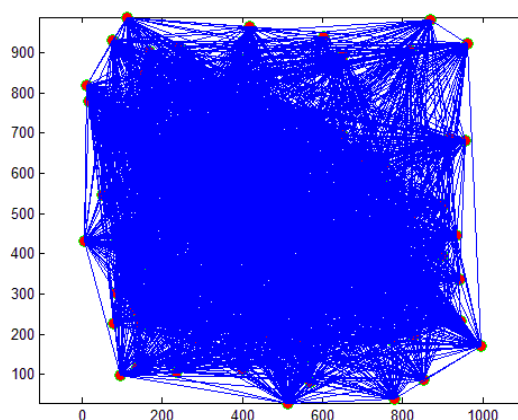


Figure 2. Network topology graph of 100 stocks

Based on the results and graphs above, we come into a verified conclusion that the directed complex network exists in the stock market of China and the structure of complex network is successfully constructed about Shanghai Stock Exchange.

4.3 Structural Characteristics of Directed Complex Network

4.3.1 Distribution of Stock Correlation Coefficient under Different Threshold

In this section, we set different threshold value intervals to do statistics of correlation coefficient matrix. Take interval $\tau=1$ and $\Delta t=1$ for example, Table 2 shows the distribution of the correlation coefficient of 450 stocks under different threshold value intervals.

Table 2. Threshold distribution of correlation coefficient

Low limit of threshold	Upper limit of threshold	The number of correlation coefficient	The percentage of correlation coefficient
-1	-0.4	0	0.00%
-0.4	-0.2	17	0.01%
-0.2	-0.1	975	0.48%
-0.1	-0.05	13399	6.62%
-0.05	0	8027	3.96%
0	0.05	151399	74.76%
0.05	0.1	25825	12.75%
0.1	0.2	2754	1.36%
0.2	0.4	87	0.04%
0.4	1	17	0.01%

As we can see in table 2, from 2012 to 2014, the correlation coefficients mainly distribute in $(-0.1, 0.1)$ and there are 87.52% in $(0, 0.1)$. It indicates that the correlation between stocks is not prominent enough, and most stocks shows positive correlation, namely, the fluctuation of price of stock a will make the price of stock b fluctuate in the same direction.

4.3.2 Out-Degree and In-Degree Distribution of Stock Nodes

Using MATLAB to calculate the out-degree and in-degree of stock nodes in the situation of threshold $\theta=0.04$, interval $\tau=1$ and $\Delta t=1$, we find that the minimum out-degree is 18 stem from MAANSHAN IRON, on the contrary, the maximum is China Minsheng Bank's 371; as for in-degree, Fangda Special Steel Technology enjoys the maximum of 377 and Huaxia Bank is 8 only. We divide the degrees into groups with the interval of every two groups is 10 per unit degree, and then draw probability distribution graphs of out-degree and in-degree respectively as figure 3 and figure 4 below.

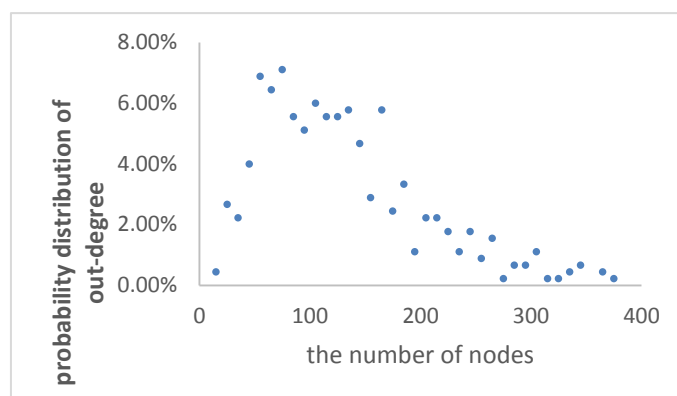


Figure 3. Probability distribution of out-degree

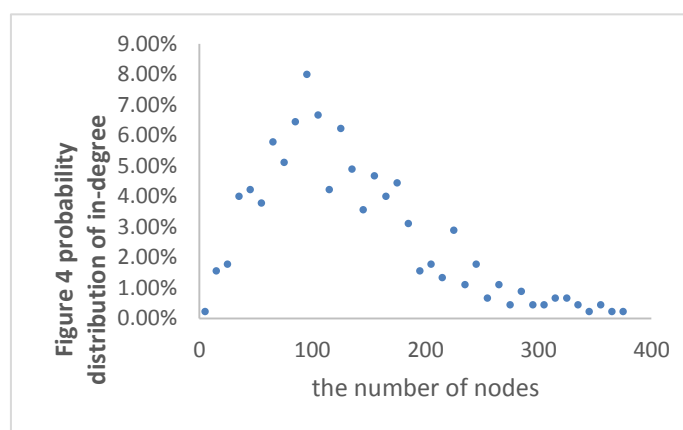


Figure 4. Probability distribution of in-degree

As we can see in the two figures above, under the situation of threshold $\theta=0.04$, time interval $\tau=1$ and $\Delta t=1$, with the increase of the number of stocks, the degree firstly increases and then decreases. We can also find that at about 128, the average value of degree, the probability reaches a maximal level.

The following tables give out the top 10 stocks which enjoy largest degree(in-degree and out-degree are separated)and their industries involved as well.

Table 3. Top 10 of out-degree

Stock code	ranking	Stock name	industry	Out-degree
600202	1	Harbin Air Conditioning	electrical equipment	371
601717	2	Zhengzhou Coal Mining machinery	special machinery	367
600015	3	Huaxia Bank	bank	360
600276	4	Hengrui Medicine	chemical pharmaceutical	346
600104	5	SAIC Motor	automobile	343
600837	6	Haitong Securities	security	340
600831	7	Guangdong Radio and TV Network Shares	web service	335
600815	8	Building Industry Share	dedicated device	331
600551	9	Times Books	media	320
601098	10	Zhongnan Media	media	316

Table 4. Top 10 of in-degree

Stock code	ranking	Stock name	industry	In-degree
600507	1	Fangda Special Steel of Science and Technology	steel	377
600760	2	Zhonghang Heibao	transport facilities	363
601107	3	Sichuan Expressway Company	expressway	357
600808	4	Magang Stocks Trading	steel	351
600307	5	Hongxing JISCO	steel	346
600979	6	Sichuan guangan AAA public	electric power industry	338
600540	7	Xinjiang Sayram Modern Agriculture	crop farming	332
600290	8	Huayi Electric	electrical equipment	327
601118	9	Hainan Rubber Group	chemical industry	327
600810	10	Shenma Industrial	chemical raw material	325

When the threshold is constant, a stock with rising value of out-degree makes a positive influence to other nodes; on the other hand, a stock with rising value of in-degree means it will receive more infection from other nodes. Based on the statement mentioned before, we can draw a conclusion that the leading stocks are those whose price fluctuation can influence more stocks. By examining and analyzing, we see that SAIC Motor is the leader of automobile stocks; and Guangdong Radio and TV Network Shares, Building Industry Share, Zhongnan Media are the leading stocks of its industries as well.

4.3.3 Small-World Effect

At time interval $\tau=1$ and $\Delta t=1$, we calculate the value of average directed path length and accumulation coefficient at different threshold by using MATLAB. Table 5 below shows the results.

Table 5. The value of average directed path length and accumulation coefficient at different threshold

threshold	path length	accumulation coefficient
0.01	1.2006	0.8062
0.02	1.3886	0.6312
0.04	1.6849	0.3765
0.08	2.3402	0.1522
0.1	Inf	0.1193
0.12	Inf	0.08
0.15	Inf	0.0479
0.18	Inf	0.0235
0.2	Inf	0.0099

According to the statistic data from paper, the average directed path length is considered to be small when it is less than 1.5; when the accumulation coefficient exceed 0.8, we regard it large. If the two indexes meet the above condition, the network shall enjoy small-world effect. Besides, the number of edges in directed complex network is possibly twice as many as that of the undirected complex network. Therefore, with networks of the same size, the path length and accumulation coefficient value of undirected network are twice as many as directed network. In our paper, we specified the average directed path length as relatively small when its value is less than 0.75; the accumulation coefficient as relative large when its value exceeds 0.4, and in this case, the network has small-world effect. Base on the data in Table 5, the value of average directed path length is larger than 0.75. Additionally, the accumulation coefficient will fall below 0.4 when the threshold exceeds 0.04. Finally, we can infer from the comparison that such kind of network doesn't enjoy the small-world effect.

4.3.4 Scale-Free Property

Under different threshold, the network of scale-free characteristics appeared inconsistent state. When the threshold is too small, there are so many edges that otherness among nodes isn't obvious enough. In this case, the network doesn't have scale-free property. On the contrary, if the threshold is too large, the otherness isn't obvious enough as well. The following page enumerates three examples as $\theta=0.04$, $\theta=0.1$, $\theta=0.2$ to represent three different situations(threshold too small, moderate, and too large)at time interval $\tau=1$, $\Delta t=1$, in which we apply the regression analysis method in dual-logarithm coordinates system with MATLAB to get the results .

For $\theta=0.1$, the regression result of out-degree is in Figure 5, with the linear equation

$\ln(p(k)) = -1.1418 - 1.2022 \times \ln(k)$, which shows the power law with the power exponent is 1.2022 and the slope of the line is -1.2022. Particularly, in this equation $R^2=0.8096$, $F=157.3544$, it means that regression results are satisfactory.

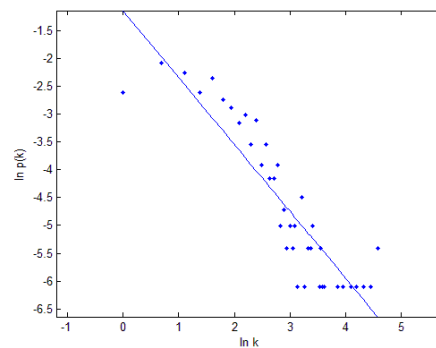


Figure 5. Distribution of out-degree in dual-logarithm coordinates system for threshold $\theta=0.1$

For $\theta=0.1$, the regression result of in-degree is in Figure 6, with the linear equation $\ln(p(k)) = -1.1771 - 1.2161 \times \ln(k)$, which shows the power law with the power exponent is 1.2161, and the slope of the line is -1.2161. Particularly, in this equation $R^2=0.8625$, $F=269.6159$, it means that regression results are satisfactory, too.

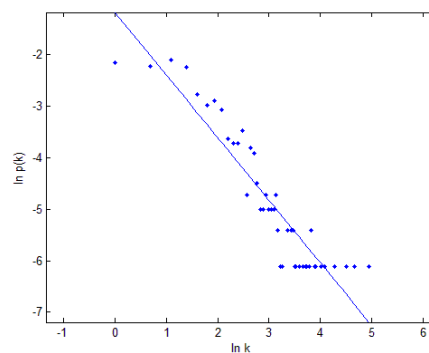


Figure 6. Distribution of in-degree in dual-logarithm coordinates system for threshold $\theta=0.1$

We draw degree distribution graph (take out-degree as an example because the difference between out-degree and in-degree is not so obvious) for $\theta=0.04$, $\theta=0.2$, respectively, as Figure 7 below. We can easily find that the function law of out-degree distribution isn't obvious.

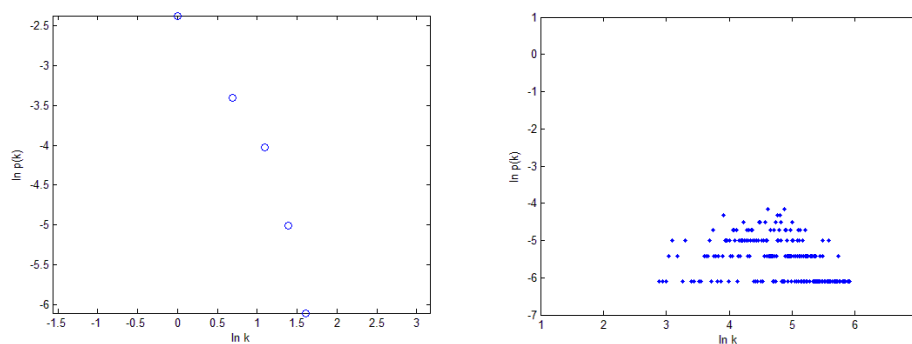


Figure 7. Distribution of out-degree in dual-logarithm coordinates system for threshold $\theta=0.2$, $\theta=0.04$

After multiple verify with program, our stock network enjoys scare-free property with threshold between $[0.08, 0.15]$.

4.3.5 The Relationship between the Amount of Nodes and Average Path Length

Under the situation with the threshold $\theta=0.04$, the interval $\tau=1$ and $\Delta t=1$, we make a statistical analysis on the number of nodes and average path length and the results are presented in Table 6. With the increase in the number of stocks, namely, the number of nodes, the average path length decreases gradually.

Table 6. Relationships between the amount of nodes and average path length

Stock amount	Average path length
20	1.773684
30	1.772414
40	1.728205
50	1.704082
75	1.694595
100	1.695455
150	1.684474
200	1.685452
250	1.687823
300	1.68379
400	1.681967
450	1.68488

Take the logarithm of number of stocks as independent variables and average path length as dependent variables to do a quadratic regression. We get a result dramatically, in which the linear equation is $\ln(L)=0.8039-0.1046\times\ln(n)+0.0096\times(\ln(n))^2$, with $R^2=0.9243$ and $F=54.9268$. The increase in the number of nodes leads to the increase of the separation degree, which means that the relationships between new issues and original stocks are not closely related.

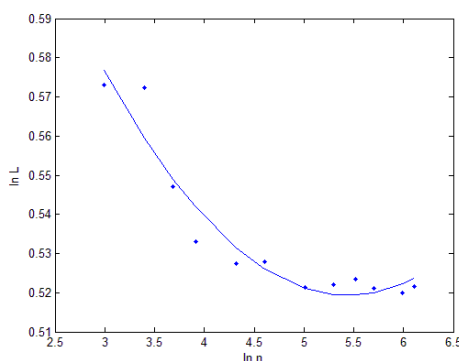


Figure 8. Regression chart of nodes amount and average path length

5. Conclusions

Our page set up a directed complex network with data of A stocks in SSE from 2012 to 2014 (data derived from CSMAR). After analyzing the intrinsic attributes and regulations in the financial market, some conclusions are enumerated as follows:

(1) During the period we study, the values of correlation coefficient mostly distribute in $(-0.1, 0.1)$. Furthermore, there are 87.52% of them distributing in $(0, 0.1)$. It indicates that the correlation between stocks is not prominent enough.

- (2) According to degree distribution, one can find out the leading stocks in financial market intuitively.
- (3) The model shows scare-free property in both out-degree and in-degree when the threshold is in $[0.08, 0.15]$. Additionally, the out-degree follows power-law distribution with the power exponent 1.2022, and the in-degree follows power-law distribution with the power exponent 1.2161, with threshold $\theta=0.04$.
- (4) The directed complex network model doesn't show small-world effect.
- (5) The number of nodes and average path length follow a logarithmic relation as follows: $\ln(L)=0.8039-0.1046\times\ln(n)+0.0096\times(\ln(n))^2$. As the increase in the number of nodes, the average path length decreases gradually.

References

- Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47. <http://dx.doi.org/10.1103/RevModPhys.74.47>
- Boginski, V., Butenko, S., & Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, 48(2), 431-443. <http://dx.doi.org/10.1016/j.csda.2004.02.004>
- Chao, L., Mingwen, W., & Yujie, M. (2014). A Study on Features of Interbank Market Based on Complex Network Theory—For Data Around Financial Crisis (2007-2009). *The Theory and Practice of Finance and Economics*, 2, 9-15.
- Ding, Z., Xintian, Z., Wenjuan, L., & Hongtao, Q. (2014). Research on Future Index Based on Complex Network Theory. *Journal of Systems and Management*, 1, 70-90.
- Hua, C. (2012). *Directed Cross-correlation Research Between Stocks Based on Complex Network*. Beijing University of Posts and Telecommunications.
- Kim, K., Kim, S., & Ha, D. (2007). Characteristics of networks in financial markets. *Computer Physics Communications*, 177, 184-185. <http://dx.doi.org/10.1016/j.cpc.2007.02.037>
- Mantegna, R. N. (1997). Hierarchical structure in financial markets. *The European Physical Journal B*, 11, 193-197. <http://dx.doi.org/10.1007/s100510050929>
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167-256. <http://dx.doi.org/10.1137/S003614450342480>
- Onnela, J. P., Chakraborti, A., Kaski, K., & Kertesz, J. (2003). Dynamic asset trees and portfolio analysis. *The European Physical Journal B*, 30, 285-288. <http://dx.doi.org/10.1140/epjb/e2002-00380-9>
- Weidong, F., Kun, L., & Jiangong, Z. (2008). Volatility analysis of Hang Seng Index. *Journal of South China University of Technology (Nature Science Edition)*, 12, 138-141.
- Xintian, Z., Zhifeng, M., & Shiyang, C. (2007). Characteristic analysis of complex network for shanghai stock market. *Journal of Northeastern University*, 28(7), 1053-1056.
- Yachun, G. (2013). *Complex System Approach to Analysis and Modelling of Financial Market Dynamics*. University of Science and Technology of China.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).