EVOLUTION OF SHANGHAI STOCK MARKET BASED ON MAXIMAL SPANNING TREES

CHUNXIA YANG*, YING SHEN and BINGYING XIA

Nanjing University of Information Science and Technology, Nanjing 210044, China
*y.cx@163.com

Received 15 May 2012
Revised 31 October 2012
Accepted 15 November 2012
Published 14 December 2012

In this paper, using a moving window to scan through every stock price time series over a period from 2 January 2001 to 11 March 2011 and mutual information to measure the statistical interdependence between stock prices, we construct a corresponding weighted network for 501 Shanghai stocks in every given window. Next, we extract its maximal spanning tree and understand the structure variation of Shanghai stock market by analyzing the average path length, the influence of the center node and the p-value for every maximal spanning tree. A further analysis of the structure properties of maximal spanning trees over different periods of Shanghai stock market is carried out. All the obtained results indicate that the periods around 8 August 2005, 17 October 2007 and 25 December 2008 are turning points of Shanghai stock market, at turning points, the topology structure of the maximal spanning tree changes obviously: the degree of separation between nodes increases; the structure becomes looser; the influence of the center node gets smaller, and the degree distribution of the maximal spanning tree is no longer a power-law distribution. Lastly, we give an analysis of the variations of the single-step and multi-step survival ratios for all maximal spanning trees and find that two stocks are closely bonded and hard to be broken in a short term, on the contrary, no pair of stocks remains closely bonded for a long time.

Keywords: Mutual information; stock network; maximal spanning tree.

1. Introduction

The study on the behavior of the Chinese stock market has become a hot topic. Su et al. found that its volatility is time-varying and government’s market intervention policies have affected it. Yan et al. objectively identified Chinese bull or bear market regimes and gave a clear outline of their statistical properties. Yao et al. claimed that the Chinese stock markets were extremely volatile during the period 2005–2008 and could become stable without extreme volatility only after Chinese companies become really commercialized and profitable and investors become rational. Generally, stock market’s behavior always associated with the complicated
relationships among stocks. In order to further reveal market’s behavior, many people investigated stock markets by establishing various stock networks, where a node denotes a stock and an edge represents a relationship between any two stocks.\textsuperscript{7–23} Huang \textit{et al.} showed a hierarchical structure of Shanghai 50 stocks based on minimal spanning tree and found that industry’s style is very obvious.\textsuperscript{20} Zhang \textit{et al.} analyzed stocks’ classification for those ones in Shanghai–Shenzhen 300 stock index and claimed that such classification can provide a useful reference to portfolio management.\textsuperscript{22} It is worth noting that most stock networks mentioned in above works are constructed by applying Pearson correlation function of a pair of financial time series to define a kind of distance between any two stocks. Such stock networks surely reveal a kind of linear physical relationships between stocks. However, there exist other relationships between stocks, which motivates us to find other methodologies to analyze stocks’ relationships. In fact, there exist highly nonlinear processes in a real stock market, which means that new methodologies are greatly needed to study stock market. So we introduce mutual information to measure the statistical interdependence between stock prices and construct undirected stock networks. In our previous study based on such undirected stock networks, we had found that the scalefreeness of the degree distribution of stock network is disrupted at turning points of the Shanghai stock market.\textsuperscript{24} Although such results are meaningful, they only present us a rough understanding of Shanghai stock market. In order to avoid the subjective error in our previous networks constructed using thresholding procedure and obtain more reliable information of the Shanghai stock market, we will construct weighted stock networks and use the method of maximal spanning tree to analyze the evolution of Shanghai stock market in this paper. The new obtained results can not only prove our previous conclusions but also give us more detailed understandings of the evolution characteristics of Shanghai stock market.

The paper is organized as follows. The variation of Shanghai stock index is shown in Sec. 2. In Sec. 3, we will extract the maximal spanning tree from every fully connected stock network. In Sec 4, we analyze the structure variation of Shanghai stock market based on maximal spanning trees. In Sec. 5, we calculate the single-step and multi-step survival ratios to help understand the evolution of maximal spanning trees. Section 6 gives our conclusions.

2. The Variation of Shanghai Stock Index

Shanghai stock market is a typical complex system, and the Shanghai stock index also has complex behavior. Figure 1 shows the variation of Shanghai stock index over a period from 2 January 2001 to 11 March 2011 where three lines indicate the dates that the stock index reaches the lowest or highest point. From Fig. 1, it is easy to see that the stock index fell to the lowest point around 11 July 2005, and then it rebounded and did not stop rising until it reached the top point around 16 October 2007, after that, it dropped again to another lowest point around 4 November 2008. Thus, we can make a bold assumption that such three periods are turning
points of the Shanghai stock market. Nevertheless, we still do not know what had happened to the relationships between stocks. Specially is the relationships between stocks at turning points different from those during normal periods? Answering such question needs a detailed study on the Shanghai stock market. So, we will establish corresponding stock correlation networks and analyze them to obtain a persuasive answer.

3. Stock Network

3.1. Mutual information

The mutual information $M_{ij}$ from information theory can be interpreted as the excess amount of information generated by falsely assuming two time series $a_i$ and $a_j$ to be independent. It offers new promising perspectives for uncovering strongly nonlinear relationships between time series.$^{25-27}$ By definition, $M_{ij}$ is large if the two time series are highly linearly (anti)correlated. Besides, a highly nonlinear relationship between $a_i$ and $a_j$ also yields large $M_{ij}$. The mutual information $M_{ij}$ between $a_i$ and $a_j$ can be calculated using:

$$M_{ij} = \sum_{\mu, \nu} p_{ij}(\mu, \nu) \log_2 \frac{p_{ij}(\mu, \nu)}{p_i(\mu)p_j(\nu)}.$$  \hspace{1cm} (1)

Here $p_i(\mu)$ is the probability density function (PDF) of the time series $a_i$, and $p_{ij}(\mu, \nu)$ is the joint PDF of a pair $(a_i, a_j)$. By definition, $M_{ij}$ is symmetric, so $M_{ij} = M_{ji}$.\hspace{1cm} (28) If logarithms to base 2 are used, the standard unit of measurement of mutual information is the bit. The principal difficulty in calculating mutual information from experimental data is how to estimate $p_{ij}(\mu, \nu)$. We use a simple histogram approach with equally sized boxes for all pairs $\{i, j\}$ to estimate the probability densities. Because the estimator [Eq. (1)] is known to depend on the
box size and partitioning, we use an identical partitioning for all pairs \( \{i,j\} \) to guarantee optimal comparability of the \( M_{ij} \). The number \( m \) of boxes is determined using a formula from:

\[
m = 1.87 \times (l - 1)^{2/5},
\]

where \( l \) is the length of time series. If a box with size \( \Delta i \Delta j \) in the \( \{i,j\} \) plane has \( N_{ij} \) points in it, we estimate \( p_{ij} \) to be \( N_{ij}/N_{\text{total}} \), where \( N_{\text{total}} \) is the total number of points in the plane. In this way, \( p_{ij}(\mu, \nu) \) and \( M_{ij} \) can be calculated. The algorithm is feasible, since the application to the network construction requires only the correct estimation of relative differences of \( M_{ij} \) between all pairs of time series.

### 3.2. Maximal spanning trees

The maximal spanning tree, a theoretical concept in graph theory has been used for filtering networks, resulting in simpler forms of graphs that can facilitate analysis. Spanning tree is a particular type of graphs without forming any loop, so a maximal spanning tree connects all the \( N \) nodes of a network with \( N - 1 \) links, so that the sum of the link weights is maximized.

### 3.3. Network construction

We use mutual information to measure the statistical interdependence between stock prices and construct a stock network, which is a weighed and fully connected one, and the mutual information represents link weights. Here, both the dependence and anti-dependence are taken into account. Such method is similar to the absolute value of the linear correlation coefficient, which takes the correlation and anti-correlation into account, when applying cross-correlation coefficient to analyze stock networks. When two stocks are independent, their mutual information is zero; when they are interdependent, the mutual information is positive. Besides, the stock market is always described as an evolving complex system, and its structure will change as time goes by. So we employ a moving window method to exploit the structure variation of stock networks. A window with 400 days width is determined according to the nonlinear behavior of stock prices, and it will slide 1 day each step. For example, we put the first window on the first 400 days of the whole series, compute mutual informations for each pair of stocks to represent link weights, and construct a weighted and fully connected stock network within the first 400-day window; then, the window slides 1 day along the daily stock price time series, new mutual information for each pair of stocks is computed, and another stock network is constructed within the next 400-day window. The same procedure is repeated until we finish scanning through the whole time series. At last, 2063 weighted and fully connected networks are constructed.

\(^a\)In fact, the results obtained by using stock price are nearly the same as that by using stock return, and the latter are introduced in “Appendix A”.

---

\(^a\)In fact, the results obtained by using stock price are nearly the same as that by using stock return, and the latter are introduced in “Appendix A”.

\[^23^2\]
Next, we extract the maximal spanning tree from every weighted and fully connected network using the Prim Algorithm. The tree connects all the 501 nodes of a network with 500 links, and the sum of the link weights is maximized. The method of maximal spanning tree is used to reduce the network complexity to grasp the market essence, it can facilitate the analysis of structure variation of the Shanghai stock market.

3.4. Topology structure of the maximal spanning tree

3.4.1. Average path length

Average path length is defined as the average of the shortest path between any two different nodes in a maximal spanning tree,

$$L = \frac{2}{N(N-1)} \sum_{i>j} d_{ij}.$$  \hspace{1cm} (3)

Here $d_{ij}$ denotes the shortest path between node $i$ and node $j$, $N$ presents the number of nodes. $L$ indicates the degree of separation between nodes.

3.4.2. The influence of the center node

We define a node who has the maximum degree is a center node in a tree. For a node, the degree shows the direct connection to other nodes, and is a reflection of its influence on other nodes. The larger degree a node has, the larger influence it gets.

3.4.3. P-value

The statistical property of degree distribution of every maximal spanning tree will be discussed by the maximum likelihood method and the Kolmogorov–Smirnov (KS) statistic proposed by Clauset et al. First, we can get the reliable estimates of $x_{\text{min}}$ and the power-law exponent by minimizing the standard KS statistic,

$$D = \max_{x \geq x_{\text{min}}} |S(x) - P(x)|.$$  \hspace{1cm} (4)

Here, $S(x)$ is the cumulative distribution function (CDF) of the data for observations with value at least $x_{\text{min}}$, $P(x)$ is the CDF for the power-law model that best fits the data in the region $x \geq x_{\text{min}}$, and $D$ is the maximum distance between the CDF of the data and the fitted model. Next, in order to rule out the non-power-law distribution, a $p$-value which quantifies a probability that the empirical data were drawn from the hypothesized power-law distribution should be computed. If the $p$-value is much less than 1, it is unlikely that the data are drawn from a power-law distribution. If it is close to 1, then the data maybe are drawn from a power-law distribution.
4. Structure Variation of the Shanghai Stock Market

4.1. Variation of statistical properties for maximal spanning tree

Figure 2 shows the variation of the average path length, where the marked time is the intermediate time of considered window. For example, a 400-day window is from 2 January 2001 to 5 September 2002, the time marked on the x axis is 5 November 2001 which is its intermediate time. Here, we do a slight adjustment when computing the average path length. As mentioned above, every maximal spanning tree has a maximal sum of link weights. As shown in Eq. (3), the distance \( d_{ij} \) between any two different nodes is affected by two impacts, one is the topology structure property and the other is link weights. Then, in order to analyze the property of tree’s topology only, we change all the link weights which are more than 0 into 1 so as to eliminate the effect by link weights. From Fig. 2, we can see that the average path length reaches three top peaks around 8 August 2005, 17 October 2007 and 25 December 2008, respectively. That is to say, over such three periods, the degree of separation between nodes increases, and the topology structure of the maximal spanning tree becomes looser.

Figure 3 shows the degree of the center node, the marked time is the same as that in Fig. 2. From Fig. 3, we can see that the influence of the center node is at a low level around 8 August 2005, 17 October 2007 and 25 December 2008. So, we can say that over such three periods, the influence of the center node gets smaller.

Figure 4 shows the \( p \)-value, the marked time is the same as that in Fig. 2. From Fig. 4, one can observe that the \( p \)-values for our distributions are above a threshold 0.05 over other periods, while they fall below the threshold over periods around 8 August 2005, 17 October 2007 and 25 December 2008, which indicates that the
Evolution of Shanghai Stock Market Based on Maximal Spanning Trees

Fig. 3. (Color online) The degree of the center node as a function of time. (The three red lines indicate 8 August 2005, 17 October 2007 and 25 December 2008, respectively.)

Fig. 4. (Color online) The p-value as a function of time. (The three red lines indicate 8 August 2005, 17 October 2007 and 25 December 2008, respectively.)

A power-law model is a good fit for the degree distributions of the maximal spanning trees over periods except for those around 8 August 2005, 17 October 2007 and 25 December 2008. Furthermore, we can say that the scalefreeness of the degree distribution of the maximal spanning tree around such three periods is disrupted.

As shown in Figs. 2–4, we can draw a preliminary conclusion that the periods around 8 August 2005, 17 October 2007 and 25 December 2008 are turning points of the Shanghai stock market, which proves the assumption in Sec. 2 and also is consistent with our previous result. At turning points, the topology structure of
the maximal spanning tree becomes looser, the influence of the center node gets smaller and the scalefreeness of the degree distribution is disrupted. Besides, there are many influential domestic or foreign events happening in such periods. Chinese government tightened macro-control and executed the share-trading reform from the year of 2005 to 2007, which led the Shanghai stock index to fluctuate rapidly.\textsuperscript{37,38} The Shanghai stock index fell to the lowest point around August 2005, and broke through 6000 point to reach the highest point around October 2007. After that, affected by the global financial crisis, it dropped to another lowest point around December 2008. In fact, such events not only affect the behavior of the index but also influence the relationships between stocks. Furthermore, the relationships between stocks at turning points are different from those during normal periods.

4.2. Variation of maximal spanning tree over different periods of Shanghai stock market

In order to know more details of Shanghai stock market, we pay a special attention to the comparison of the maximal spanning tree over a normal period with that at turning points. A normal period means the time when the Shanghai stock market does not change radically. Here, a period from 21 May 2002 to 3 January 2004 whose intermediate time is 17 March 2003 and another period from 3 January 2003 to 31 October 2004 whose intermediate time is 6 November 2003 are two normal periods. The corresponding maximal spanning tree around 17 March 2003 is shown in Fig. 5 (left), where a node represents a stock of Shanghai stock market. In order to show the topology clearly, we do not mark the node labels. The right of Fig. 5 shows the tree’s degree distribution. We can see that the topology structure of the maximal spanning tree is star-like form (left) and the tree’s degree distribution is a power-law distribution, where the dotted line is the fit curve. Furthermore, we calculate the KS goodness-of-fit statistic using the method proposed by Clauset

![Fig. 5. (Color online) Maximal spanning tree over a normal period around 17 March 2003 (left) and its degree distribution (right).](image)
et al. and get the estimates of $x_{\min}$ and power-law exponent for the degree distribution in Fig. 5 (right). They are 1 and 2.37 respectively, and $D = 0.0198$. The $p$-value of the maximal spanning tree is 0.23, more than 0.05. Similarly, Fig. 6 (left) shows the maximal spanning tree around 6 November 2003 and its corresponding degree distribution (right). One can find that the tree’s topology structure is also star-like. The corresponding obtained values are as follows: $x_{\min} = 1$, power-law exponent $\gamma = 2.29$, $D = 0.0387$, and $p$-value is 0.63. From such two figures, we can believe that the degree distribution of the maximal spanning tree in such normal period is a power-law distribution. Figure 7 (left) shows the maximal spanning tree at turning point around 8 August 2005, from which we can see that the topology structure of the maximal spanning tree turns to be chain-like. The right of Fig. 7 shows its degree distribution. Here, $x_{\min} = 1$, power-law exponent $\gamma = 2.27$, $D = 0.0482$. Although $D$ is not large, the $p$-value is just only 0.04, so we can
conclude that the degree distribution of the maximal spanning tree around 8 August 2005 is no longer a power-law distribution. Figure 8 shows the maximal spanning tree at the turning point around 17 October 2007 and its degree distribution. $x_{\text{min}}$, power-law exponent, $D$ and $p$-value of the degree distribution are 2, 2.58, 0.0659 and 0, respectively. Figure 9 shows the maximal spanning tree at the turning point around 25 December 2008 and its degree distribution. $x_{\text{min}}$, power-law exponent, $D$ and $p$-value of the degree distribution are 1, 2.17, 0.1469 and 0, respectively. From Figs. 8 and 9, one can find that both of their topology structures are chain-like and their maximal spanning trees are not scalefreeness too.

Therefore, from Figs. 5–9, we can draw three conclusions: First, during the period from 2 January 2001 to 11 March 2011, the topology structure of the maximal spanning tree changes from star-like to chain-like around the first turning point,
then it turns to be star-like, while at the second turning point, it turns to be chain-like once again, after that it is back to star-like, and at the third turning point, it turns to be chain-like again. Second, the degree distribution of the maximal spanning tree is no longer a power-law distribution during the three turning points. Third, the maximal spanning tree shows the clustering of stocks where a link indicates that two connected nodes are closely contact, from which people can reduce risk by unselecting stocks which are closely contact with the selected ones. So it can be a good guide to the risk management of stock investment.

4.3. **Comparison between maximal spanning tree and threshold value method**

There exist differences between maximal spanning tree and threshold value method. By the latter, stock networks are constructed based on the judgment of the investigator who retains the links with weights above a certain threshold. However, maximal spanning tree method retains those edges that fit the maximal spanning tree criterion. So the method of maximal spanning tree can avoid the subjective error happening in threshold value method. Besides the results obtained by the two methods are different. Although the same conclusion that the degree distribution of stock network at turning points is no longer a power-law distribution is obtained, we still do not know the details of variation of stock network topology by the latter method. Therefore, we use the maximal spanning tree method to give a further analysis of the structure variation of Shanghai stock market, particularly at turning points. Not only do we prove the previous conclusion that the degree distribution of the network at turning point is no longer a power-law distribution, but also find that the topology structure becomes looser and changes from star-like to chain-like. The difference also happened to the $p$-values. Here, the $p$-values at the turning points are smaller than 0.05, which are consistent with the previous results. However, the $p$-value of the maximal spanning tree over a normal period is larger than that of networks constructed by the latter method, which means that the probability for the degree distribution of the maximal spanning tree to be a power-law distribution is larger than that of network constructed by the latter method.

5. **Other Variations of Maximal Spanning Trees**

5.1. **Single-step survival ratio**

In order to investigate the robustness of the maximal spanning tree, we calculate the single-step survival ratio, which is a fraction of edges in common found in two adjacent trees at times $t$ and $t - 1$, and the ratio is defined as follows:

$$\sigma(t) = \frac{1}{N-1} |E(t) \cap E(t-1)| .$$  \hspace{1cm} (5)

Here $E(t)$ refers to the cardinality of the set at time $t$, $|E(t) \cap E(t-1)|$ presents the number of common edges presenting in the intersection between two adjacent
trees at times $t$ and $t-1$. Figure 10 shows the single-step survival ratio. The marked time is the same as that in Fig. 2. Obviously, we can see that the single-step survival ratio is always very high, and the average is about 0.9, which means that a large majority of connections survives from one time window to the next. In other words, we can say that two stocks are closely bonded and hard to be broken in a short term. Although the single-step survival ratio is always high during the whole period, there are still a few variations between stocks from one maximal spanning tree to the next. People can understand the relationships between stocks that they are concerning by analyzing such changes, then determine whether to adjust their investment strategy or not.

5.2. Multi-step survival ratio

In order to further study the long term evolution of the maximal spanning tree, we introduce the multi-step survival ration at time $t$ as follows:\[17\]

$$
\sigma(t,k) = \frac{1}{N-1} \left| E(t) \cap E(t-1) \cap \cdots \cap E(t-k+1) \cap E(t-k) \right|.
$$

(6)

Fig. 10. (Color online) The single-step survival ratio as a function of time.
Evolution of Shanghai Stock Market Based on Maximal Spanning Trees

Fig. 11. (Color online) The multi-step survival ratio as a function of time.

Fig. 12. (Color online) The mean multi-step survival ratio as a function of time.

of $k$ as follows:

$$\sigma(k) = \frac{1}{M-k}[\sigma(k+1,k) + \sigma(k+2,k) + \cdots + \sigma(M-1,k) + \sigma(M,k)].$$  \hspace{1cm} (7)

Here

$$M = 2063, \sigma(M,k) = \frac{1}{N-1}|E(M) \cap E(M-1) \cap \cdots \cap E(M-k+1) \cap E(M-k)|.$$

When $k=1$, we obtain the mean of 2062 multi-step survival ratios, and $\sigma(1) = \frac{1}{2062}[\sigma(2,1) + \sigma(3,1) + \cdots + \sigma(2063,1)]$. Similarly, when $k=320$, $\sigma(320) = \frac{1}{1743}[\sigma(321,320) + \sigma(322,320) + \cdots + \sigma(2063,320)]$. Figure 12 shows the mean multi-step survival ratio. From Fig. 12, one can observe that when $k=60$, the mean multi-step survival ratio is nearly 0.1, which means that only 10% of the total edges are in common in trees of any neighborhood 61 windows (about three months), while most of the edges change. Such result gives us some significant
guidance. Investors could adjust their investment portfolio every three months by analyzing the variation of stocks’ clustering based on maximal spanning trees.

6. Conclusion

Currently, further study on the behavior of stock market by establishing various stock networks has received considerable attention. In this paper, we have investigated the Shanghai stock market by analyzing its corresponding stock network’s topological changes. First, we use a moving window to scan through every stock price time series over a period from 2 January 2001 to 11 March 2011 and mutual information to measure the statistical interdependence between stock prices, and construct a corresponding weighted network of 501 Shanghai stocks for every given window. Second, we extract its maximal spanning tree and understand the variations of statistical properties by analyzing the average path length, the influence of the center node and the p-value for every maximal spanning tree. Then, we give detailed analyses of the variation of maximal spanning trees at different periods of Shanghai stock market. The obtained results are as follows: the periods around 8 August 2005, 17 October 2007 and 25 December 2008 are turning points; at turning points, the degree of separation between nodes increases, the structure becomes looser, the influence of the center node gets smaller, the maximal spanning tree’s topology structure is more chain-like and its degree distribution is no longer a power-law distribution. All of such results not only enrich our previous study but also give us a more persuasive answer than that just obtained by analyzing the variation of the Shanghai stock index. Third, we give a comparison between maximal spanning tree method and threshold value method, and find that the former method is better for network analysis because it avoids the subjective error and gives a deeper understanding of Shanghai stock market. At last, we give an analysis of the variations of the single-step and multi-step survival ratios, from which we draw a conclusion that two stocks are closely bonded and hard to be broken in a short term, on the contrary, no pair of stocks remains closely bonded after a long time. All the above studies could help us learn more about the Shanghai stock market characteristics, meanwhile motivate us to mine its underlying economic meaning.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant Nos. 60874111, 61273229), Qing Lan Project of Jiangsu Province, College Science Foundation of Jiangsu Province (07KJD120128) and Social Science Institute Program of Jiangsu Province (B0808).

Appendix A

From price time series, we obtain log-return series given by

\[ S_t = \ln(P_t) - \ln(P_{t-1}) \]  \hspace{1cm} (A.1)
Evolution of Shanghai Stock Market Based on Maximal Spanning Trees

Fig. A.1. (Color online) Return’s nonlinear behavior for a stock labeled SH600007.

Here $P_t$ is the price of a stock at day $t$. Then, we analyze the nonlinear behavior of stock returns. Palus introduces a method to detect the nonlinearity for a given time series $x(t)$, which is usually considered as a realization of a stochastic process $X(t)$.\textsuperscript{39,40} First, two stochastic variables $X(t)$ and $X(t + \tau)$ are constructed. Then mutual information $M(\tau)$ and another value $L(\tau)$ are computed,

$$L(\tau) = \frac{1}{2} \left( \log_a C_{11} + \log_a C_{22} \right) - \frac{1}{2} \left( \log_a \sigma_1 + \log_a \sigma_2 \right).$$  \hspace{1cm} (A.2)

Here $\sigma_1$ and $\sigma_2$ are eigenvalues of the covariance matrix of the above two-dimensional random variables $X(t)$ and $X(t + \tau)$. $C_{11}$ and $C_{22}$ are the diagonal elements of the covariance matrix. When $M(\tau)$ and $L(\tau)$ change similarly, this can reflect a linear correlation between $X(t)$ and $X(t + \tau)$. In contrast, when the variable patterns of $M(\tau)$ and $L(\tau)$ are significantly different, $X(t)$ and $X(t + \tau)$ are nonlinearly related. By this method, we detect the nonlinear behavior of the return for a stock labeled SH600007 over a period from 2 January 2001 to 11 March 2011. Figure A.1 shows the corresponding results. It shows that the variable pattern of $M(\tau)$ and $L(\tau)$ is obviously different when the delay time $\tau$ is small ($\tau \approx 590$). Figure A.2 shows return’s nonlinear behavior for two stocks labeled SH600007 and SH600008 over the same period. Similarly, one can find that the variable tendency of $M(\tau)$ and $L(\tau)$ is obviously different when the delay time $\tau$ is small ($\tau \approx 590$). Thus such two experiments indicate that there surely exist nonlinear processes in a real stock market, but also the length of the nonlinear correlation is approximately 590 days, so the size of moving window is 590 days.
We also use mutual information to measure the statistical interdependence between stock returns and employ a moving window with the size of 590 days which can slide 1 day each step to scan through every stock return time series. Then, we construct every weighted and fully connected networks and extract each maximal spanning tree from them. We also calculate the average path length, the influence of the center node, and $p$-value for every maximal spanning tree to analyze the structure variation of Shanghai stock market. Figure A.3 shows the variation of the average path length. Similarly we change all the link weights which are more than 0 into 1 so as to analyze the property of tree’s topology only. The influence
Evolution of Shanghai Stock Market Based on Maximal Spanning Trees

![Degree of the center node over time](image1)

**Fig. A.4.** (Color online) The degree of the center node as a function of time. (The three red lines indicate 16 August 2005, 18 October 2007 and 22 December 2008, respectively.)

![P-value over time](image2)

**Fig. A.5.** (Color online) The p-value as a function of time. (The three red lines indicate 16 August 2005, 18 October 2007 and 22 December 2008, respectively.)

of the center node and the p-value are shown in Figs. A.4 and A.5, respectively. In Figs. A.3–A.5, the marked time is the intermediate time of the considered window. Obviously, one can find that the average path length reaches three peaks around 16 August 2005, 18 October 2007 and 22 December 2008, which indicates that the topology structure turns to be looser around such three periods than other periods. Besides, as shown in Fig. A.4, the center node has a lower degree around such three periods. The p-value in Fig. A.5 shows that the power-law model is a good fit for the degree distributions of the maximal spanning trees over periods except for those around 16 August 2005, 18 October 2007 and 22 December 2008. As shown in Figs. A.3–A.5, we can conclude that the periods around 16 August 2005, 18 October 2007 and 22 December 2008 are turning points of the Shanghai stock market; at turning points, the topology structure of the maximal spanning tree...
C.-X. Yang, Y. Shen & B.-Y. Xia

becomes looser, the influence of the center node gets smaller and the probability for the degree distribution to be a power-law distribution is lower. Such conclusions obtained by using stock return are nearly the same as that by using stock price.

References