Proceedings of the 2017 IEEE/SICE International Symposium
on System Integration, Taipei, Taiwan, December 11-14, 2017

Excessive Co-Movement Effect and Evolution Network Analysis of Chinese Stock Market*

Shang Zhen[1] and Qinxin Wu[2] and Xiong Wang[3]

***Abstract*— In order to better understand mutual influences among stock price fluctuations, we treat stocks on the market as nodes, collecting closing price data of The Shanghai and Shenzhen 300 Index from 2014 to 2016. And we use the partial correlation coefficient to measure the linkage effect between stocks. Specifically, through the selection of a certain threshold, we get a financial complex network with stock linkage effect. Furthermore, we use Newman Fast analysis algorithm to divide Shanghai and Shenzhen 300 Index network into 15 communities. It turns out that the correlation within the community is significantly closer than the correlation between the community and the outside community, which is called close price-fluctuating plate. Moreover, through the time division that is used to analyze the structure of stock network community, we may conclude that banks, securities and real estate stocks' prices are basically a synchronous change during this period. And there are also some stocks which have excessive co-movement effect changes asynchronously. The results of our work can not only provide inner information and linkage on the stock market capital flows, but also can explore the evolution of the stock market rules.**

I. INTRODUCTION

Complex networks are part of the theory of complexity, which is able to correctly study the system in other fields. The nodes in the network are abstracted from the physically interacting individuals, and the edges between the nodes represent the interrelationships among the individuals. Because the interaction between individuals is not random, a variety of interesting features (clustering features, degree correlations, synchronization features, etc.) have emerged in complex networks, attracting more and more researchers' attention. In general, financial markets are considered to be a complex system model[1]. Financial market has accumulated massive high-frequency data for complex network theory research and empirical analysis to provide data support[2], [3]. With the increasingly stronger computer skills, scientists can search for more convenient and effective observation laws and modes with a large number of data and experience. In many truly complex networks, the community structure has a vital character. Thus, searching and analyzing the structure of the community is of great significance to the study of network structure and characteristics. Over the past few decades, such problems have received considerable amounts of attention, such as the detection and characterization of community structures in networks, clustering algorithms, and so on.

Currently, there are diverse analysis of the stock sector community based on complex networks. In order to understand the mutual influence of stock price fluctuation, Wang et al.[4] use the improved Newman greedy algorithm to divide the Shanghai stock market into 13 communities, and get a lot of market information from the community structure. Based on the theory and research method of complex network, the paper by Kang, Zhang, Liu[5] analyzes CSI300 stocks using the correlation coefficient of logarithmic yield rate. Based on the theory of complex network and community, by constructing the quantitative model, the stock is divided according to the correlation size and the distinction between the plate division. By this method[6], the mainstream division is found out. Zhang,Yang, Lu [7] use DFA for data screening to calculate the absolute correlation coefficient of the stock index of the filtered data. They build the network topology; calculate the stock index of the network statistical characteristics of indicators, analysis of yield, volume, price-earnings ratio of the network structure. In order to analyze the evolution of the securities market, Han, Liu, Wang[8] put forward the dynamic network model construction and topology analysis method and establishes the dynamic network model of the securities market.They also explores the basic characteristics and the community structure of the network. That means, it is possible to analyze the evolution of dynamic network by constructing complex network theory.

In our work, we propose a method to establish a complex network stock market with correlation coefficient matrix and use the community structure algorithm to analyze the linkage effect of the financial market. The proved classic GN algorithm [9] under the Newman fast algorithm [10] is a beneficial tool to look up for the stock network in the largest modular value Q. Thus, we could get the optimal community network structure. We then use the NetDraw tool to draw detailed community relationships, use Python to get structural information about the degree, weight, and neighbor nodes of each network community, and then analyze the importance of the stock nodes in the network[11]. After that, we further divided the time into several parts to analyze the structure of the stock network community. Finally, we explain and analyze the causes and characteristics of the domestic plate and the combined effect. In addition, we discuss investment strategies and feasibility based on our empirical results, which makes the combination of theory and

*This work was supported by the National Natural Science Foundation of China (No. 61601306) and Shenzhen Overseas High Level Talent Peacock Project Fund (No. 20150215145C)

[1]Institute for Advanced Study, Mathematics and Applied Mathematics, Shenzhen University, Shenzhen 518060, China seanzhen52@gmail.com

[2]Institute for Advanced Study, Mathematics and Applied Mathematics, Shenzhen University, Shenzhen 518060, China 562078230@qq.com

[3]Institute for Advanced Study Shenzhen University, Shenzhen 518060, China wangxiong8686@szu.edu.cn

978-1-5386-2263-6/17/$31.00 ©2017 IEEE

practice meaningful. These studies help us to understand the evolution of financial markets, especially the stock market, from the perspective of complex networks

II. DATA PROCESSING

Through the Shanghai and Shenzhen 300 Index, we selected 300 stocks with high value and good fluidity as a sample unit, which covers about 60% of the stock market value. It can strongly reflect the overall situation and the fluctuation characteristics of the Chinese stock market, especially Shanghai and Shenzhen A-share market trend. Time period is from Nov. 11,2014 to Jan. 11,2016, a total of 286 trading days. The stock data for this period covered the bull market stage and the bearish market stage, as well as the smooth phase, which helped us to analyze the potential relevance of the stock in different stages.

The selected data are first preprocessed. The weighted network model of non-scale is established, and the effective trading day price of stock $i(i=1,2,...,N)$ is $P_i^T(1), P_i^T(2), ..., P_i^T(N)$, where N represents the number of days of effective trading of stock i in time period T; $P_i^T(t)$ represents the closing price of stock i on the t-day effective trading day in period T. We then define the log yield of stock i as:

$$r_i^T(t) = lnP_i^T(t) - lnP_i^T(t-1) \tag{1}$$

At the beginning, we consider the selected time period as the total time period for analysis, i.e T = 1. So that each stock will form a 285 closing price logarithmic yield sequence. In addition, during this period of time, due to the reasons for their suspension, some valid trading days have no transaction data for some stocks i. In order to ensure that the valid trading days of all stocks are of the same length, the data processing in this paper is based on averaging the data of the previous valid trading day and the subsequent effective trading day.

Pearson linear correlation formula is usually used in analyzing the correlation between objects. We can compute the Pearson linear correlation coefficient between any two stocks i and j in time period T:

$$\rho_{XY}^T = \frac{Cov(r_i^T, r_j^T)}{\sqrt{D(r_i^T)D(r_j^T)}} \tag{2}$$

Among them, X and Y are random array of variables, which are accurately expressed as stock prices of a sequence of data in the stock market. The linear correlation coefficient is used to characterize the linear correlation between the random variables Y and X. When the value of $|\rho_{XY}|$ is close to 1, the linear degree of correlation between Y and X is close to the highest level; the closer the value of $|\rho_{XY}|$ is to 0, the weaker the linear correlation between Y and X is. When using linear correlation coefficients to measure linkage, it is easy to put it into practice. But this method is not so accurate, and the linkage effect may not be so simple to be linearly related only. It is because the correlation between the stocks is not only directly decided by the internal relations of the two stocks, but also by the various factors in the market. At this point to accurately reflect the linkage between the two stocks, we can not simply calculate the linear correlation coefficient, but need to consider the partial correlation coefficient, the formula is as follows:

$$\rho_{ij(k)} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{1-r_{ik}^2}\sqrt{1-r_{ijk}^2}} \tag{3}$$

Where r_{ij} is the Pearson linear correlation coefficient between stock i and stock j, and k represents the data of the CSI 300 index at the selected time, that is, r_{ik} is the linear correlation coefficient between stock i and the CSI 300 index. It's more accurate to use the CSI 300 index as another factor to influence the correlation between two stocks.

The value of the partial correlation coefficient is the same as that of the simple correlation coefficient, and the linkage between the logarithmic yields is calculated. We may conclude that the correlation coefficient is larger, indicating that the linkage between the variables is closer, and vice versa.

Among them, the edges between stocks under various partial correlative values is shown as Figure. 1, which x-axis refer to partial correlative coefficient and y-axis refer to the edges.

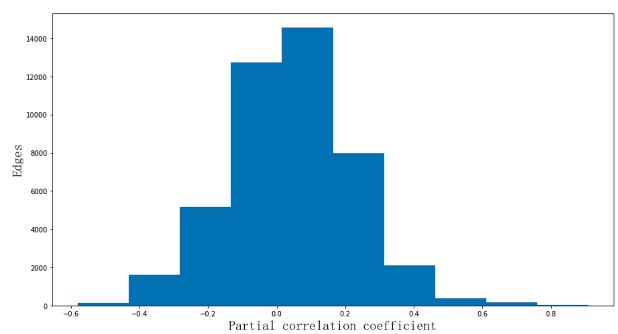

Fig. 1. The histogram of partial correlative coefficient

For example, we compare the relative increase from the Figure. 2. We can see that the increases of stocks with relatively high linkage correlation coefficient become more consistent.

Based on the analysis of the number of edges between the vertices in the stock network, the average clustering coefficients of the networks respectively reflect the clustering of the network, and the graph is the change curve of the average aggregation coefficient with the partial correlation coefficient threshold setting, which is shown as Figure. 3.

And we get that the larger the threshold, the less the edge of the network. The average aggregation coefficient is 0.517 when the partial correlation coefficient is 0.4, we choose 0.40 as the threshold, and the aggregation result is satifying. That is, only when the side of the coefficient is greater than 0.40, the two stocks are considered related, which means they share a strong linkage. Analyze the left 239 communities, nearly 1000 edges; we receive a better clustering effect.

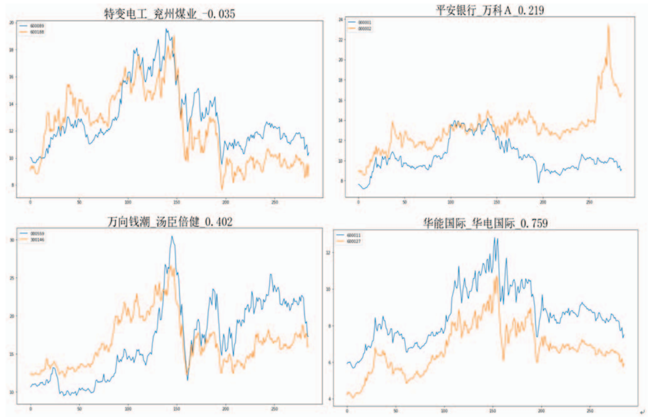

Fig. 2. The linkage between the stocks under various values(closing price/day)

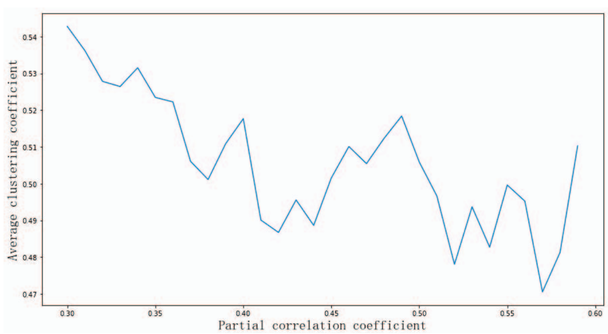

Fig. 3. The average aggregation variation curve

Netdraw tools can be used to get the network effect as shown in Figure. 4:

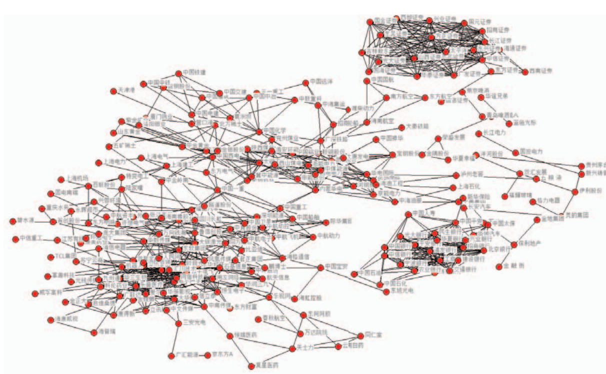

Fig. 4. Correlation of stocks network

III. DESCRIPTION OF THE SPLITTING ALGORITHM FOR COMMUNITY STRUCTURE SEARCH

M.E.J. Newman and M. Girvan[9] proposed an algorithmic network for extracting community structures from complexities as a method of division. The algorithm is based on the characteristics that high cohesion exists within the community and low cohesion between community, then gradually removes the community between the sides, to obtain a relatively cohesive community structure. GN algorithm is based on this idea to repeatedly calculate the shortest path of the current network, calculate the betweenness of each edge, deleting the edge of the largest betweenness. Finally, under certain conditions, the algorithm stops, we can get the network community structure.

We implement this algorithm to program and try to find out some relevant information, like instructions blocks and community structures. In general, the GN algorithm has the following basic process:

(1) Calculate the betweenness between each edge in the network
(2) Find the edge of the highest betweenness and remove it from the network
(3) Repeat step 2 until the system is broken down into N unconnected vertices.

Where the betweenness of a certain edge is defined as the number of shortest paths passing through every edge in the network. Obviously, if there are many shortest paths going through one edge, this edge is crucial to the network and it is quite possible that if we remove this edge, the whole network will split into two parts (communities) with strong structural characteristic.

For defining a good community and the steps we end the split, we use the Newman algorithm in the modular Q[10] to quantify how our community split. Best number of classes (Q function):

$$Q = \frac{1}{2m}\sum_{ij}[A_{ij} - \frac{k_i k_j}{2m}]\delta(c_i, c_j) \qquad (4)$$

Where A_{ij} is the element in the network adjacency matrix, k_i is the degree of i point, and

$$\delta(c_i, c_j) = \begin{cases} 1, i \in c \ \ j \in c \\ 0, others \end{cases} \qquad (5)$$

We use the above algorithm to process the stock data in order to try to find block relationships. We draw the line graph of the value of the Q function in the process. In Figure. 5, the Q diagram, we found that when the number of communities is 15, and the maximum Q value appears at 0.517.

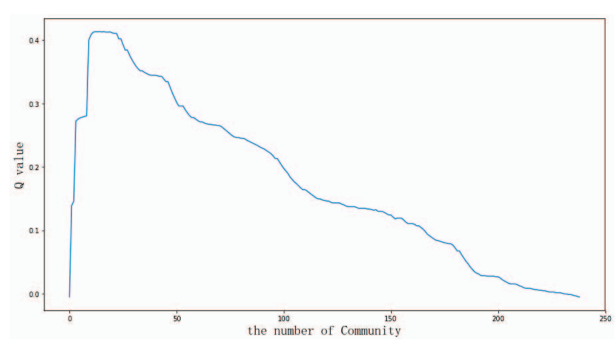

Fig. 5. The relationship curve between Modularity Q and the number of communities

The Q value obtained is larger than that of the classical community network in some papers (Zacharys karate club network[12] gets a Q value at about 0.3). This shows that the reality network is similar to the Zachary's network that has a relatively independent community. The relationship between the communities is close and there is a certain network logic.

The NetDraw tool relocates the nodes according to 15 communities to get the following Figure. 6:

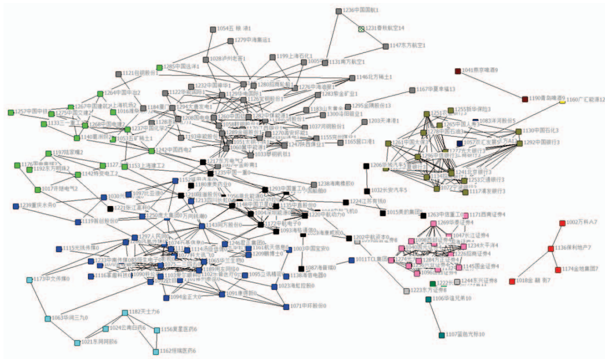

Fig. 6. The clustering structural correlation of stocks network

IV. COMMUNITY ANALYSIS

We got a block diagram of the 15 community structures in the clustered form. In the stock market, the devision of nominal industry is often based on industry, location and concept. For example, it includes energy blocks, steel blocks, etc. Our goal is to find accurate blocks with strong correlations for all shares.

In 15 societies, according to the size and characteristics of the community, First, class by size:

(1) the number of members of the community greater than or equal to 21 is divided into large associations
(2) the number of members of the community less than or equal to 20 is divided into small associations

Secondly, by character:

(1) Members belong to a same community of nominal plate; the members of the community are more relevant, sharing more accurate range.
(2) some members of the community belong to different plates, establishing a strong correlation between the new real plate. The stocks make up the new community will provide valuable information about the potential relationships among the Capital Flow, Investment Institute's portfolio and the underlying stocks.

The large number of stocks in large community is mainly due to the fact that the price fluctuation in this period is not significant; the discrimination is not distinct, leading to a higher correlation without analysis value. According to the size and character of the classification, we will focus on the analysis and display of some of the new real plate from the small stock community.

The following are stocks that come from different nominal industries and form a new potential comprehensive concept block.An example of a new community plate that combines banking and insurance(Figure. 7).

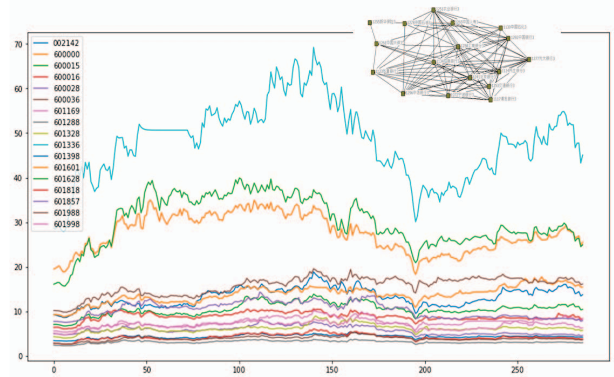

Fig. 7. The Community Structure and Trend Line of new community belonging to banking and insurance(closing price/day)

New portfolio of securities industry plate network and stock price fluctuations during this period.(Figure. 8)

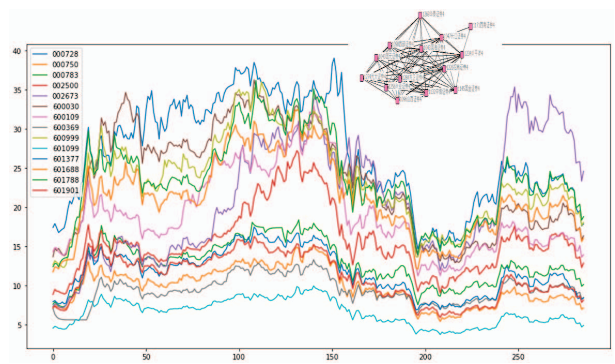

Fig. 8. The Community Structure and Trend Line of new community belonging to securities(closing price/day)

The new combination of the real estate industry sector network and the stock price fluctuations during this period.(Figure. 9)

Fig. 9. The Community Structure and Trend Line of new community belonging to the real estate industry(closing price/day)

Through the above example, we can easily find that stocks in the same new community are changing at the same time, even if they come from a variety of nominal industries.

V. RESEARCH ON EVOLUTION LAW OF STOCK NETWORK

In order to study the evolution of the stock market during this period, the 286 trading days are divided into six time periods with 60 days as an interval based on the observation of stock prices during this period, which can divided into the time periods including a stable phase of the stock price, the bull market rose stage and the bear market crash phase, using the same direction as former to deal with time period data: calculate the partial correlation coefficient; set the coefficient threshold in order to build a strong network, gradually remove the edges in the network. These make the network aggregation and other structural features more obvious. Then we can analyze the short time network and the structure of the community in the network respectively.

The average clustering coefficient of the network and the number of nodes of the maximum connected subgraph of the network respectively reflect the aggregation and connectivity of the network. Figure. 10 shows the curve of the average aggregation coefficient changing with the threshold setting each time period.

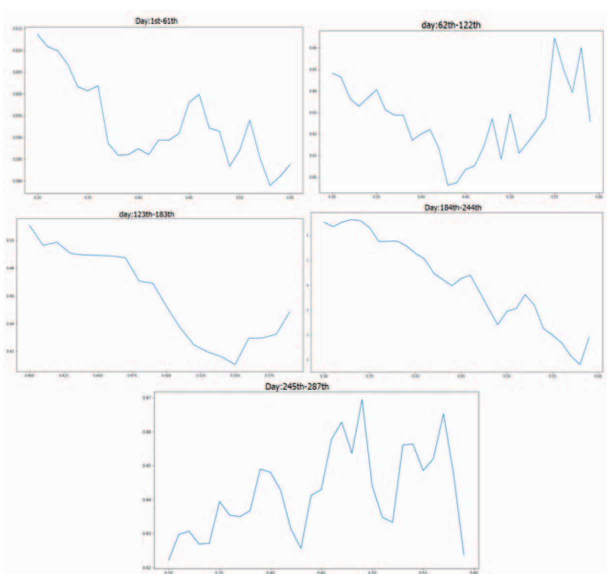

Fig. 10. The average aggregation variation curve for different time periods

From the Figure. 10, we can see that the average aggregation coefficient will have a maximum value within a certain range. With reference to the number of nodes and the average number of aggregations, the corresponding optimal threshold for each time period can be obtained, shown as follows:

Time (days)	14/11/11 - 15/2/5	15/2/6 - 15/5/12	15/5/13 - 15/8/6	15/8/7 - 15/11/10	15/11/11 - 16/1/11
Threshold	0.46	0.5	0.46	0.52	0.49

Then, according to stock network community structure division for each time using Newman greedy algorithm, we can get corresponding optimal Q value, shown as follows:

Time (days)	14/11/11 - 15/2/5	15/2/6 - 15/5/12	15/5/13 - 15/8/6	15/8/7 - 15/11/10	15/11/11 - 16/1/11
Optimal Q	0.538	0.409	0.29	0.495	0.332

The larger the Q value is, the better the division of the stock section community will be. The initial Q value is up to 0.5 or more, in the first 123 days to 183 days, the best Q value is smaller, close to 0.3, while other best Q values are all greater than 0.3. It indicates that there is a more significant division of the community structure. During this period of time, the trend of the development of socialization is obvious.

The Newman greedy algorithm divides the stock market community for each time period shown as follows:

Time (days)	Number of Community	Number of Major Community	Major Community and Involving Industry
14/11/11 - 15/2/5	8	4	Securities, banking, real estate, aviation
15/2/6 - 15/5/12	17	8	Energy, securities, banking, aviation, real estate and so on
15/5/13 - 15/8/6	8	3	Securities, banking, aviation
15/8/7 - 15/11/10	15	9	Energy, biology, real estate, aviation, construction, securities, banking and so on
15/11/11 - 16/1/11	19	11	Energy, banking, electricity, securities, pharmaceuticals, real estate and so on

There is a small number of societies divided between days 1 and 61 and days 123 to 183, which indicates that most of the stock in the two time share similar price. And these two periods are from the period of rising and falling, indicating that the most of the stock price fluctuations are similar in significant ups and downs period.

We can find that, through analysis of mainly related fields of divided community in each time, some banks, securities and real estate industry stocks can be defined as a representative community in every time period. It indicates that the share relationship among these members of the community is very strong. If we are familiar with the change trend of a stock in a community, then we could invest on other stocks in this community under the restrciton of community division, showing the value of portfolio transactions.

VI. EXCESSIVE COMOVEMENT ANALYSIS

From the above analysis, some industry stocks with time changes can always be built into a community. That is to say, stocks in this community are basically changing synchronically or reducing during the same period,such as securities industry stock-based community, which is shown as Figure. 8.

But some of the securities divided by traditional industries are not classified as securities societies, and these separate industry stocks are generally divided into a large community. The shares of this new community belong to the nominal industry including different sources. why the linkage between them is higher than the linkage with the stock of the same name industry? After the evolution analysis of the community over time, we found that:

(1) In the bull market or bear market stage, many stocks are in the state of rising or falling, as we analyzed from 1 days to 61 days and 123 days to 183 days. Because they are in the state of rising or falling, the variances of stock price fluctuations are not significant. The stocks with small degree of variation in fluctuations will easily be divided into the same community. Thus, we may say that in this state, large community has stocks from various different nominal industry, and they share a high linkage. Figure. 11 shows the Securities community stock price volatility in the day of 123th-183th.

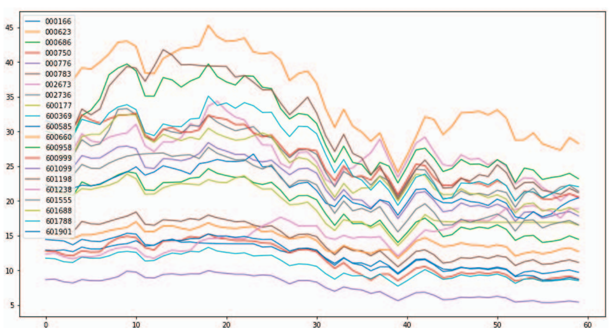

Fig. 11. 123th-183th Securities community stock price volatility(closing price/day)

(2) Compared with other stocks in the same industry, some stocks have delay effects. In analyzing process of some communities evolution during a period of time, stocks in securities industry share this delay effect, for instance. Other industry stocks also have such problems, the general delay of the cycle is short, lasting for 3-10 days or so. Figure. 12 shows the price volatility of two stocks which each code of stock is 002500 and 601198, belonging to securities industry in different communites.

VII. CONCLUSION

The study of complexity is another breakthrough in the field of economics. It transforms the inherent pattern of thinking in the field of economics, prompting people to better understand the evolutionary mechanisms of society.

In this paper, we firstly propose and establish the partial correlation matrix of CSI300 constituent stocks. Then, through the community structure Newman greedy algorithm and the network modular Q function calculation process, we get the stock network structure. The clustering and connectivity of the nodes show the hierarchical structure of the

Fig. 12. Stock price volatility(closing price/day)

network. By calculating the network clustering coefficient, the financial network has been proved to have a hierarchical structure. The stock is closely linked to each other to develop a small group, and these small groups develop into a larger whole. We found that the stock has a strong correlation and thus more accurately divided plates can be found in Chinese stock market, which is shown the value of portfolio transactions.

And then we build a complex network model to divide stock market according to time interval to study the working rule. This article mainly researches on the community nodes of stock network, analyzing involved industry of communities and comovement effect. However, the selected time period is not long enough. The next step in research would be the control on the time period selection of the complex network structure.

REFERENCES

[1] Bonanno, G., Lillo, F., Mantegna, R.N.: Levels of complexity in financial markets,arXiv: cond-mat/0104369 (2001)
[2] Complex network theory and its application [M]. Tsinghua University Press, Wang Xiaofan, Li Xiang, Chen Guanrong , 2006
[3] Guo Shizhe, Lu Zheming (2012): Complex network basic theory. Beijing. Science Press. Pp.1-340
[4] Wang Juan, Wang Weihua. Based on the complex network of stock community analysis [J] Journal of Wuhan University of Technology (Information and Management Engineering). 2010 (05)
[5] Analysis of Complex Network Characteristics of Shanghai and Shenzhen 300 Stock Networks [J]. Contemporary Economy, 2015,33: 153-155.
[6] Yu Zhanqiu, Li Lulu, Xu Dongmei, Zhou Wei.Study on stock correlation analysis based on complex network [J]. China Market, 2015,22: 42-45.
[7] CZHANG Li-jun, YANG Zhi-hui, LU Fei-fei.Experimental Analysis of Stock Index Relevance Based on Complex Network Theory [J]. China Management Science, 2014, 12: 85-92.
[8] HAN Hua, LIU Wan-lu, WANG Jin-shui.Study on the Construction and Evolution of Dynamic Network Model in Securities Market [J]. Chinese Journal of Management, 2013, 10 (2): 299-304.
[9] Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks.Phys. Rev. E 69, 026113 (2004)
[10] Fast algorithm for detecting community structure in networks M. E. J. Newman. 2004
[11] Du C, Wang X, Qiu L. Block & comovement effect of stock market in financial complex network[J]. Complex Sciences, 2009: 1248-1260.
[12] Zachary W W. An information flow model for conflct and fission in small groups. Journal of Anthropological Research, 33, 452-473, 1977