Network evolution of the Chinese stock market:

a study based on the CSI 300 index

Abstract

As an emerging market, Chinese stock market is playing an increasingly important role in global financial system. This market deserves more detailed studies for its potential to develop with the progress of Chinese financial reform. By analyzing topological properties and their temporal changes, this paper provides a new perspective of network evolution for Chinese stock market with the emphasis on interdependencies among stocks. The sample of this study is the selected constituent stocks of CSI 300 index. We empirically analyze correlation matrices and correlation-based networks by employing rolling window approach. In the study, the small world property of the network and positive correlations between stocks are found and some key stocks even play important roles to exert more influences on the others. Further study demonstrates the close relationship between network structure and market fluctuation.

JEL classification numbers: C13, C53, G11, G17

Keywords: Correlation matrix, Network analysis, Influence strength, Centrality, CSI 300 index

1 Introduction

Globalization is integrating the national or regional economies that were ever loosely connected with each other. A complex network is coming into being, which interconnects these economies worldwide. Economic network has provided a new approach that stresses the complexities and interdependencies between economic entities (Schweitzer et al., 2009).

Financial markets are ever expending to the scale of globalization with tremendous volumes that can even give a shock to the world economy. Financial stability is raising an extraordinary attention from the academia and political societies since the breakout of this financial crisis. The network approach actually shows a new way to study the interconnected financial systems.

Regarding economic networks, much previous research has shown us a variety of empirical results from specific perspectives. For example, Mahutga (2006) studies the international trade network and further maps this network to the structure of the international division of labor (Mahutga & Smith, 2011). Some studies have investigated risk and contagion in interbank markets by analyzing the network structure in different countries (Boss et al., 2004; Iori et al., 2008; Li et al., 2010).

As one of the most important financial markets, the stock market network has also attracted researchers from the disciplines of economics, physics and systems science. This multidisciplinary feature provides us a new paradigm to further understand the structural and dynamic characteristics of stock markets. Usually, the pairwise relationship between stocks can be used to construct the stock network. And correlation coefficients are computed based on time series of the stock prices or their logarithmic values.

Mantegna (1999) shows the hierarchical structure of the stocks selected from the constituent stocks of the Dow Jones Industrial Average (DJIA) index and the Standard and Poor's 500 (S&P 500) index. The subsequent research efforts continue to study the stock network by considering time horizons (Bonanno et al., 2001; Bonanno et al., 2004; Tumminello et al., 2007). Besides the previous study on daily prices, the time horizon is decreased to cover the 1/20, 1/10, 1/5 and 1/2 of one trading day time horizon, that is 6 hours and 30 minutes (23400 seconds) in the New York Stock Exchange (NYSE). The structure of the minimum spanning tree (MST) varies with the time horizon (Bonanno et al., 2004). The Epps effect is observed and the correlation weakens when the time horizon decreases (Epps, 1979; Bonanno et al., 2001).

The dynamics of the stock network can be investigated from the temporal changes of structural properties such as average path length, clustering coefficients and centralities, which are fundamental concepts in network analysis. Liu & Tse (2012) study the world stock market by considering the cross-correlations between 67 stock market indices and daily closing values of these indices are from Morgan Stanley Capital Investment (MSCI). Their research demonstrates similar behavior in developed markets while independent in emerging markets. The rolling window approach is used to reflect the network evolution (Liu & Tse, 2012; Peron et al., 2012; Bury, 2013). The topological properties between abnormal status (crisis period) and normal status are also compared (Onnela et al., 2002; Kumar & Deo, 2012).

Construction of correlation matrix is usually the start point of network analysis on stock market. The correlation matrix can describe a fully connected network where there exist linkages between each stock. Filtering techniques can be employed to remove some redundant or less important information. For example, minimum spanning tree (MST) and planar maximally filtered graph (PMFG) can be used to simplify the network (Mantegna, 1999; Tumminello et al., 2005). Threshold method is another way to remove the weak connections (Boginski et al., 2005; Huang et al., 2009).

This paper attempts to investigate the Chinese stock market, which is an emerging market and will potentially play a more important role with the progress of Chinese financial reform. By far most of the aforementioned research focuses on the NYSE. The emerging markets are only examined by a few researchers (Pan & Sinha, 2007; Huang et al., 2009; Gałązka, 2011; Kantar et al., 2012). Besides, this paper focuses on the network evolution of the Chinese stock market by investigating the dynamic changes in topological properties. The rest of this paper is organized as follows: Section 2 explains the dataset used in this study; Section 3 constructs correlation matrices and conducts analysis on them; Section 4 derives the network structure based on the correlation matrices and studies the network dynamics by analyzing topological properties and their temporal changes; Section 5 further investigates the relationship between network structure and market fluctuation; Section 6 discusses the empirical results and gives some economic explanations; Section 7 concludes with limitation and future work.

2 Dataset

The CSI 300 index is a comprehensive stock index that includes 300 stocks traded in the Shanghai Stock Exchange (SSE) or the Shenzhen Stock Exchange (SZSE). Table 1 shows the 2012 market value of the 300 constituent stocks. The CSI 300 stocks totally account for 72.5% of the aggregate value in the tradable

A-share market. Thus it is meaningful to study the Chinese stock market by investigating this stock index.

Since the constituent stocks are adjusted semiannually or sometimes temporarily, we collect 176 stocks out of 300 by comparing the two dates – 5 January 2009 and 31 December 2013 – and selecting the stocks that are common in these two days. This procedure is also meaningful to remove some influences from abnormal events caused by specific companies. As shown in Table 1, the market value of these 176 stocks still accounts for 54.5% of the total tradable A-share market. Moreover, among the 176 stocks, the aggregate market value of the top 30 stocks has reached 36.4% of the Chinese A-share market, and the top 90 stocks account for 47.7% of the A-share. These statistics about market value ensure that we can study the Chinese stock market by investigating the sampled stocks.

The daily closing prices of each stock in the sample are collected from the Wind database. The corresponding time period is from January 2009 to December 2013 with 1211 observations. The time series of the logarithmic return are used to construct the correlation matrices for analysis.

Table 1 CSI 300 market value (31 December 2012)

	Assets	Sales Revenue	Market Value (Total)	Market Value (A share)	Market Value (Total Tradable)	Market Value (A Tradable)
Listed Stock (A) (Billion CNY)	119357	24623	24298	21073	19954	16778
CSI300 (Billion CNY)	109774	18213	16613	13511	15267	12169
CSI300(176) (Billion CNY)	89690	14880	12537	9692	11989	9148
CSI300/A (%)[a]	0.920	0.740	0.684	0.641	0.765	0.725
CSI300 (176)/A (%)[b]	0.751	0.604	0.516	0.460	0.601	0.545
176/300 (%)[c]	0.817	0.817	0.755	0.717	0.785	0.752

a) the aggregate value of the 300 constituent stocks of CSI 300 index divided by the aggregate value of the A-share stocks;
b) the aggregate value of the 176 constituent stocks of CSI 300 index divided by the aggregate value of the A-share stocks;
c) b divided by a.

3 Correlation Matrices

3.1 Rolling window approach

Correlation matrices contain information of interactions. The rolling window approach is employed to investigate temporal changes of the stock market. Cross-correlation coefficients between stocks are computed in each time window. This process can be formalized as the following:

n_o − the number of observations

w − the window size, i.e., the number of observations contained in a time window

n_w − the number of windows

For the case of one-day interval for the rolling window,

$$n_w = n_o - w + 1 \ .$$

Let S denote the set of the stocks in the sample, then

for stock i and j, $i, j \in S$, in the k-th window, $k = 1, 2, \dots, n_w$,

$$X_{ik} = (x_{i,k}, x_{i,k+1}, \dots, x_{i,k+w-1})$$

$$X_{jk} = (x_{j,k}, x_{j,k+1}, \dots, x_{j,k+w-1})$$

where the vector X_{ik} denotes the set of the observations of stock i falling in the k-th time window while the small $x_{i,k}$ denotes the k-th observation in the time series for stock i. Similar is X_{jk}. Thus, the correlation between stock i and stock j can be computed,

$$\rho_{i,j,k} = \frac{cov(X_{ik}, X_{jk})}{\sigma_{X_{ik}} \cdot \sigma_{X_{jk}}}$$

where cov is the covariance and σ is the standard deviation.

Accordingly, for each time window with the size w, the correlation coefficients between each stock in the sample can be calculated and the correlation matrix is obtained. Finally, there are n_w correlation matrices for further analysis. In this study, 1210 observations of logarithmic returns for each stock are used (one less than the number of the closing prices for return calculation). The window size is set to 100 trading days, covering about four and a half months in calendar dates. Thus the total number of the time windows is 1111, corresponding to the dates from 5 June 2009 to 31 December 2013.

3.2 Probability density function

From the correlation matrix in a certain time window, we can depict the corresponding probability density function (PDF) and then plot the PDFs along time windows, as shown in Figure 1-1.

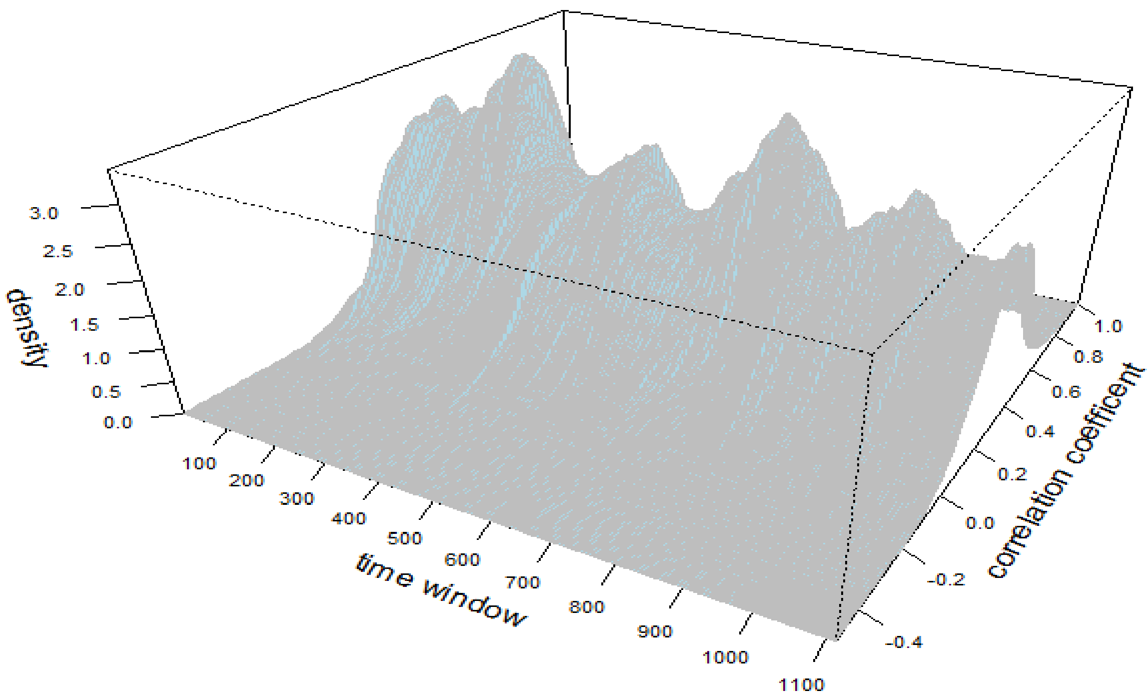

Figure 1-1 Probability density functions along the time windows

For each time window, we can have kurtosis and skewness of the PDF as well as mean, median and standard deviation (sd) of correlation coefficients, which form new time series by rolling the time window, shown in Figure 1-2.

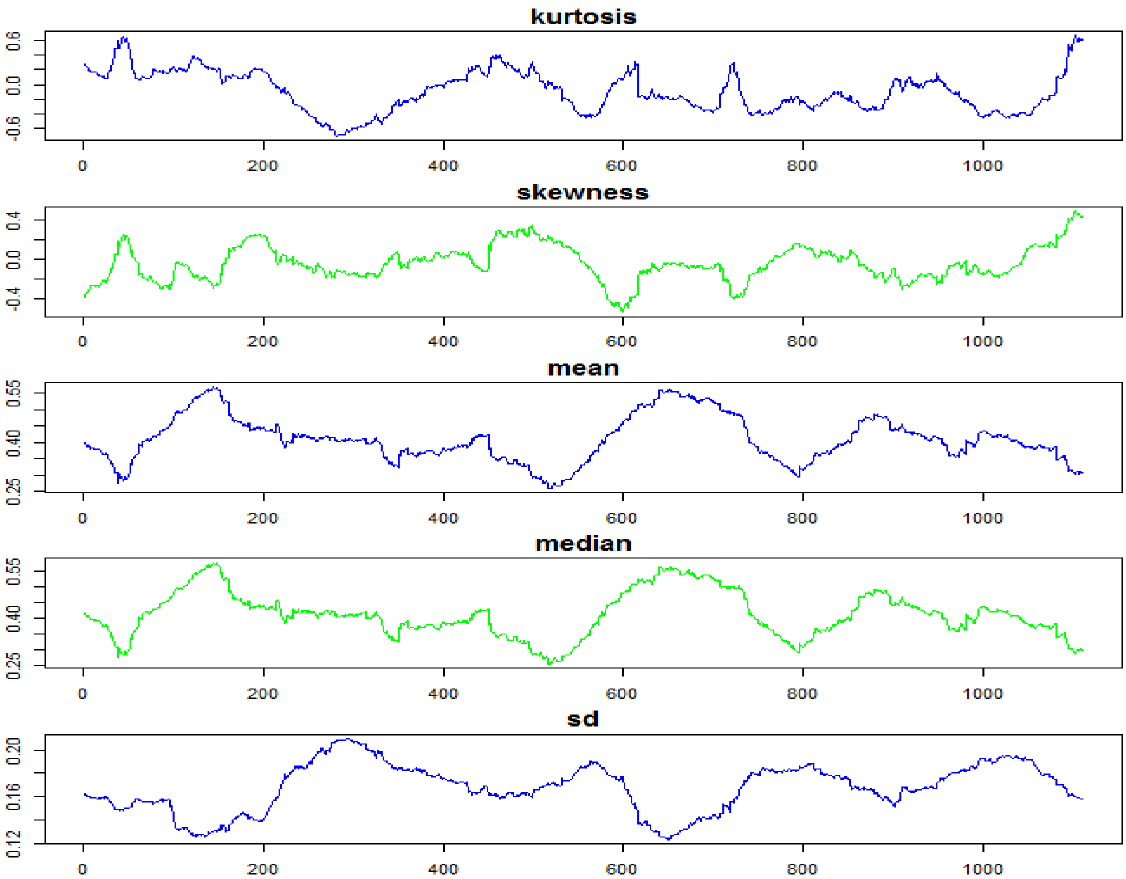

Figure 1-2 Temporal variation of statistics for the correlation matrices

Table 2 Statistics on all the time windows

	mean	median	std	max	q3	q1	min
corr-mean	0.4096	0.4046	0.0695	0.5701	0.4453	0.3634	0.2605
corr-sd	0.1679	0.1694	0.0205	0.2101	0.1830	0.1558	0.1229
corr-max	0.9435	0.9439	0.0167	0.9762	0.9559	0.9332	0.8999
corr-min	-0.2221	-0.2730	0.1370	0.1174	-0.1561	-0.3263	-0.4120
corr-median	0.4133	0.4103	0.0717	0.5780	0.4526	0.3672	0.2564
skewness	-0.0413	-0.0615	0.1766	0.4913	0.0697	-0.1542	-0.5371
kurtosis	-0.0952	-0.1288	0.2758	0.6724	0.1158	-0.3144	-0.7063

* q1 and q3 represent the first and third quartile respectively.
* corr-mean is the mean of the correlation coefficients in each time window; similarly for standard deviation (corr-sd), maximum (corr-max), minimum (corr-min), median(corr-median); skewness and kurtosis are the parameters of the probability density function for each time window.

Table 2 further lists the statistics of the variables in Figure 1-2. The kurtosis is the deviation from the normality value 3, showing the fluctuation between positivity and negativity but mostly lower than the normal distribution. The mean of the skewness (-0.0413) shows the positively skewed PDFs, along with the correlations mean (0.4096), illustrating the positive correlation or synchronization in the Chinese stock market. The similar profile of the mean and the median along the time window in Figure 1-2 shows the similar temporal variation, also with the very similar statistics in Table 2.

3.3 Influence strength

Influence strength is introduced to analyze the influence of an individual stock on the other stocks in the market (Kim et al., 2002; Gałązka, 2011). For stock *i* in the window *k*, the influence strength (*IS*) is defined as the following,

$$IS_{i,k} = \frac{\sum_j \rho_{i,j,k}}{N-1}, i \neq j \; and \; i,j \in S$$

where *N* denotes the number of stocks in the sample set *S*.

The average influence of stock *i* on the whole temporal windows is

$$IS_i = \frac{\sum_k IS_{i,k}}{n_w}, k = 1,2,\dots,n_w.$$

Table 3-1 The stocks with the top 10 largest IS values

Code	Code2	MV_Rank	IS_Mean	IS_Rank	N_Top10	N_Top20	N_Top30
601899	ZJKY	49	0.5381	1	760	858	897
600886	GTDL	66	0.5295	2	534	762	963
601398	GSYH	2	0.5258	3	490	798	966
600664	HYGF	152	0.5208	4	455	694	836
600497	CHXZ	124	0.5168	5	352	520	736
600804	PBS	96	0.5081	6	197	459	776
600997	KLGF	176	0.5067	7	201	680	833
600811	DFJT	161	0.5002	8	226	438	670
600649	CTKG	89	0.4995	9	167	366	661
600018	SGJT	16	0.4980	10	250	452	582

* Code is the security code listed in the stock exchange; Code2 consists of the initials of Chinese Pinyin;
* MV_Rank gives the corresponding ranking for market value in the sample;
* IS_Mean is the mean of the IS values along the time window; IS_Rank gives the corresponding ranking of IS_Mean in the sample;
* N_Top10 is the times that the rankings of IS position within the top 10; similary for N_Top20, N_Top30.

Table 3-2 Statistics of IS on stock groups

IS	nobs	mean	median	std	max	q3	q1	min
sample	176	0.4096	0.4176	0.0623	0.5381	0.4556	0.3678	0.1922
g1(q0~q1)	44	0.3259	0.3340	0.0396				
g2(q1~q2)	44	0.3954	0.3990	0.0146				
g3(q2~q3)	44	0.4345	0.4321	0.0113				
g4(q3~q4)	44	0.4825	0.4784	0.0210				
h1(mv1~44)	44	0.3908	0.3947	0.0647				
h2(mv45~88)	44	0.4068	0.4108	0.0645				
h3(mv89~132)	44	0.4231	0.4344	0.0616				
h4(mv133~176)	44	0.4168	0.4218	0.0552				

* q1 and q3 represent the first and third quartile respectively;q0 = min, q2 = median, q4 = max;
* g1~g4 respectively corresponds with the group of stocks whose IS value falls within the certain quartiles;
* h1~h4 respectively corresponds with the group of stocks whose market value ranks at certain positions.

Table 3-1 lists the stocks with the top 10 largest IS values while Table 3-2 shows statistics on stock groups. The IS values varies from the largest value of 0.5381 to the smallest value of 0.1922. The variation among the top 10 is from 0.5381 to 0.4980, showing their stronger influences on the other stocks in the market. From the stock groups divided by quartiles, the stocks in the group g1 have the weaker influence on the others with the mean value 0.3259 (std=0.0396) while the stocks in the group g4 show stronger influence with the mean value 0.4825 (std=0.0210).

Besides, the stock groups divided by market value rankings reveal that the group h3 consisting of the stocks with market value rankings from 89 to 132 (smaller market value group in the sample) has the greatest influence than the other stock groups. The smallest market value group h4 still has the higher influence value than the other two bigger market value groups, h1 and h2. The stocks with large market values may not have strong influence on other stocks.

The temporal changes of IS values for the stocks ranking top 3 are illustrated in Figure 2. The left vertical axis labels the IS values (the curve) while the right vertical axis shows the rankings (the scattered points) in the graph. As shown in Figure 2, the profiles of the IS value variation for the three stocks are very similar, indicating the similar trends of their influences on the other stocks, while the rankings show more difference in temporal variation. The dashed horizontal lines mark the position of Ranking 30. Details about the times of rankings in Top 10, Top 20 and Top 30 can be found from the last three columns in Table 3-1.

For example, for the stock ZJKY (Zijin Mining Group Co., Ltd.), in the 1111 time windows, there are 760 times when the ranking is Top 10, accounting for 68.4%. Similarly, the number of the times is 858 (77.2%) for Top 20 and 897 (80.7%) for Top 30. The stock GSYH (Industrial and Commercial Bank of China) shows the largest number of the times for Top 30, i.e., 966 times (about 86.9% of the whole time windows), thus showing its stronger influence on the other stocks in the market. The subsequent stock is GTDL (SDIC Power Holding Co., Ltd.) with 963 (86.7%) times ranking Top 30.

4 Correlation-based Networks

For the correlation matrix in a certain time window, there exists a corresponding network, called correlation-based network. Formally, let $A = (A_1, A_2, \ldots, A_k, \ldots, A_{n_w})$ denote the correlation matrix series for the time windows. For the matrix A_k,

$$A_k = (\rho_{i,j,k})\,.$$

The threshold method is used to construct a network from the correlation matrix. Let θ denote the threshold value to filter the matrix, and then a new matrix $A_k^* = (a_{i,j,k})$ is generated where

$$a_{i,j,k} = \begin{cases} 1, & \rho_{i,j,k} \geq \theta \\ 0, & \rho_{i,j,k} < \theta \end{cases}.$$

By considering this filtered matrix as the adjacency matrix, we can construct an unweighted network and denote it as $\mathbb{N}_k$. Similarly, we can obtain $\mathbb{N}_1, \mathbb{N}_2, \ldots, \mathbb{N}_k, \ldots, \mathbb{N}_{n_w}$, which respectively corresponds with each time window. By network analysis to these networks, time series of network properties can show the temporal changes in the network structure of the stock market.

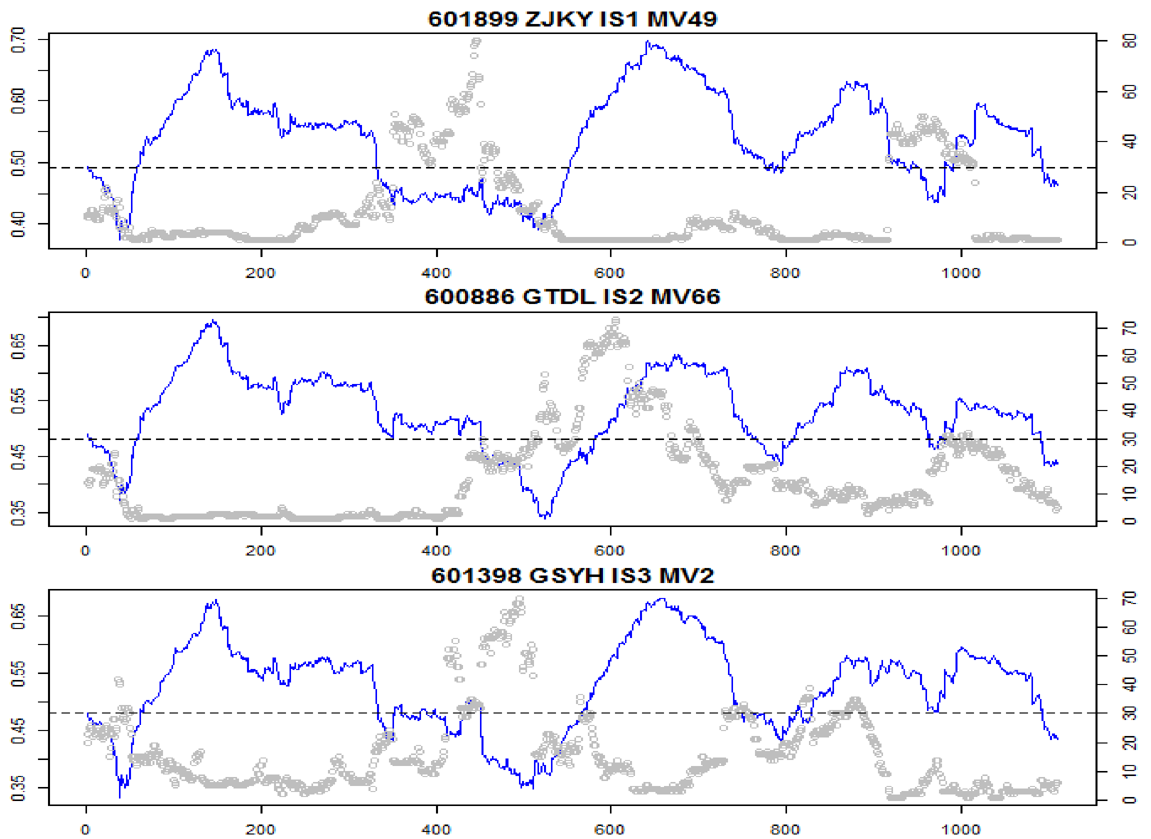

Figure 2 Temporal variation of IS values for the stocks ranking top 3

* The number following IS represents the ranking of IS value; the number following MV represents the ranking of market value.

* The left vertical axis labels the IS values (the curve) while the right vertical axis shows the rankings (the scattered points) in the graph.

The variation of the threshold value will change the density of network connection (Boginski et al., 2005; Huang, et al., 2009; Tse et al., 2010; Brida & Risso, 2010). Different from previous research, this study uses a dynamic threshold to obtain the filtered matrix. For a certain time window, the median of the correlation coefficients in the correlation matrix is used as the threshold. In this way, the density of network connection is kept nearly as a constant (i.e. the normalized value 0.5) since nearly half of the connections are removed from the original fully-connected networks.

Network metrics such as clustering coefficient, average path length and centralities are employed to investigate the topological changes in the network structure (Bonacich, 1987; Wasserman & Faust, 1994; Watts & Strogatz, 1998; Newman, 2003; Newman, 2008).

4.1 Clustering coefficient

Consistent with the graph theory, the network $\mathbb{N}_k$ corresponds to a graph $G = (V, E)$. V is the set of vertices in the graph while E is the set of edges. For $v_i, v_j \in V$, $e_{ij} \in E$, let V_i^{NE} denote the set of neighbor vertices of the vertex v_i, and then

$$V_i^{NE} = \{v_j: e_{ij} \in E\} \ .$$

$|V_i^{NE}|$ represents the size of this set, i.e., the number of the neighbor vertices of the vertex v_i.

The clustering coefficient of the vertex v_i can be formally defined as

$$C_i = \frac{2|\{e_{jm}:\ e_{jm} \in E,\ \ v_j, v_m \in V_i^{NE}\}|}{|V_i^{NE}|\,(|V_i^{NE}|-1)}.$$

The clustering coefficient of the network $\mathbb{N}_k$ is the average of the clustering coefficients for all the vertices, i.e.,

$$C = \frac{\sum_i C_i}{|V|}.$$

Similarly, the temporal changes can be shown by the time series $C_{\mathbb{N}_1}, C_{\mathbb{N}_2}, \dots, C_{\mathbb{N}_k}, \dots, C_{\mathbb{N}_{nw}}$.

The middle panel in Figure 3 shows the temporal changes of the clustering coefficients. It can be seen that the clustering coefficients still vary even though the connection density is controlled as a constant in this research.

4.2 Average path length

Let d_{ij} denote the length of the shortest path between v_i and v_j, that is also called the distance between v_i and v_j, and it can be measured by the number of edges in the shortest path. Then average path length (APL) can be defined as

$$APL = \frac{2\sum_{v_i, v_j \in V} d_{ij}}{|V|(|V|-1)}.$$

The bottom panel in Figure 3 shows the temporal variation of average path length. The correlation between APL (clustering coefficient) and CSI index is -0.4289 (-0.5727), revealing the clustering effect during the bearish period of the Chinese stock market.

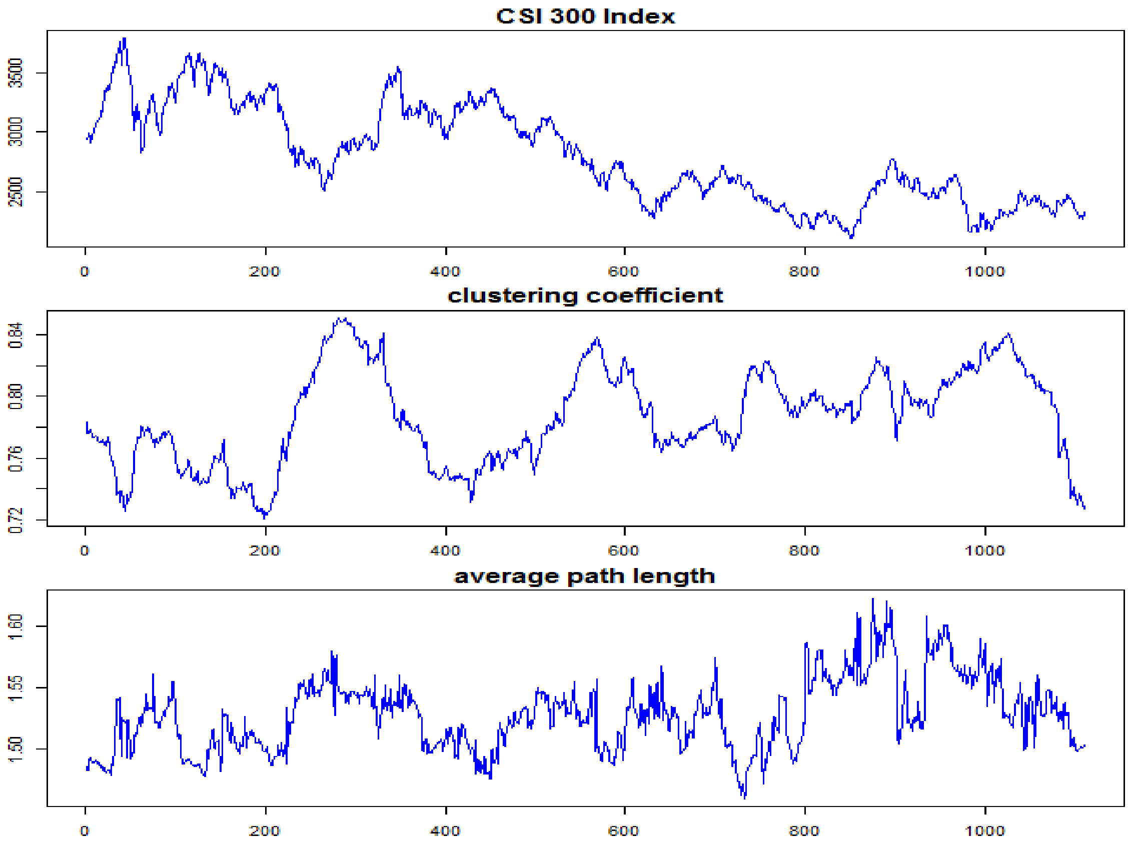

Figure 3 Temporal variation of CSI 300 index, clustering coefficient and APL

4.3 Centrality

Centrality metrics such as degree, closeness, betweenness and eigenvector are used to gauge the topological properties of a network.

Degree Centrality measures the number of links between a vertex and the other vertices. In the graph G, the degree of a vertex v_i is defined as

$$degree(v_i) = \left|\{e_{ij}: e_{ij} \in E, v_i, v_j \in V, v_i \neq v_j\}\right|.$$

Degree centrality can also measure the connection density of the network when the sum of the degrees of each vertex is divided by the number of vertices.

Closeness centrality measures the average geodesic distance from a vertex to all the other vertices. The closeness centrality of the vertex v_i is defined as

$$closeness(v_i) = \frac{1}{\sum_{\substack{v_j \in V \\ v_i \neq v_j}} d_{ij}},$$

where d_{ij} is the shortest path length between the vertex v_i and v_j. The higher the closeness score is, the closer the vertex is to the other vertices.

Betweenness centrality measures the intermediating role of a vertex. For the vertex v_i,

$$betweenness(v_i) = \sum_{v_p, v_q \in V} \frac{NSP_{pq}^i}{NSP_{pq}}.$$

NSP_{pq} denotes the number of all the shortest paths linking the vertex v_p and v_q. As part of NSP_{pq}, NSP_{pq}^i is the number of the shortest paths that satisfy: (i) linking v_p and v_q; (ii) passing through v_i.

Eigenvector centrality measures a vertex by considering the importance of its neighbor vertices. A neighbor vertex with more neighbors usually contributes more than the ones with less neighbors. Let $\mathbb{e}_i$ denote the eigenvector centrality of the vertex v_i, considering the contribution of neighbor vertices (Bonacich, 1987; Newman, 2003),

$$\mathbb{e}_i = \frac{1}{\lambda} \sum_{v_j \in V} a_{ij}\, \mathbb{e}_j,$$

where λ is constant. By transformation, the eigenvector equation is as the following:

$$\mathbb{A}\mathbb{E} = \lambda \mathbb{E}.$$

For the k-th time window, $\mathbb{A}$ is equally to the matrix A_k^*. There exist many solutions and usually λ is preferred to be the largest eigenvalue while $\mathbb{E}$ is the corresponding eigenvector.

Figure A-1 in Appendix shows the temporal variation of the mean and standard deviation of the network centralities, and Figure A-2 shows the median values. Table 4 gives the statistics of the network metrics along the time window.

Table A-1 gives the detailed information about the stocks ranking top 10 for centrality measures such as Panel A for degree, Panel B for closeness, Panel C for betweenness and Panel D for eigenvector. Table A-2 further lists the 12 stocks with the four centrality measures all ranking top 30. There are only 3 stocks, ZJKY, GSYH and DFDQ, with the four centrality measures ranking top 10. Comparing Table A-2 with Table 3-1, seven of twelve stocks with the high-score centralities also have strong influences on the other stocks.

Table 4 Statistics of network metrics

	mean	median	std	max	q3	q1	min
clustercoef	0.7865	0.7867	0.0305	0.8504	0.8107	0.7647	0.7208
apl	1.5279	1.5271	0.0291	1.6228	1.5465	1.5043	1.4597
degree-mean	0.5000	0.5000	0.0005	0.5000	0.5000	0.5000	0.4944
degree-sd	0.2500	0.2524	0.0160	0.2751	0.2631	0.2389	0.2111
degree-max	0.8441	0.8400	0.0299	0.9257	0.8571	0.8229	0.7771
degree-min	0.0002	0.0000	0.0012	0.0114	0.0000	0.0000	0.0000
degree-median	0.5747	0.5714	0.0291	0.6514	0.5957	0.5543	0.5086
close-mean	0.2193	0.1785	0.1177	0.6733	0.2821	0.1489	0.0756
close-sd	0.0339	0.0280	0.0182	0.1299	0.0369	0.0254	0.0182
close-max	0.2488	0.1953	0.1494	0.8929	0.3170	0.1619	0.0824
close-min	0.0143	0.0057	0.0512	0.3676	0.0057	0.0057	0.0057
close-median	0.2265	0.1845	0.1199	0.6917	0.2897	0.1549	0.0809
betw-mean	0.0029	0.0029	0.0002	0.0035	0.0030	0.0028	0.0023
betw-sd	0.0035	0.0034	0.0006	0.0059	0.0038	0.0031	0.0025
betw-max	0.0233	0.0217	0.0077	0.0476	0.0267	0.0176	0.0124
betw-min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
betw-median	0.0018	0.0018	0.0003	0.0028	0.0020	0.0017	0.0013
eigen-mean	0.0675	0.0674	0.0008	0.0693	0.0680	0.0668	0.0660
eigen-sd	0.0336	0.0338	0.0016	0.0364	0.0349	0.0327	0.0297
eigen-max	0.1056	0.1053	0.0032	0.1126	0.1075	0.1030	0.0989
eigen-min	0.0000	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000
eigen-median	0.0800	0.0798	0.0029	0.0872	0.0820	0.0778	0.0737

* q1 and q3 represent the first and third quartile respectively.

* degree-mean is the mean of the degrees in the network corresponding to a time window; similarly for standard deviation (xx-sd), maximum (xx-max), minimum (xx-min), median(xx-median) where xx represents the centrality such as degree, closeness, betweenness and eigenvector.

5 Structural Changes and Market Fluctuation

Previous sections make analysis on the market structural dynamics from correlation matrices and correlation-based networks. This section attempts to explore the relationship between structural changes and market fluctuation.

The variation of topological properties such as clustering coefficient, average path length and centrality metrics are proxies for the structural changes. The logarithmic returns (LGRn) of the CSI index are proxies for market fluctuation, picking n from 1, 55 and 100, which respectively represents the one-day, 55-day and 100-day logarithmic return, i.e.,

$$LGRn_t = log(p_t) - log\ (p_{t-n}).$$

Further, the future logarithmic returns (FRn) are used to investigate the predicting role of the structural changes on market fluctuation.

$$FRn_t = log(p_{t+n}) - log\ (p_t).$$

5.1 Correlations between structural properties and market fluctuation

Table A-3 gives the correlations between the time series of structural properties and market fluctuation along the time window. The absolute values of the correlation coefficients tend to increase from LGR1 to LGR100, revealing that the structural properties seem to have stronger long-range relationship with market fluctuation. A similar tendency remains in the future logarithmic returns series, i.e., FR1, FR55 and FR100.

By contrast with mean and median of cross-correlations, standard deviation seems to have a stronger negative relation with market fluctuation. This means that more heterogeneity occurs when the market enters into a bearish or crisis period. The kurtosis seems to have a positive relation with the logarithmic return.

The clustering coefficient shows a negatively stronger relation with market fluctuation. A significant negative value shows the clustering effect when the market is bearish, consistent with the US stock market (Peron et al., 2012) and global stock indices (Sensoy et al., 2013).

The mean of eigenvector centrality shows the stronger positive relation with market fluctuation. The standard deviations of degree and eigenvector centrality show the significant negative relation with market fluctuation.

5.2 Logistic modeling

The correlation analysis has shown the relationship between structural properties and market fluctuation. This section uses logistic model to demonstrate the predicting function.

Table 5 The predicting variables in the logistic models

		Model I	Model II	Model III
1	corr-mean	●		
2	corr-median		○	
3	corr-sd			●
4	skewness	●	●	●
5	kurtosis	●	●	●
6	Cluster Coeff.	○	●	●
7	APL	●	●	●
8	degree-median		●	
9	degree-sd			●
10	close-mean	○		
11	close-median		○	
12	close-sd			○
13	betw-mean	●		
14	betw-median		●	
15	betw-sd			●
16	eigen-mean	◎		
17	eigen-median		●	
18	eigen-sd			◎

* For example, Model I has 8 predicting variables initially (marked by three types of circles). The predicting variables marked by ● is significant in both the stepwised model and the initial model; the predicting variables marked by ○ is only in the initial model with no significance; the predicting variables marked by ◎ is not significant in the initial model but is remained in the stepwised model with significance. The significance level is 0.10. The dependent variables are severn UTn and the predicting variable is considered as significant when it is significant in the models for more than three UTn.

Table 5 lists the predicting variables used in the logistic models. Besides some common variables, the three models take into consideration the mean, median and standard deviation separately. The dependent variables are UTn, dichotomized from FRn where n is picked from 1, 5, 10, 21, 55, 100 and 200. If FRn is less than 0, then UTn is 0; otherwise, UT1 is 1.

Table 6 Predicting market up and down by using network structural properties

			UT1	UT5	UT10	UT21	UT55	UT100	UT200
Model I	correct ratio	all	0.550	0.564	0.609	0.648	0.691	0.644	0.806
		up	0.615	0.589	0.550	0.585	0.554	0.504	0.516
		down	0.484	0.540	0.664	0.701	0.781	0.733	0.915
	AIC		1536	1503	1459	1405	1187	1137	820
	test for dispersion		0.451	0.444	0.463	0.315	0.042	0.899	0.070
Model II	correct ratio	all	0.542	0.583	0.631	0.639	0.693	0.679	0.801
		up	0.602	0.607	0.615	0.528	0.542	0.499	0.500
		down	0.481	0.560	0.646	0.735	0.792	0.793	0.915
	AIC		1534	1495	1449	1393	1177	1110	828
	test for dispersion		0.441	0.394	0.417	0.431	0.762	0.938	0.018
Model III	correct ratio	all	0.539	0.575	0.627	0.694	0.732	0.743	0.811
		up	0.641	0.580	0.606	0.607	0.583	0.606	0.560
		down	0.433	0.570	0.646	0.770	0.829	0.830	0.906
	AIC		1536	1499	1442	1381	1176	1049	785
	test for dispersion		0.461	0.409	0.427	0.251	0.172	0.443	0.010

* up is the correct ratio to predict the market ups; down is the correct ratio to predict the market downs; all is the overall correct ratio.
* test for dispersion: null hypothesis – there exists dispersion in the sample; it can assert no dispersion when p is greater than 0.1.
* the dependent variables (UTn) are dichotomized from the future logarithmic returns (FRn).
* the predicting result above is from the stepwise model by removing some close correlated variables in Model I, II, and III.

Table 6 gives the results about the logistic modeling. The prediction for market down seems to have a higher ratio of correctness, which probably means stocks in the market usually fall rather than rise simultaneously. There exists no much difference between mean and median values. The UT100 seems to be better predicted for market direction while the UT200 fails in the dispersion test.

6 Discussion

The dataset in this study is the 176 stocks selected from the CSI 300 index. It is shown that A-share market in China is still monopolized by giant companies in some extent. The data from the Chinese stock exchanges show that there were totally 2472 listed companies in Chinese A-share market in December 2012. All the 176 stocks in the sample, covering only about 7.1% of the listed companies, have accounted for more than half of the stock market from the perspectives of assets, sales revenue and market value.

We construct correlation matrices and compute the related statistics by using rolling window approach. Particularly, the positive mean of correlation coefficients and the negative mean of the skewness denote that the stocks in Chinese A-share market are generally positively correlated and the market shows some synchronization as a whole.

From the correlation matrices and the corresponding networks, some specific stocks are found to play more important roles in the whole market. The top 3 stocks with the largest influence strength in Table 3-1 are ZJKY, GTDL and GSYH. ZJKY is a large state holding mining group specializing in gold and mineral resource exploration and development; GTDL is one of the largest power companies in China; GSYH is the largest commercial bank in China and also ranks top 10 among the worldwide banks based on assets or market capitalization. These stocks are all large companies belonging to the key industries (resource, energy and finance) in the national economy.

By comparing the ranking of influence strength with market value in Table 3-1, most of the stocks with the high-score influence strength, are not stocks with huge market value. This shows that large companies cannot necessarily have strong influence in the stock market. In other words, some small ones can also exert significant influence. Table 3-2 further confirms that the stock group with the smaller market value has the strongest influence.

From the perspective of centrality measures such as degree, closeness and eigenvector, ZJKY, GSYH and GTDL are also ranking top 3 among the sample of 176 stocks, illustrating their stable positions and strong influence strength in the stock network. ZJKY and GSYH still remain the top 10 position for betweeness centrality, showing their important intermediating roles while GTDL becomes a little weaker but still ranks the 16^{th} in this centrality.

The Chinese stock market exhibits the small-world property. Table 4 displays the network metrics, featuring the high cluster coefficient (0.7865) and small average path length (1.5279). The small-world property means that market information can be rapidly disseminated within the network, thus raising the market efficiency and leading to the aforementioned market synchronization.

Furthermore, we investigate the relationship between structural changes and market fluctuation. The absolute values of correlation coefficients increase with the time horizon for computing logarithmic returns, revealing the long range relationship between network structure and market behavior. Further analysis on the correlations, such as the coefficients listed below the columns of LGR100 and FR100, shows that most of the network metrics are negatively related to LGRn. During the bearish or crisis period, the stocks in the network are more tightly connected and negative impact on parts of the network can be easily spread to the whole system. On the contrary, most of these metrics are positively correlated with FRn. One possible explanation of this result is that the high level of correlations during the crisis would remain for a period of time, accompanied with the synchronized recovery of the stocks in the market. And the network would also experience some dynamic changes during this time until it reaches a new state of connection.

Results of logistic modeling demonstrate the relation between network structure and market fluctuation from the predicting perspective. The correct ratio increases with the increasing time lag, indicating the stronger long range influence. Moreover, the network structure properties seem to predict market down much better, which indicates the asymmetry of Chinese stock market between the bearish and bullish period.

7 Conclusion

This paper studies the Chinese stock market by investigating the evolution of the correlation-based network that is constructed from the 176 constituent stocks in the CSI 300 index. The stocks in this market show positive correlations and some stocks play more important roles in strongly influencing other stocks in the market. The stock network has a high clustering coefficient and a short average path length, indicating the feature of a small-world network model. We have also shown the close relation between network structure and market fluctuation. This research can be applied to portfolio management where correlations and market fluctuation are key issues. Network analysis on stock market can provide a new paradigm and further help to develop new tools for portfolio management.

Some limitation exists for the future research. Industrial sectors can be investigated so as to explain clustering feature or stock community in details. In the developed markets such USA, some industrial sectors are shown to have stronger influence in the market. Thus it deserves further analysis on the stocks according to their industrial sectors and thus we can better understand the Chinese stock market and its evolution. It will further help to understand some fundamental and dynamic characteristics of this stock market by comparison with the other markets classified as developed or emerging ones.

References

Boginski, V., Butenko, S., & Pardalos, P. M. (2005). Statistical analysis of financial networks. Computational Statistics & Data Analysis, 48(2), 431-443.

Bonacich, P. (1987). Power and centrality: A family of measures. American journal of sociology, 1170-1182.

Bonanno, G., Caldarelli, G., Lillo, F., Miccichè, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. The European Physical Journal B-Condensed Matter and Complex Systems, 38(2), 363-371.

Bonanno, G., Lillo, F. & Mantegna, R.N. (2001). High-frequency cross-correlation in a set of stocks. Quantitative Finance, 1(1), 96-104.

Boss, M., Elsinger, H., Summer, M., & Thurner, S. (2004). Network topology of the interbank market. Quantitative Finance, 4(6): 677-684.

Brida, J. G., & Risso, W. A. (2010). Dynamics and structure of the 30 largest North American companies. Computational Economics, 35(1), 85-99.

Bury, T. (2013). Market structure explained by pairwise interactions. Physica A: Statistical Mechanics and its Applications, 392(6), 1375-1385.

Epps, T. W. (1979). Comovements in stock prices in the very short run. Journal of the American Statistical Association, 74(366a), 291-298.

Gałązka, M. (2011). Characteristics of the Polish Stock Market correlations. International Review of Financial Analysis, 20(1), 1-5.

Huang, W. , Zhuang, X., & Yao, S. (2009). A network analysis of the Chinese stock market. Physica A: Statistical Mechanics and its Applications, 388(14), 2956-2964.

Iori, G., De Masi, G., Precup, O. V., Gabbi, G., & Caldarelli, G. (2008). A network analysis of the Italian overnight money market. Journal of Economic Dynamics and Control, 32(1), 259-278.

Kantar, E., Keskin, M., & Deviren, B. (2012). Analysis of the effects of the global financial crisis on the Turkish economy, using hierarchical methods. Physica A: Statistical Mechanics and its Applications, 391(7), 2342-2352.

Kim, H. J., Kim, I. M., Lee, Y., & Kahng, B. (2002). Scale-free network in stock markets. Journal of the Korean Physical Society, 40(6), 1105-1108.
Kumar, S., & Deo, N. (2012). Correlation and network analysis of global financial indices. Physical Review E, 86(2), 026101.
Li, S., He, J., & Zhuang, Y. (2010). Network efficiency analysis of Chinese inter-bank market. Joumal of Southeast University, 26(3), 494-497.
Liu, X. F., & Tse, C. K. (2012). A complex network perspective of world stock markets: synchronization and volatility. International Journal of Bifurcation and Chaos, 22(06), 1250142.
Mahutga, M. C. (2006). The persistence of structural inequality? A network analysis of international trade, 1965–2000. Social Forces, 84(4), 1863-1889.
Mahutga, M. C., & Smith, D. A. (2011). Globalization, the structure of the world economy and economic development. Social Science Research, 40(1), 257-272.
Mantegna, R. N. (1999). Hierarchical structure in financial markets. The European Physical Journal B-Condensed Matter and Complex Systems, 11(1), 193-197.
Newman, M. E. (2003). The structure and function of complex networks. SIAM review, 45(2), 167-256.
Newman, M. E. (2008). The mathematics of networks. The new palgrave encyclopedia of economics, 2, 1-12.
Onnela, J. P., Chakraborti, A., Kaski, K., & Kertiész, J. (2002). Dynamic asset trees and portfolio analysis. The European Physical Journal B - Condensed Matter and Complex Systems, 30(3), 285-288.
Pan, R. K., & Sinha, S. (2007). Collective behavior of stock price movements in an emerging market. Physical Review E, 76(4), 046116.
Peron, T. K. D. M., da Fontoura Costa, L., & Rodrigues, F. A. (2012). The structure and resilience of financial market networks. Chaos: An Interdisciplinary Journal of Nonlinear Science, 22(1), 013117.
Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., & White, D. R. (2009). Economic networks: The new challenges. Science, 325(5939), 422.
Sensoy, A., Yuksel, S., & Erturk, M. (2013). Analysis of cross-correlations between financial markets after the 2008 crisis. Physica A: Statistical Mechanics and its Applications, 392(20), 5027-5045.
Tse, C. K., Liu, J., & Lau, F. (2010). A network perspective of the stock market. Journal of Empirical Finance, 17(4), 659-667.
Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. Proceedings of the National Academy of Sciences of the United States of America, 102(30), 10421-10426.
Tumminello, M., Di Matteo, T., Aste, T., & Mantegna, R. N. (2007). Correlation based networks of equity returns sampled at different time horizons. The European Physical Journal B, 55(2), 209-217.
Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and Applications. Cambridge University Press.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. Nature, 393(6684), 440-442.

Appendix

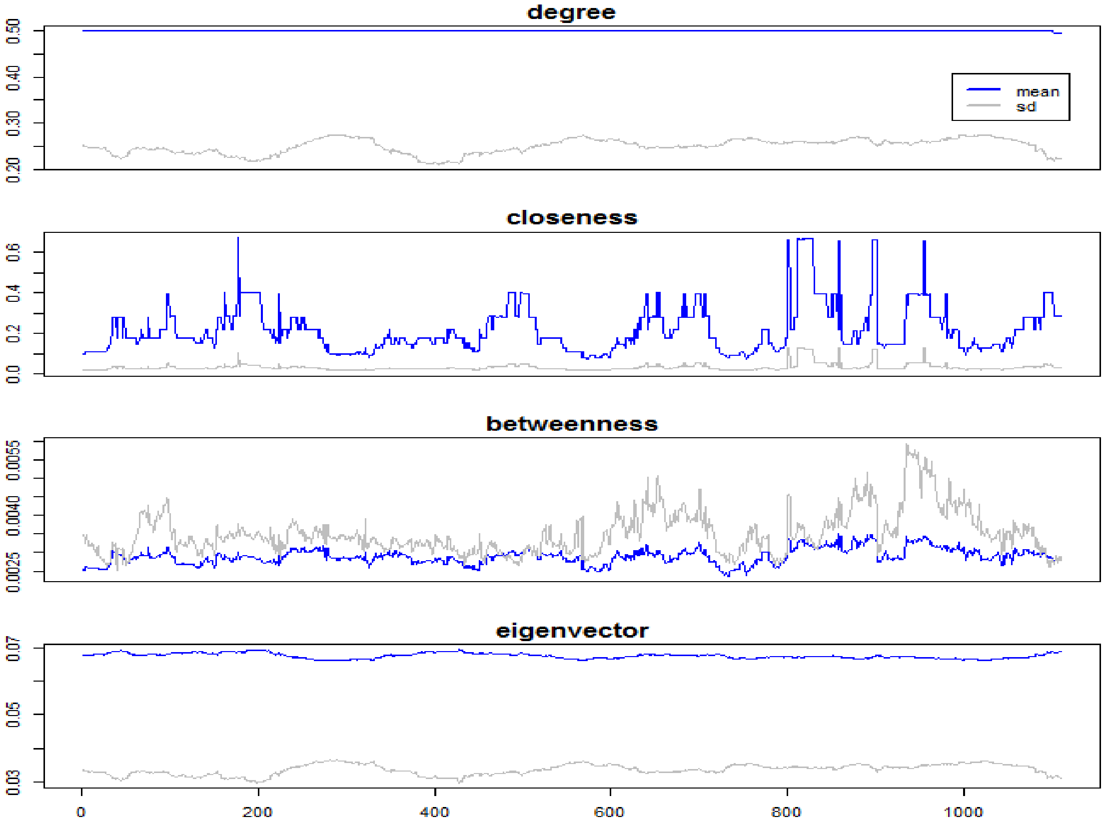

Figure A-1 Temporal variation of network centralities (mean and sd)

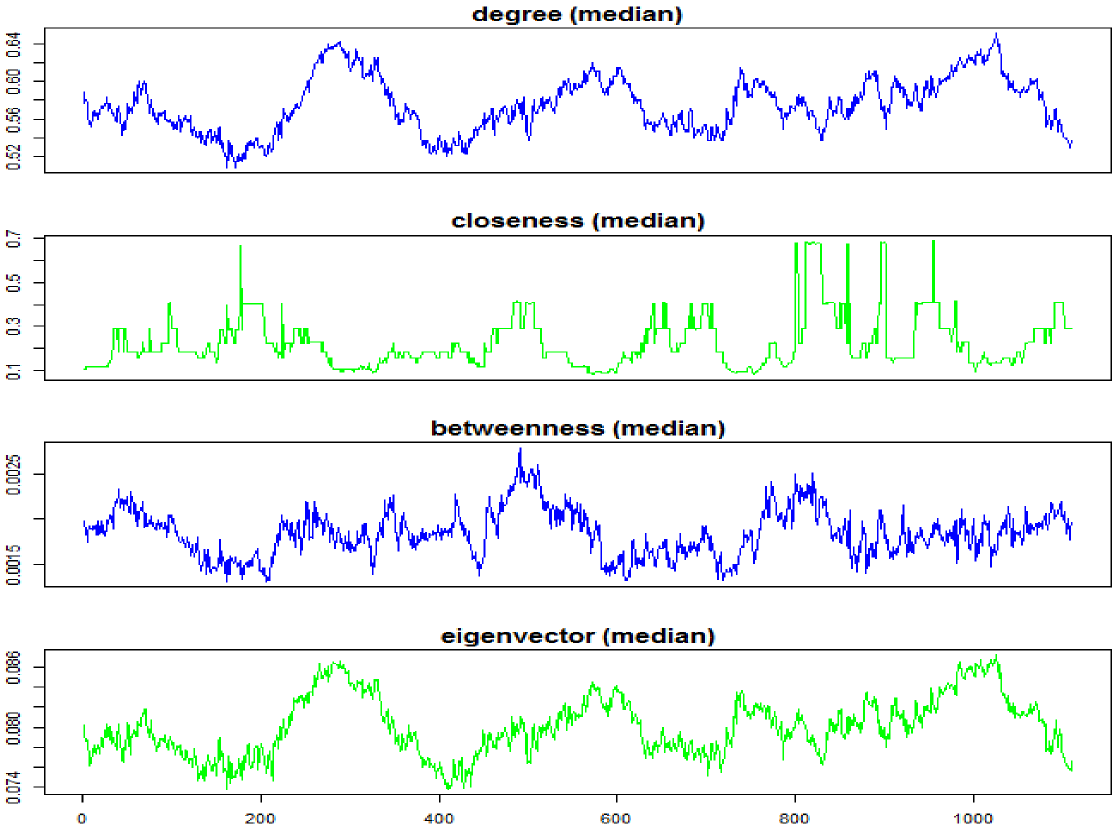

Figure A-2 Temporal variation of network centralities (median)

Table A-1 The stocks with the top 10 largest centrality

Panel A: Degree Centrality

Code	Code2	MV_Rank	Degree_mean	Rank	N_Top10	N_Top20	N_Top30
601899	ZJKY	49	0.7813	1	681	830	882
601398	GSYH	2	0.7806	2	560	813	974
600886	GTDL	66	0.7704	3	351	697	888
600664	HYGF	152	0.7577	4	332	596	753
600497	CHXZ	124	0.7366	5	208	426	571
600875	DFDQ	102	0.7349	6	289	380	466
600804	PBS	96	0.7265	7	77	362	648
600997	KLGF	176	0.7262	8	89	388	679
600739	LNCD	101	0.7177	9	289	385	489
600649	CTKG	89	0.7106	10	106	224	538

Panel B: Closeness Centrality

Code	Code2	MV_Rank	Close_mean	Rank	N_Top10	N_Top20	N_Top30
601899	ZJKY	49	0.2442	1	678	831	882
601398	GSYH	2	0.2429	2	550	813	976
600886	GTDL	66	0.2425	3	332	707	897
600664	HYGF	152	0.2416	4	336	596	755
600997	KLGF	176	0.2407	5	92	373	682
600739	LNCD	101	0.2400	6	296	401	498
600875	DFDQ	102	0.2398	7	290	375	451
600497	CHXZ	124	0.2395	8	209	421	572
600811	DFJT	161	0.2390	9	177	405	562
000629	PGFT	119	0.2386	10	275	441	582

Panel C: Betweenness Centrality

Code	Code2	MV_Rank	Betw_mean	Rank	N_Top10	N_Top20	N_Top30
600739	LNCD	101	0.0094	1	397	655	784
601699	LAHN	112	0.0074	2	279	355	431
600123	LHKC	166	0.0071	3	262	372	499
601899	ZJKY	49	0.0061	4	120	427	715
600875	DFDQ	102	0.0059	5	215	358	486
000878	YNTY	157	0.0058	6	270	370	433
600741	HYQC	78	0.0058	7	186	271	385
601398	GSYH	2	0.0058	8	104	303	547
600150	ZGCB	62	0.0057	9	257	409	516
600118	ZGWX	85	0.0055	10	251	363	435

Panel D: Eigenvector Centrality

Code	Code2	MV_Rank	Eigen_mean	Rank	N_Top10	N_Top20	N_Top30
601398	GSYH	2	0.1017	1	590	814	974
600886	GTDL	66	0.1015	2	388	679	923
601899	ZJKY	49	0.1013	3	687	823	869
600664	HYGF	152	0.1003	4	334	617	771
600497	CHXZ	124	0.0981	5	182	424	610
600804	PBS	96	0.0973	6	68	341	646
600875	DFDQ	102	0.0973	7	295	369	460
600997	KLGF	176	0.0972	8	111	431	702
600649	CTKG	89	0.0959	9	102	299	600
600018	SGJT	16	0.0941	10	249	441	562

Table A-2 The stocks with four centralities all ranking top 30

Code	Code2	MV_Rankk	degree_meann	rank1	close_mean	rank2	betw_mean	rank3	eigen_meann	rank4
600642	SNGF	97	0.6634	25	0.2362	18	0.0053	12	0.0886	28
600739	LNCD	101	0.7177	9	0.2400	6	0.0094	1	0.0925	14
000630	TLYS	147	0.6867	14	0.2365	15	0.0044	25	0.0911	18
000651	GLDQ	23	0.6838	16	0.2385	11	0.0043	26	0.0910	19
600875	DFDQ	102	0.7349	6	0.2398	7	0.0059	5	0.0973	7
601398	GSYH	2	0.7806	2	0.2429	2	0.0058	8	0.1017	1
600497	CHXZ	124	0.7366	5	0.2395	8	0.0045	24	0.0981	5
600804	PBS	96	0.7265	7	0.2384	12	0.0043	29	0.0973	6
600664	HYGF	152	0.7577	4	0.2416	4	0.0051	15	0.1003	4
601899	ZJKY	49	0.7813	1	0.2442	1	0.0061	4	0.1013	3
600886	GTDL	66	0.7704	3	0.2425	3	0.0050	16	0.1015	2
600018	SGJT	16	0.7090	11	0.2379	14	0.0045	23	0.0941	10

* rank1 is the ranking of degree_mean in the sample; similarly for rank2, rank3 and rank4.

Table A-3 Correlations between network structural properties and market fluctuation

	LGR1	LGR55	LGR100	FR1	FR55	FR100
corr-mean	0.04613	0.04959 *	-0.10836 ***	0.05701 *	0.07218 **	-0.15526 ***
corr-median	0.0474	0.06069 **	-0.08552 ***	0.05957 **	0.07421 **	-0.14480 ***
corr-sd	-0.00148	-0.21325 ***	-0.29530 ***	-0.00062	0.26809 ***	0.47333 ***
skewness	-0.04636	-0.08310 ***	-0.03667	-0.06252 **	-0.16684 ***	-0.16270 ***
kurtosis	-0.03524	0.22247 ***	0.51259 ***	-0.04835	-0.34984 ***	-0.41106 ***
clustercoef.	0.02078	-0.20550 ***	-0.43231 ***	0.03253	0.33686 ***	0.41933 ***
apl	0.04016	-0.13653 ***	-0.20374 ***	0.03239	0.33895 ***	0.28395 ***
degree-median	0.01221	-0.11294 ***	-0.28037 ***	0.0192	0.30297 ***	0.37773 ***
degree-sd	0.02596	-0.14333 ***	-0.40284 ***	0.03919	0.28411 ***	0.33648 ***
close-mean	-0.00153	-0.09828 ***	-0.01054	-0.01847	0.00065	-0.02572
close-median	-0.00169	-0.09874 ***	-0.0115	-0.01869	0.00367	-0.02065
close-sd	-0.00254	-0.05411 *	-0.03368	-0.02003	0.05728 *	0.05495 *
betw-mean	0.02613	-0.14138 ***	-0.14048 ***	0.01523	0.25729 ***	0.18329 ***
betw-median	-0.02853	-0.05148 *	0.05674 *	-0.00683	0.10607 ***	0.19428 ***
betw-sd	0.0392	-0.11840 ***	-0.05562 *	0.04579	0.12978 ***	0.08220 ***
eigen-mean	-0.01696	0.20677 ***	0.43561 ***	-0.03031	-0.32634 ***	-0.40953 ***
eigen-median	0.00471	-0.21498 ***	-0.35926 ***	0.01531	0.28684 ***	0.40037 ***
eigen-sd	0.01636	-0.20291 ***	-0.43171 ***	0.03045	0.32814 ***	0.40957 ***

significant level: *** 0.01; ** 0.05; * 0.10; otherwise, not significant.