

Research Article

Dynamic Communities in Stock Market

Xiangquan Gui,^{1,2} Li Li,² Jie Cao,² and Lian Li¹

¹ School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China

² College of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

Correspondence should be addressed to Xiangquan Gui; guixq@lzu.edu.cn

Received 19 March 2014; Accepted 16 April 2014; Published 8 May 2014

Academic Editor: Suohai Fan

Copyright © 2014 Xiangquan Gui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The stock market has the huge effect and influence on a country or region's economic and financial activities. But we have found that it is very hard for the prediction and control. This illustrates a critical need for new and fundamental understanding of the structure and dynamics of stock markets. Previous research and analysis on stock markets often focused on some assumptions of the game of competition and cooperation. Under the condition of these assumptions, the conclusions often reflect just part of the problem. The stock price is the core reflections of a stock market. So, in this paper, the authors introduce a methodology for constructing stock networks based on stock prices in a stock market and detecting dynamic communities in it. This strategy will help us from a new macroperspective to explore and mine the characteristics and laws hiding in the big data of stock markets. Through statistical analysis of many characteristics of dynamic communities, some interesting phenomena are found in this paper. These results are new findings in finance data analysis field and will potentially contribute to the analysis and decision-making of a financial market. The method presented in this paper can also be used to analyze other similar financial systems.

1. Introduction

The stock market is a dynamic complex system formed from many enterprises, institutions, and individuals, which are connected with each other by trade, investment, and so forth. The stock market has a great effect and influence on a country or region's economic and financial activities. But we have noticed that it is very difficult to predict and control it. There is an urgent need to have a new and global understanding of the structures and dynamic characteristics of stock market.

Previous researches on stock markets mainly focused on some competition and cooperation games under specified conditions. Because of the high complexity of the stock market, the conclusions under the limited conditions often only reflect a part of the problems. This forces us to adopt a new study mode, from a more macroperspective, to explore the characteristics and reasons behind stock markets complex changes.

From chaos to complexity, from the molecular activities of cells in our body to the communications between people

in our entire planet, complex network theory provides a new method to explore the world for us. Particularly, from 15 years ago, Barabási published his pioneering paper of scale-free network [1]; complex networks have attracted the interest of many researchers in different fields of the world. And a large number of research results have been produced in recent years. These research achievements provide a powerful tool and reference for our understanding of the real world complex systems, such as protein interaction networks in the field of biology, social networks, and scientists collaboration networks in the field of sociology [2, 3]. The theory and tools of complex networks also provide us with a new perspective to study stock markets. The price of stocks is the final and most core reflection of a stock market; therefore, in this paper, we construct stock networks based on stock prices and study the evolution characteristics of the community structure in it, by using complex network theory and tools. The evolution characteristics of community structure in time series not only reflect the changes of a stock group, but also reflect the stock market's global features. Through such a new research

approach, from a more macropoint of view, we can mine and reveal the underlying characteristics and laws hiding in the big data of stock markets.

In this paper, we constructed stock networks based on prices of stocks in a stock market. The stock network refers to the graph consisting of nodes (vertices) and edges, where nodes correspond to stocks (companies) and edges between them to price fluctuation relationships, which are constructed by computing a correlation coefficient of each pair of stocks. Mantegna is the first person to construct stock networks based on stock price correlations [4]. After that, many papers based on stock price correlations were presented. For example, Onnela et al. studied split-adjusted daily closure prices for a total of $N = 477$ stocks traded at the New York Stock Exchange over the period of 20 years, from 2 January 1980 to 31 December 1999 [5, 6]. They constructed dynamic asset graphs and dynamic asset trees based on price correlations and discussed their properties and differences. Kullmann et al. studied the clustering of companies within a specific stock market index, like the Dow Jones (DJ) or the Standard & Poor's 500 (S&P 500), by using the Potts superparamagnetic method [7]. They constructed an appropriate q-state Potts model, where the spins correspond to companies and the interactions are functions of stock price correlations. Boginski et al. studied characteristics of the stock network representing the structure of the US stock market and detected cliques and independent sets in it [8]. Jallo et al. constructed three kinds of stock networks based on American and Swedish stock markets and compared the characteristics of three construction methods [9]. Vizgunov et al. constructed the stock network for different time periods from 2007 to 2011, based on the Russian stock market. They found that for the Russian market there is a strong connection between the volume of stocks and the structure of maximum cliques for all periods of the observations [10, 11]. Huang et al. constructed a correlation network of the China stock market using the threshold method and then studied the structural properties and the topological stability of the network [12].

More specifically, in this paper, we study the split-adjusted daily closure prices for $N = 400$ stocks which were traded at the Hong Kong Exchanges (HKEx) over the period of 10 years and construct stock networks based on stock price correlations. Different from the literatures mentioned above, we focus on the properties of dynamic communities in the networks. Since one of the most relevant features of networks representing real systems is community structure, which is the organization of nodes with many edges joining nodes of the same communities and comparatively few edges joining nodes of different communities [13]. Moreover, a financial market is characterized as an evolving complex system [14]. So the evolution (or change) of communities is analyzed in this paper. Basic events that may occur in a community evolution are birth, growth, contraction, merger with other communities, split, and death, which were systematically proposed by Palla et al. in the literature [15]. Therefore, we believe that the analysis of dynamic communities in a stock market is more meaningful than static ones and that is a new macroperspective to understanding a stock market. Through the analysis, we find several phenomena as follows. First,

the evolution of communities in stock networks is different from other networks, such as social networks in the literature [15] and, second, correlativity exists between the characters of dynamic community structure and the fluctuation of the stock market. These results potentially contribute to market analysis and decision-making.

The paper is structured as follows. Section 2 describes how to construct stock networks. In Section 3, we describe how to detect and match communities in the networks. The analysis of dynamic communities is then offered in Section 4. Finally, in Section 5, we summarize our findings and present some thoughts on future researches.

2. Constructing Stock Networks

In this paper, the term stock networks refers to a set of undirected graphs, where the nodes correspond to stocks and the edges correspond to correlation coefficients between them. The data set is stocks' daily closure prices traded at the Hong Kong Exchanges (HKEx). We chose $N = 400$ stocks and collect the stock data over the period of 10 years, from 3 January 2000 to 6 August 2010. We construct networks by the split-adjusted daily closure prices of stocks, in a total of 2616 price quotes per stock. The data is divided into M windows of width T in order to uncover dynamic characteristics of the networks. The window width T corresponds to the number of daily returns included in the window. A number of consecutive windows overlap with each other. The starting time of a window is determined by the window step length parameter δT , which describes the displacement of the window, measured in trading days. The data windowing method and some associated parameters are illustrated in Figure 1.

Let $P_i(\tau)$ be the closure price of the stock i at time τ , where τ refers to a date. Given a time window t , $t = 1, 2, \dots, M$, let the return vector of stock i in the window t be r_i^t , whose components are logarithmic returns of the stock i in the window t ; that is, $r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - 1)$, where the value extension of τ is extended from the second trading day in the window t to the first trading day in the next window $t + 1$. In order to investigate correlations between stocks in the window t , the correlation coefficients between stocks i and j are defined as

$$\rho_{ij}^t = \frac{\langle r_i^t r_j^t \rangle - \langle r_i^t \rangle \langle r_j^t \rangle}{\sqrt{[\langle r_i^{t2} \rangle - \langle r_i^t \rangle^2] [\langle r_j^{t2} \rangle - \langle r_j^t \rangle^2]}}, \quad (1)$$

where $\langle \cdot \rangle$ indicates a time average over T days. The correlation coefficient ρ_{ij}^t fulfills the condition $-1 \leq \rho_{ij}^t \leq 1$, and the value of ρ_{ij}^t reflects the level of correlations between the stock i and stock j , from the perfect correlation ($\rho_{ij}^t = 1$) to the perfect anticorrelation ($\rho_{ij}^t = -1$). Those correlation coefficients form an $N \times N$ correlation matrix C^t , which is the basis of stock networks constructed in this paper.

To construct stock networks, we need to discuss two parameters, T and δT , first. Onnela et al. have used this

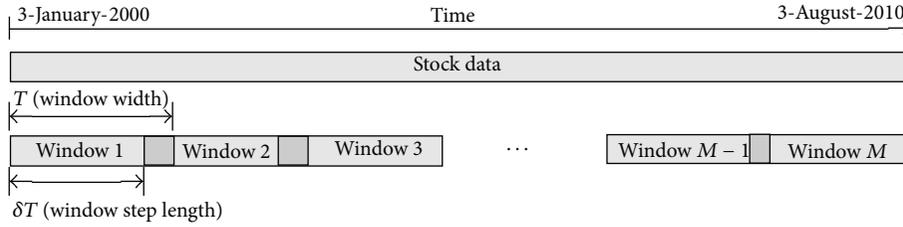


FIGURE 1: The data windowing method and parameters.

kind of correlation coefficients, and the method of time windows division to construct asset graphs can be found in the literature [6]. They also point out that the choice of T is a trade-off between too noisy and too smoothed data for small and large T , respectively. They find out that $\delta T = 21$ days and $T = 1000$ days are optimal values [16]. However, we find that $T = 1000$ is not suitable for our data, although $\delta T = 21$ is a good choice through many experiments. Some experimental results can be seen in Figures 2 and 3.

Let $T = \{1000, 400, 200, 100\}$ and let the window step length $\delta T = 21$ (fixed at about one month). Figure 2 shows four plots of the mean correlation coefficient as a function of time, defined as

$$\bar{\rho}(t) = \frac{1}{N(N-1)/2} \sum_{\rho_{ij}^t \in C^t} \rho_{ij}^t \quad (2)$$

To have a clearer picture of correlation coefficients, Figure 3 shows four contour plots of probability density functions for the correlation coefficients with different T values. From the visual point of view, it is difficult to say which is the optimal T value. It seems that set $T = 1000$ in Figures 2(a) and 3(a) makes the data too smooth, which may lose too much market information. On the contrary, setting $T = 100$ in Figures 2(d) and 3(d) makes the data seems too noisy. So we consider two periods of the HSI (Hang Seng Index, which is the most widely quoted indicator of the performance of the Hong Kong stock market) as two reference points to choose the T value. From 2000 to 2010, there are two large fluctuations of the HSI. One is during the period from March 2004 to February 2005; the HSI fell to 10918 points from 14058 points, down 22%, and returned to the former level. The other is from October 2007 to August 2010 (the end date of stock data used in this paper); the HSI fell to 13674 points from 31958, down 57%, and still has not returned to the former level now. This one can be considered as a bear market, since stocks trend downwards for a long period. Taking these two periods as reference points, we believe setting $T = 200$ is a good choice, since Figures 2(c) and 3(c) show quite clearly different regions during the two periods. From Figures 2 and 3, we can also find that stocks are more closely associated with each other when the stock market fell. This phenomenon is described by the commonly heard phrase of “decline is characterized by the stocks moving together.”

Setting $T = 200$ and $\delta T = 21$, the overall number of windows is $M = 116$; that is, $t = 1, 2, \dots, 116$. With

these choices, we can construct the stock networks S^t based on the correlation matrix C^t , by simply considering C^t as the adjacent matrix of S^t . Then S^t are weighted undirected complete graphs. However, it is hardly to analyze the community structure in these complete graphs. Since these graphs represent the market, it is natural to construct some graphs by including only the strongest connections in it. But how many edges (connections) should be included in such graphs? From Figure 4(a), we can find that the fewer edges included the fewer nodes incident with at least one edge. It means that if we include few edges, then many nodes will become isolated nodes. This will lose a lot of useful information. On the contrary, if we include too many edges, then graphs will not have distinct community structure, measured by modularity values, as Figure 4(b) shows, where the modularity is used as an indicator of community structure, which measures the density of links inside communities as compared to links between communities. It is defined as

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(D_i, D_j), \quad (3)$$

where A_{ij} represents the weight of the edge between nodes i and j , $k_i = \sum_j A_{ij}$ is the sum of weights of edges attached to node i , and D_i is the community to which node i is assigned. It is detected by the algorithm of Blondel et al. [17], which will be introduced in the next section. Consider $\delta(\mu, \nu) = 1$ if $\mu = \nu$ and 0 otherwise; consider $m = (1/2) \sum_{i,j} A_{ij}$ is the total weights of edges of the graph.

In practice, modularity values of many real networks typically fall in the range from about 0.3 to 0.7 [18]. When the modularity value of a network is below 0.3, it can be considered to have no distinct community structure. Therefore, in this paper, we include 1.2% of total edges to construct stock networks, according to the correlation values from large to small and deleting all isolated nodes. Then the average modularity value of the networks is 0.302 and the average coverage of nodes reaches 35.2%. Finally, stock networks S^t are weighted undirected graphs with the fixed edges number 957 and average 140.68 nodes. Figure 5 shows several pictures of stock networks S^t . In the figure, different node colors represent different communities. As we expected, nodes in the same community are basic stocks belonging to the same industry, which also fits the stock movements in the real market.

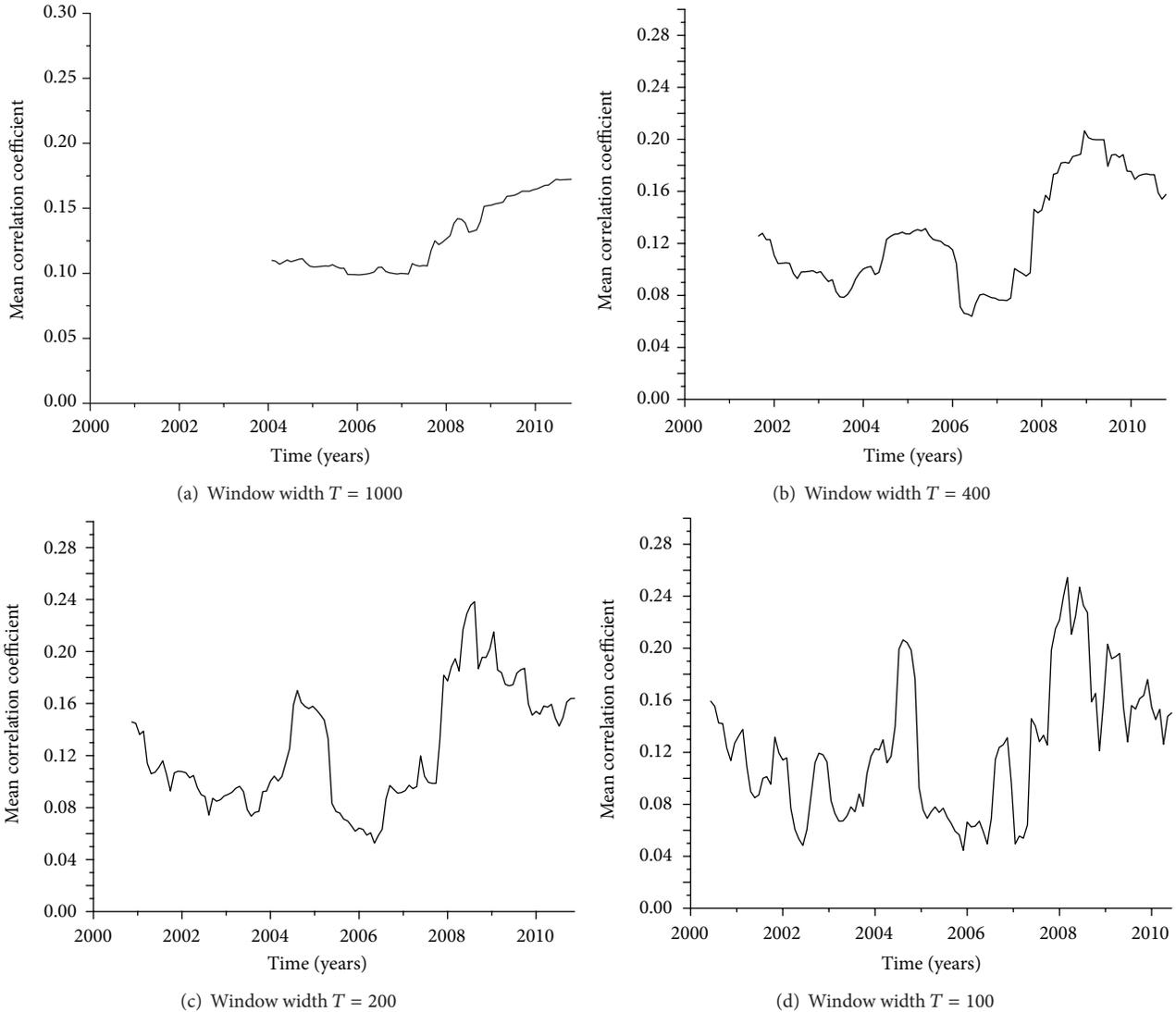


FIGURE 2: Some experiment results of $\bar{\rho}(t)$ as a function of time.

3. Detecting and Matching Communities in Time Windows

After constructing the stock networks, we detect communities in each time window using the algorithm of Blondel, which is introduced in the literature [17] and has been widely used in many cases of weighted graphs. There are several reasons why we choose this algorithm. First, stock networks are weighted graphs, and the quality of communities detected by Blondel's algorithm is very good, as measured by the weighted modularity. Second, this algorithm can unfold a complete hierarchical community structure for a network, which is very useful for further studies of hierarchical structures in stock networks. Third, the algorithm is extremely fast. It is shown that this algorithm outperforms all other known community detection methods in terms of computation time in the literature [17].

Blondel's algorithm uses a greedy method based on weighted modularity optimization. Initially, all nodes of a graph are put in different communities. Then, the algorithm is divided into two phases that are repeated iteratively. The first phase consists of a sequential sweep over all nodes until no further improvement of modularity achieved. At the end of the first phase, the first level partition is obtained. In the second phase, a new network is built whose nodes are the communities found during the first phase. The two phases of the algorithm are then iterated, yielding new hierarchical level partitions, until there are no more changes and a maximum of modularity is attained. The details of the algorithm can be seen in the literature [17].

After communities have been detected in each time window separately, to analyze characters of dynamic communities, communities in succeeding time windows have to be matched with each other. We use the match method posed

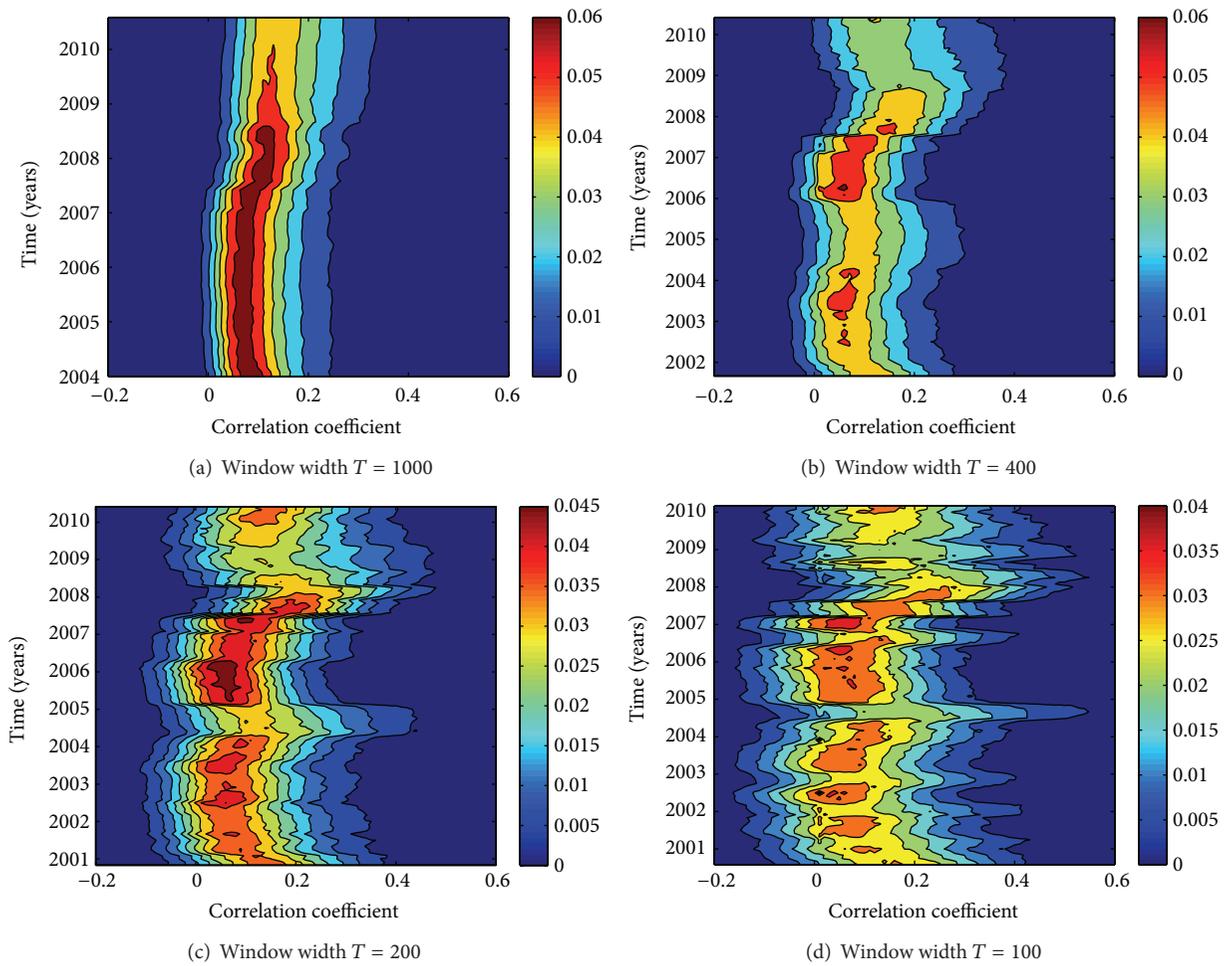


FIGURE 3: (Color online) Contour plots of probability density functions for correlation coefficients with different T values.

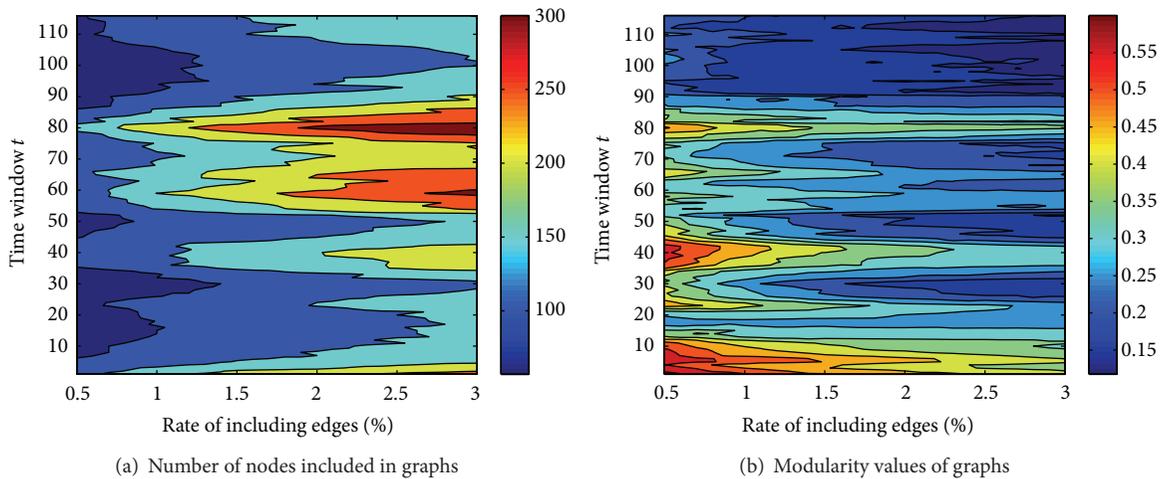


FIGURE 4: (Color online) Contour plots of node numbers and modularity values of graphs versus rate of including edges and t .

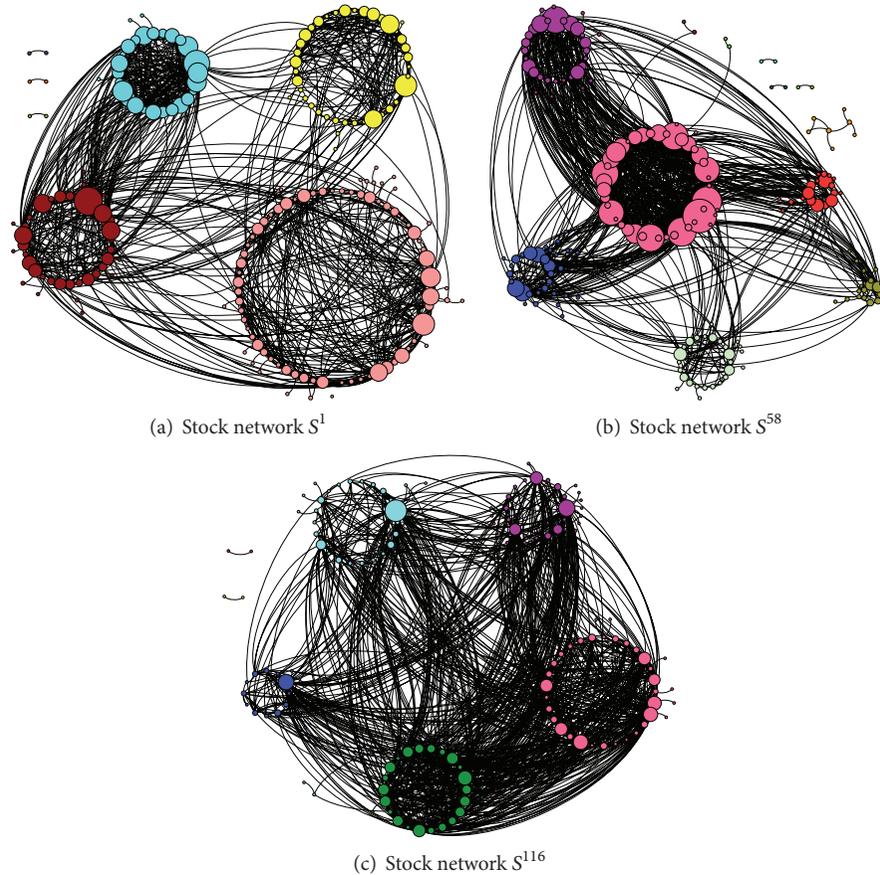


FIGURE 5: (Color online) Several pictures of stock networks S^t . Different node colors represent different communities and node sizes reflect its degree.

in the literature [15]. The method is a process of finding counterparts. Communities are matched from consecutive time windows in descending order of their relative node overlap (i.e., Jaccard similarity coefficient). The relative node overlap between communities A and B is defined as $J(A, B) = |A \cap B|/|A \cup B|$, where $|A \cap B|$ is the number of nodes in the intersection of A and B and $|A \cup B|$ is the number of nodes in the union of two communities. When a community has no counterpart from communities in the previous or the next time window, it is considered as a newborn community or finished its life, respectively.

4. Characters of Dynamic Communities in Stock Networks

First, we investigate some basic statistic properties characterizing the dynamic of stock networks, which are distributions of the coverage of networks (the ratio of nodes contained in a network), the community number, the modularity value Q , and the overall community size. The results are shown in Figure 6. There are overall 961 communities that can be detected, in all time windows. The maximum size of communities is 83 and the minimum is 2. In Figure 6(a),

we show the overall community size distribution, which resembles a power-law distribution.

From Figures 6(b), 6(c), and 6(d), we can find that these three curves have similar variation tendency. When the coverage ratio gets smaller, as fixed edges number of networks, connections of nodes will become denser and community structure will become more indistinct. Using the bear market in the period from October 2007 to August 2010 as a reference again, it is clearly shown in the figures that these three curves are in a low level at this period. This implies that when the market declines a few stocks will own stronger connections with each other and these connections are so tight that the network cannot make distinct community structure. Conversely, when the market is good, more stocks will own strong connections and easily form communities. Furthermore, this gives us a new inspiration; the modularity level of stock networks can reflect that the market is good or bad.

Second, we consider a basic quantity characterizing a dynamic community with its age α , representing the time passed since its birth. There are 243 dynamic communities that can be extracted from all communities in time windows. The average age of dynamic communities is 3.95. Figure 7 illustrates the age distribution of dynamic communities, and we can find that it displays power-law shape. Most age of

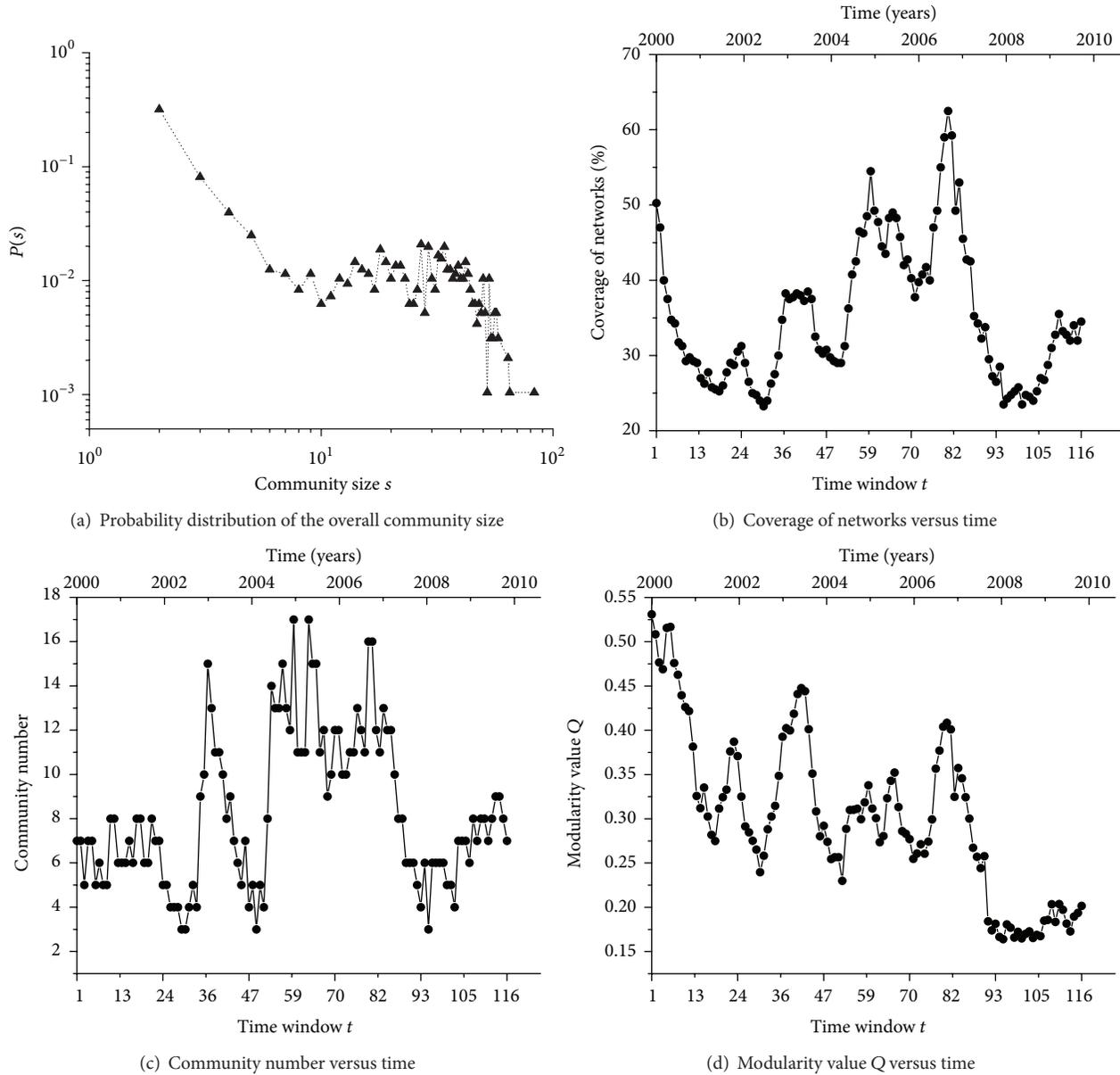


FIGURE 6: Plots of several basic statistic properties of stock networks.

dynamic communities is very small, 92.6% less than 10, which reveals the high dynamic nature of stock market in part. Figures 8 and 9 show the correlations of the dynamic community age α with the start size s of a dynamic community and the dynamic community stationarity ζ , respectively. The stationarity ζ_i of a dynamic community (say $d c_i$) is defined as the average correlation between subsequent states, $\zeta_i \equiv [\sum_{t=t_0}^{t_{\max}-1} J(d c_i^t, d c_i^{t+1})] / (t_{\max} - t_0)$, where t_0 denotes the birth of the community $d c_i$, t_{\max} is the last step before the extinction of the community $d c_i$, and J denotes the Jaccard similarity coefficient mentioned in Section 3. The stationarity ζ represents the stability of community components during the lifetime of a dynamic community. The larger the ζ is,

the smaller the change of components is. From Figure 8, we cannot found clear correlations between the size and the age. It is not as we expected; we thought larger communities may be on average older, just like social relation networks in the literature [15]. The correlation between the stationarity and the age is relatively more obvious, as shown in Figure 9. It is suggested that the value around 0.7 is easy to form older dynamic communities.

Intuitively, the tightly linked community will probably have longer lifetime. To verify this guess, for each community, we measured the total weight inside the community (W_{in}) as well as outside the community (W_{out}). Then, we calculated the average age ($\langle \alpha \rangle$) as a function of $\langle W_{in} / (W_{in} + W_{out}) \rangle$

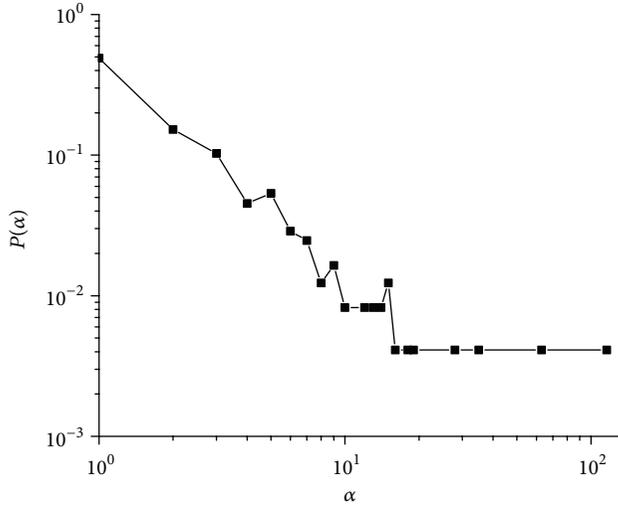


FIGURE 7: Probability distribution of the dynamic community age α .

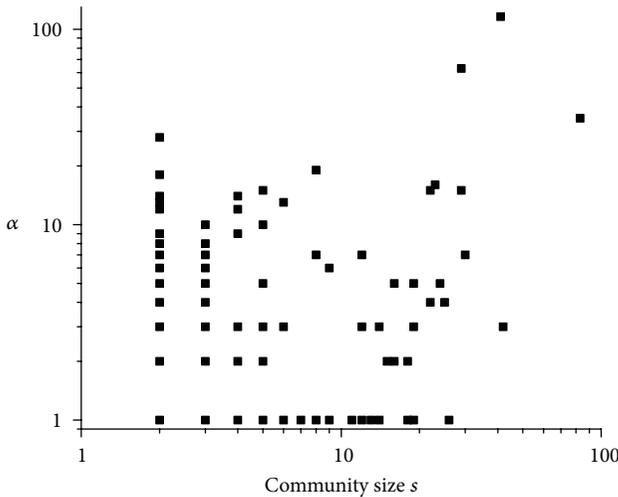


FIGURE 8: Scatter plot of correlations between the dynamic community age α and the start size s of a dynamic community.

of a dynamic community. But, from Figure 10, the curve reaches its peak at 0.6; we find that more tightly dynamic communities are not necessarily older.

5. Summary and Further Studies

In summary, we have introduced some characters of dynamic communities in stock networks, which we have studied recently. The way of constructing stock networks and detecting communities can also be used in other similar complex systems. There are two main results obtained in this paper. First, the clarity of community structure in stock networks can reflect fluctuation of the market. The modularity value Q can reflect clarity of community structure well. So the fluctuation of the modularity value Q gives us a new viewpoint for observing economic changes. Second, from statistical analyses, we find that the tight link and the slow change

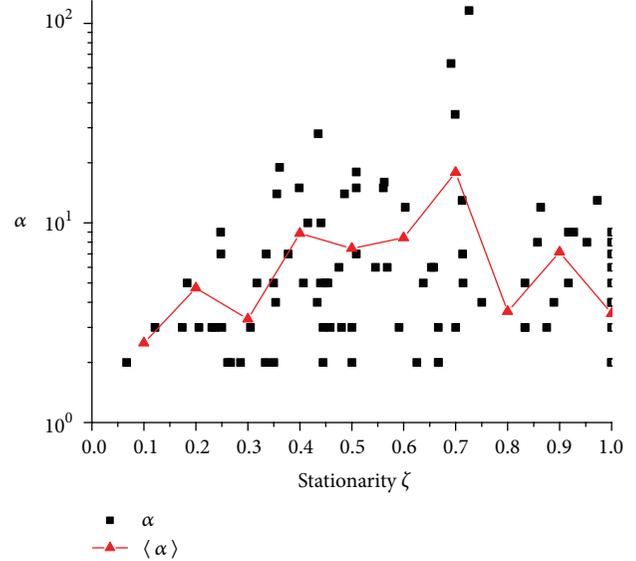


FIGURE 9: Scatter plot of correlations between the dynamic community age α and the dynamic community stationarity ζ . $\langle \alpha \rangle$ represent the arithmetic average value of α .

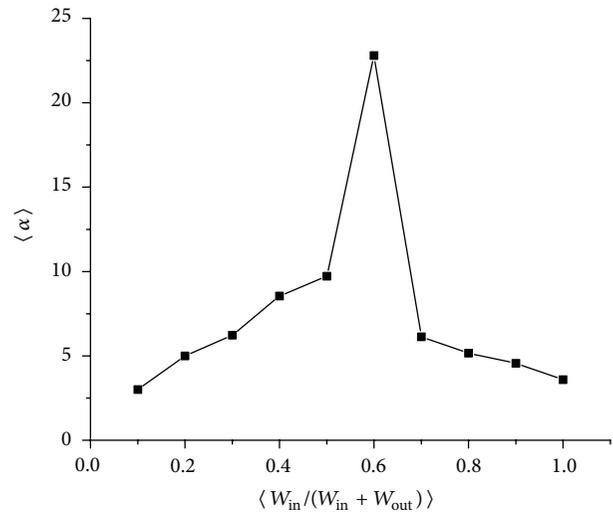


FIGURE 10: Plot of correlations between the average dynamic community age $\langle \alpha \rangle$ and the average community tightness parameter $\langle W_{in}/(W_{in} + W_{out}) \rangle$ in a dynamic community.

are both not good for a dynamic community's long life. Instead, when tight extent of a community ($W_{in}/(W_{in} + W_{out})$) and stationarity of a community (ζ) are about 0.6 or 0.7, a community will be easier to have long life. By the way, this is a coincidence or there are some correlations with the famous golden ratio (about 0.618) or 30 : 70 Pareto Principle, which need further research.

Our results potentially contribute to financial market analysis and decision-making. Furthermore, there are many problems that need further research. For example, from Figure 5, we can find that there are dense links between several communities and some nodes densely link to not only

one community. This implies that stock networks have hierarchical and overlapped community structure. The study of hierarchical and overlapped communities in stock networks may reveal more interesting phenomena. We also study some statistical properties of a single node, such as the probability of leaving its community, the lifetime of a single node in a dynamic community, and the weight ratio between inside a community and outside a community. But we have not found a clear correlation between these properties of a single node. This is because of the high dynamic nature of financial market or the lack of our empirical data, which will be studied in future research. In this paper, empirical data are based on the Hong Kong stock market; characters of dynamic communities in other areas or economies are also worthy of further study.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the China Postdoctoral Science Foundation (no. 2013M542392), the National Science and Technology Support Program (no. 2012BAF12B19), and the National Natural Science Foundation of China (no. 11361033).

References

- [1] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [2] A.-L. Barabási and J. Frangos, *Linked: The New Science of Networks*, Basic Books, 2002.
- [3] B. Barzel and A.-L. Barabási, "Universality in network dynamics," *Nature Physics*, vol. 9, no. 10, pp. 673–681, 2013.
- [4] R. N. Mantegna, "Hierarchical structure in financial markets," *European Physical Journal B—Condensed Matter and Complex Systems*, vol. 11, no. 1, pp. 193–197, 1999.
- [5] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, "Dynamics of market correlations: taxonomy and portfolio analysis," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 5, Article ID 056110, 12 pages, 2003.
- [6] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, "Asset trees and asset graphs in financial markets," *Physica Scripta*, vol. 2003, no. 6, article 48, 2003.
- [7] L. Kullmann, J. Kertész, and R. N. Mantegna, "Identification of clusters of companies in stock indices via Potts superparamagnetic transitions," *Physica A: Statistical Mechanics and Its Applications*, vol. 287, no. 3–4, pp. 412–419, 2000.
- [8] V. Boginski, S. Butenko, and P. M. Pardalos, "Statistical analysis of financial networks," *Computational Statistics & Data Analysis*, vol. 48, no. 2, pp. 431–443, 2005.
- [9] D. Jallo, D. Budai, V. Boginski, B. Goldengorin, and P. M. Pardalos, "Network-based representation of stock market dynamics: an application to american and swedish stock markets," in *Models, Algorithms, and Technologies for Network Analysis*, pp. 93–106, Springer, New York, NY, USA, 2013.
- [10] A. Vizgunov, B. Goldengorin, V. Kalyagin, A. Koldanov, P. Koldanov, and P. M. Pardalos, "Network approach for the Russian stock market," *Computational Management Science*, vol. 11, no. 1–2, pp. 45–55, 2014.
- [11] A. Vizgunov, B. Goldengorin, V. Zamaraev et al., "Applying market graphs for russian stock market analysis," *Journal of the New Economic Association*, vol. 15, no. 3, pp. 66–81, 2012.
- [12] W.-Q. Huang, X.-T. Zhuang, and S. Yao, "A network analysis of the Chinese stock market," *Physica A: Statistical Mechanics and Its Applications*, vol. 388, no. 14, pp. 2956–2964, 2009.
- [13] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.
- [14] W. B. Arthur, S. N. Durlauf, and D. A. Lane, *The Economy as an Evolving Complex System II*, vol. 28, Addison-Wesley, Reading, Mass, USA, 1997.
- [15] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, no. 7136, pp. 664–667, 2007.
- [16] J.-P. Onnela, *Taxonomy of financial assets [M.S. thesis]*, Helsinki University of Technology, Helsinki, Finland, 2002.
- [17] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Article ID P10008, 2008.
- [18] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 2, Article ID 026113, 2004.