Physica A 388 (2009) 2956–2964

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A network analysis of the Chinese stock market

Wei-Qiang Huang *, Xin-Tian Zhuang, Shuang Yao

School of Business Administration, Northeastern University, Shenyang, Liaoning 110004, China

A R T I C L E I N F O

Article history:
Received 7 December 2008
Available online 26 March 2009

Keywords:
Stock correlation network
Stock price fluctuations
Complex network theory
Topology structure
Scale-free degree distribution
Clustering coefficient
Component
Clique
Independent set
Topological stability
Portfolio investment

A B S T R A C T

In many practical important cases, a massive dataset can be represented as a very large network with certain attributes associated with its vertices and edges. Stock markets generate huge amounts of data, which can be use for constructing the network reflecting the market's behavior. In this paper, we use a threshold method to construct China's stock correlation network and then study the network's structural properties and topological stability. We conduct a statistical analysis of this network and show that it follows a power-law model. We also detect components, cliques and independent sets in this network. These analyses allows one to apply a new data mining technique of classifying financial instruments based on stock price data, which provides a deeper insight into the internal structure of the stock market. Moreover, we test the topological stability of this network and find that it displays a topological robustness against random vertex failures, but it is also fragile to intentional attacks. Such a network stability property would be also useful for portfolio investment and risk management.

1. Introduction

In nature, a large number of complex systems can be described by complex networks [1–3].

In the network, the vertices are elements in the system, and two vertices are connected by an edge. After the innovative researches of small-world networks by Watts and Strogatz [4] and a scale-free network by Barabasi and Albert [5], people do extensive empirical researches on topology characteristics of actual networks from different domains, such as the WWW, the Internet, the movie actor collaboration networks, cell networks and telephone networks etc. These networks have a common topology character of being small-world and scale-free [2]. Traditional network models such as regular and stochastic networks cannot embody these characters. In recent years, many network modes are established such as small-world models [4,6,7] and dynamic evolvement scale-free models [5,8,9].

In essence, the stock market is a complex system [10]. The price fluctuations among vast stocks have complicated relationships. Lately, people have investigated stock markets by establishing the corresponding stock correlation networks, of which the vertices are stocks and edges between vertices are price fluctuation relationships of stocks [11–16]. From a different angle of view, people put forward much network construction arithmetic such as the minimum-cost spanning tree(MST) [11–13], the planar maximal filtering graph (PMFG) [14,15] and the correlation threshold method [16]. The research contents include basic topology characteristics and intrinsic hierarchical structures of stock correlation networks etc. Mantegna was the first [11] to construct networks based on stock price correlations. He investigated the portfolio of the stocks used to compute the Dow Jones Industrial Average (DJIA) index and the portfolio of stocks used to compute the Standard and Poor's 500 (S& P 500) index in the time period from July 1989 to October 1995. He found a hierarchical arrangement of stocks in the stock correlation network which was built by investigating the daily time series of the logarithm of the stock price. Onnela et al. [13] constructed a network called the asset graph according to the rank of 477 stock's

* Corresponding author.
E-mail address: hwqneu@hotmail.com (W.-Q. Huang).

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2009.03.028

correlation strength in the New York stock market. They studied its properties, such as topologically different growth types, number and size of clusters and the clustering coefficient. These properties, calculated from empirical data, are compared against those of a random graph. Tumminello et al. [15] investigated the planar maximally filtered graphs of the portfolio of the 300 most capitalized stocks traded at the New York Stock Exchange during the time period 2001–2003. Topological properties such as the average length of shortest paths, the betweenness and the degree were computed on different planar maximally filtered graphs generated by sampling the returns at different time horizons. Boginski et al. [16] constructed a stock market graph by analyzing daily fluctuations of 6546 financial instruments in the US stock markets during 500 consecutive trading days in 2000–2002. They conducted the statistical analysis of this graph and detected cliques and independent sets in this graph.

The topological stability of a network against random failures and attacks is a very important research aspect of a complex network. The failures or attacks refer to removal of vertices or edges. The research results indicate that regular and random networks display comparative effects between error and attack tolerance. A small-world network is sensitive to long distance attacks [17]. Scale-free networks display a topological robustness against random node failures, but they are also fragile to intentional attacks [18]. The listed companies may have some extreme risks such as delisting or bankruptcy, to which corresponds removals of vertices in the stock correlation network. The removal of certain stocks will lead to topological changes of the holistic stock correlation network.

The study on such topological changes can help us understand correlation patterns among stocks, thus it can be a good guide for risk management of stock investment. Nowadays, a great number of different stocks are traded in the Shanghai and Shenzhen stock market in China. According to trading data of these stocks, we construct a corresponding stock correlation network with a correlation threshold method. We present a detailed study of the properties of this network, including the topology structures and topological stability against random failures and attacks. Different from other network construction arithmetic, the correlation threshold is convenient for studying inherent relationships between network characteristics and correlation threshold. The rest of this article is organized as follows. In Section 2 we describe the process of network construction and topology statistical quantities of the network. Section 3 is the empirical study and results. In the last section we present a few conclusions.

2. Construction and structure of the stock correlation network

2.1. Constructing the network

Let $P_i(\tau)$ be the stock-price of a company $i(i = 1, \ldots, N)$ at time τ. Then the return of the stock-price at a time interval Δt is defined as

$$r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - \Delta t) \tag{1}$$

meaning the geometrical change of $P_i(\tau)$ during the interval Δt. We take Δt as one day in the following analysis throughout this paper. The cross-correlations between individual stocks are considered in terms of the matrix $\mathbf{C}$, whose elements are given by using the following formula

$$c_{i,j} \equiv \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{(\langle r_i^2 \rangle - \langle r_i \rangle^2)(\langle r_j^2 \rangle - \langle r_j \rangle^2)}} \tag{2}$$

where the brackets mean a temporal average over the period we studied. Then $c_{i,j}$ can vary between $[-1, 1]$. The case of $c_{i,j} = 1(-1)$ means that two companies i and j are completely correlated (anti-correlated), while $c_{i,j} = 0$ means that they are uncorrelated.

The main idea of constructing the stock correlation network is as follows. Let the set of stocks represent the set of vertices of the network. Also, we specify a certain threshold value θ, $-1 \le \theta \le 1$ and add an undirected edge connecting the vertices i and j if the correlation coefficient $c_{i,j}$ is great than or equal to θ. Obviously, different values of θ define the networks with the same set of vertices, but different sets of edges.

Let graph $G = (V, E)$ represent the stock correlation network, where V and E are the set of vertices edges respectively. E is defined as

$$E = \begin{cases} e_{ij} = 1, & i \ne j \text{ and } c_{ij} \ge \theta \\ e_{ij} = 0, & i = j. \end{cases} \tag{3}$$

2.2. Topology structure of the network

2.2.1. Degree distribution

The degree of vertex i is $k_i = \sum_{j \ne i} e_{ij}$ which denotes the vertex number connecting with i. The vertex degree distribution function P_k tells us the probability that a randomly selected vertex is connected with k edges. The results of many empirical studies on actual networks demonstrate that the vertex degree distribution obeys a power law [2]

$$P_k \propto k^{-\gamma} \tag{4}$$

where γ is the parameter. Formula (4) can be equivalently expressed as

$$\ln P_k \propto -\gamma \ln k. \tag{5}$$

The network having power law vertex degree distribution is a scale-free network. In such a network, most of the vertices have small degree and a small quantity of vertices have large degree. We call the latter as "Hub" vertices. Especially, in the stock correlation network stock i having a large degree means that it is correlated with many other stocks in the sense of price fluctuation. So we can dig out the very stocks that can most accurately reflect market behaviors by the degree of stocks.

2.2.2. Clustering coefficient

If k_i nearest neighbors of vertex i have m_i edges among them, the ratio of m_i to $k_i(k_i - 1)/2$ is the clustering coefficient of vertex i. The network clustering coefficient is calculated by averaging through the clustering coefficient of all vertices. The results of many empirical studies on actual networks demonstrate they have larger network clustering coefficients than stochastic networks with the same size [2]. The phenomenon of large clustering in actual networks motivates the appearance of small-world network models [4,6,7]. Specially, the network clustering coefficient of a stock correlation network indicates the clustering property of stocks in the meaning of price fluctuation correlation.

2.2.3. Connected components

The graph $G = (V, E)$ is connected if there is a path from any vertex to any vertex in the set V. If the graph is disconnected, it can be decomposed into several connected subgraphs, which are referred to as connected components of G. We define the size of a connected component $|CO|$ as the vertex number in it. Erdos and Renyi [19] studied how the connected components of random graph $G = G_{N,p}$ change with parameter p. If $p < 1/N$, almost surely all components are either trees or clusters containing exactly one cycle; if $1/N < p < (\ln N)/N$ almost surely the graph G has only one giant component; if $p > (\ln N)/N$ the graph G is almost all connected. Thus at $p \approx 1/N$ the random graph changes its topology abruptly from a loose collection of small components to a system dominated by a single giant component.

Aiello et al. [20] introduced a two-parameter random-graph model $P(\alpha, \gamma)$ defined as follows: Let n_k be the number of nodes with degree k. $P(\alpha, \gamma)$ assigns uniform probability to all graphs with $n_k = e^{\alpha} k^{-\gamma}$. It is demonstrated that when $\gamma < \gamma_0 = 3.47875$ there is almost surely a unique infinite component; when $\gamma > \gamma_0$ the random graph almost surely has no infinite component. In the stock correlation network, the connected component indicates the stock portfolio in which the price fluctuations are correlated with each other. The structural characteristics of components in the stock correlation network reflect the correlation modes of stocks in the market. In this paper, we will investigate how the connected components of stock correlation network depend on the threshold θ.

2.2.4. Clique

Given a subset $S \subseteq V$, by $G(S)$ we denote the subgraph induced by S. A subset $C \subseteq V$ is a clique if $G(C)$ is a complete graph, i.e. it has all possible edges. The size of a clique $|CL|$ could be denoted by the vertex number in it. The maximum clique problem is to find the largest clique in a graph. The financial interpretation of the clique in the stock correlation network is that it defines the set of stocks whose price fluctuations exhibit a similar behavior. If $\theta \geq 0$, all the stocks in a clique are positively correlated with each other. In such a case a price fluctuation of any one stock will make all other stocks' price in this clique fluctuate towards the same direction.

In practice, the portfolio in which price fluctuations of stocks are negatively correlated with each other could disperse investment risks. So another concept closely related to a clique is an independent set which is a subset $I \subseteq V$ such that the subgraph $G(I)$ has no edges. From the financial point of view, the independent set in the stock correlation network represents "the most diversified" portfolio, where all stocks are negatively correlated with each other. The size of an independent set $|IS|$ could be expressed by the vertex number in it. In this paper, we will study the relationship between maximum clique, maximum independent set in the stock correlation network and threshold θ.

The maximum independent set problem in the graph $G(V, E)$ can be easily reformulated as the maximum clique problem in the complementary graph $\bar{G}(V, \bar{E})$, which is defined as follows. If an edge $(i, j) \in E$, then $(i, j) \notin \bar{E}$, and if $(i, j) \notin E$, then $(i, j) \in \bar{E}$. Clearly, a maximum clique in $\bar{G}$ is a maximum independent set in G, so in this sense these problems are equivalent, and the complexity results obtained for the maximum clique problem are also valid for the maximum independent set problem.

3. Empirical study and results

3.1. Network construction

The network that we study in this paper represents the set of stocks traded in the Shanghai and Shenzhen stock markets in China. More specifically, we consider 1080 stocks (697 in Shanghai, 383 in Shenzhen) and analyze daily changes of their prices over a period of 1198 consecutive trading days in 2003–2007. Based on this information, we calculate the cross-correlation coefficients c_{ij} between each pair of stocks using formula (2). The correlation coefficients c_{ij} can vary from −1 to 1.

Fig. 1 shows the distribution of the correlation coefficients based on the price data for the years 2003–2007. It can be seen that this plot has a shape similar to the normal distribution with a mean of 0.38. Most coefficients are at intervals of

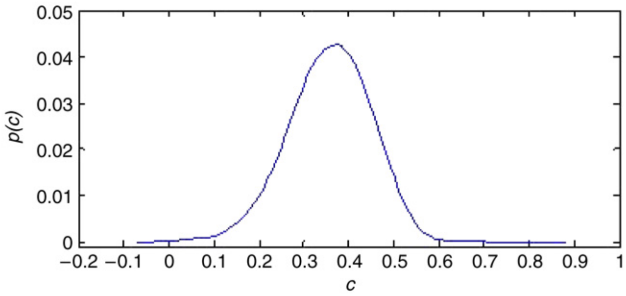

Fig. 1. Distribution of correlation coefficients in the stock market. The distribution shape is similar to the normal distribution. Most of the coefficients are between 0.1 and 0.6, few are negative. This distribution indicates "up and down in the same direction" phenomena in Chinese stock market.

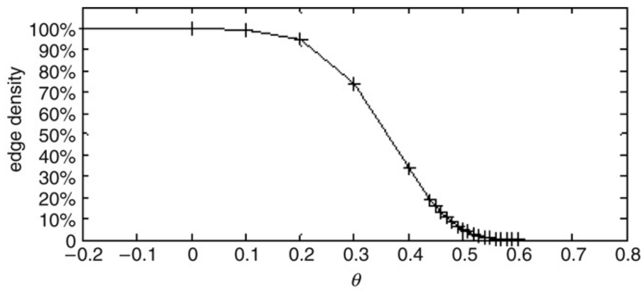

Fig. 2. Edge density of the stock correlation network for different values of the correlation threshold.

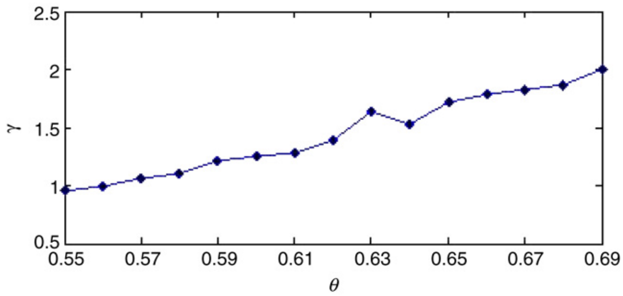

Fig. 3. The network degree distribution parameter for different values of the correlation threshold. Firstly, we do a regression analysis on formula (6). If the linear relationship is significant, formula (5) is valid, that is the network is scale-free. We find that only when $\theta \in [0.55, 0.69]$, does the network degree distribution obey a power-law model.

[0.1, 0.6], and all of them are larger than −0.1. It tells us the prices of most stocks in the Shanghai and Shenzhen markets fluctuate towards the same direction. It is easy to see that the number of edges in the stock correlation network decreases as the threshold value θ increases. The corresponding graph is presented on Fig. 2 which demonstrates the edge density of the network (existing edge number/maximal edge number C_N^2) with respect to θ. We can see that the edge density drops sharply from 99.35% to 0.12% as θ increases from 0.1 to 0.6. It is because most of the correlation coefficients of stock price are virtually at this threshold interval.

3.2. Network topology structure

3.2.1. Degree distribution

The power-law model fairly well describes some of the real-life massive networks, such as the WWW, the Internet, movie actor collaboration networks, cell networks and telephone networks etc. In this section, we will show that the stock correlation network also obeys the power-law model. It should be noted that since we consider a set of networks, where each network corresponds to a certain value of θ, the degree distributions will be different for each θ.

The results of our empirical study turned out to be rather interesting. For $\theta \in [0.55, 0.69]$ the degree distribution is approximately a straight line in the log–log scale, which is exactly the power law distribution. Fig. 3 demonstrates the degree distribution parameter γ with respect to different $\theta \in [0.55, 0.69]$. As can be seen from it, the parameter γ increases with θ. A larger γ indicates a more uniform vertex degree and fewer vertices having a relatively large degree.

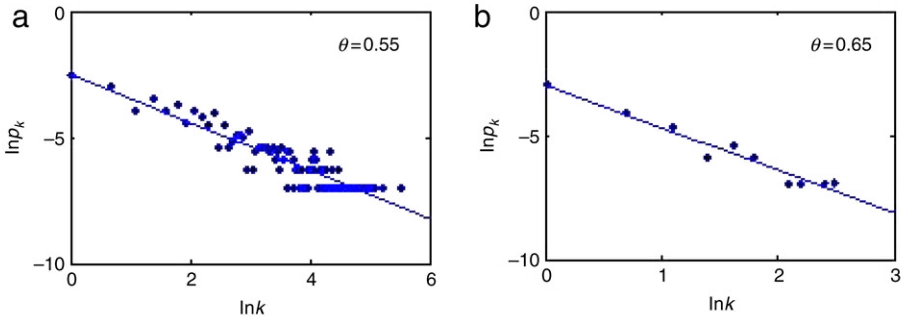

Fig. 4. Log–log linear fitting of network degree distribution for (a) $\theta = 0.55$; (b) $\theta = 0.65$, When θ equals to other values at interval [0.55, 0.69], the degree distribution is similar to that of $\theta = 0.55, 0.65$ except for the distribution parameter value.

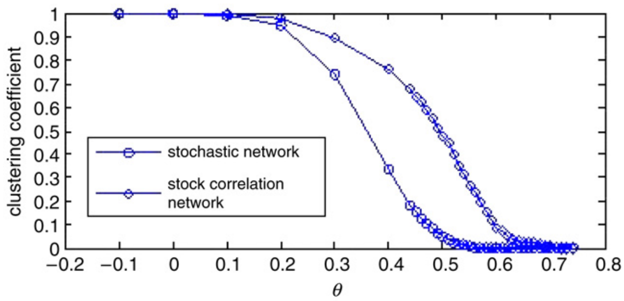

Fig. 5. The clustering coefficients comparison between the stock correlation network and a stochastic network of the same size. We can see that at certain correlation threshold intervals, the stock correlation networks are highly clustered.

If we specify other correlation thresholds, the distribution of the degrees of the vertices is very "noisy" and does not have any well-defined structure. Note that for these values of correlation threshold the network has too low ($\theta > 0.69$) or too high ($\theta < 0.55$) an edge density. The network structure seems to be very difficult to analyze in these cases. Especially, we need to calculate the clique in the complementary network when calculating the independent set in the stock correlation network as $\theta < 0$. We find that the degree distribution in the complementary network when $\theta < 0$ is similar to that of $\theta > 0.69$ or $\theta < 0.55$ in the original network. So we can conclude that only when the network edge density is moderate, is the stock correlation network a scale-free network.

The larger a stock's degree is, the more stocks it is correlated with in the sense of price fluctuation. From the point of price fluctuation influence, the stocks that have large degree generally have a "higher status" and "stronger market influence power". So the scale-free property of the stock correlation network demonstrates that most of the stocks are at the same level, at the same time, a small quantity of stocks have a "higher status" and "stronger market influence power". The latter play a very important role in the price fluctuation correlations of the whole network.

Fig. 4 demonstrates the degree distributions of the networks for some values of the correlation threshold ($\theta = 0.55, 0.65$).

3.2.2. Clustering coefficient

Fig. 5 shows the clustering coefficients of the stock correlation network and the stochastic network with the same size under different correlation thresholds. The clustering coefficients of stock correlation networks become smaller with an increase of thresholds. If $\theta \in [0.54, 0.71]$, the clustering coefficients of the stock correlation networks are 20 times (or more) greater than stochastic networks. Especially, if $\theta \in [0.58, 0.71]$ and $\theta \in [0.62, 0.71]$, the multiples are 50 and 100 respectively. So the stock correlation networks are highly clustered as $\theta \in [0.54, 0.71]$. Moreover, the complementary networks' clustering coefficients are 0 when $\theta < 0$. Integrating with above results, we know that the stock correlation networks are both scale-free and highly clustered when $\theta \in [0.55, 0.69]$. This property is consistent with many other complex systems.

Comparing Fig. 2 with Fig. 5, we find that the decrease rate of clustering coefficients is much smaller than that of edge densities. The stock correlation networks are highly clustered even when the edge density is rather small. For example if $\theta \in [0.61, 0.71]$, the edge densities are much smaller than 1% while the clustering coefficients are relatively large.

3.2.3. Component structure

In the stock correlation network, components represent stocks that are correlated with each other. We find that the component structure of a stock correlation network depends on thresholds θ. Fig. 6 shows the component number of the

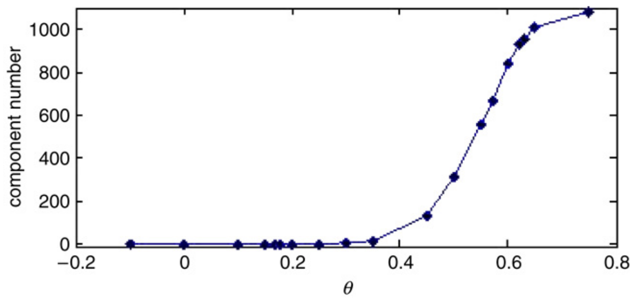

Fig. 6. The component number in the stock correlation network under different correlation threshold. If $\theta > 0.75$, the vertices are nearly all isolated and the component number is approximately the vertices number; if $\theta < -0.1$, the networks are full connected, the component number is just one.

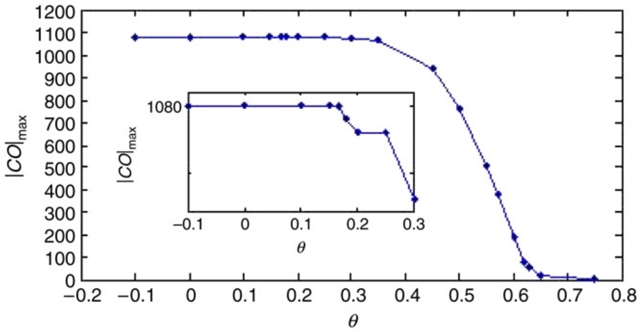

Fig. 7. The maximum component size of stock correlation network under different correlation threshold. The inset figure is an enlarged drawing of original figure when $\theta \in [-0.1, 0.3]$. From it we can clearly find that $\theta = 0.17$ is a critical point whether the network is full connected.

network under different thresholds. As can be seen, the component number increases with the threshold. Fig. 7 is the maximum component size under different thresholds. Comparing Fig. 6 with Fig. 7, we can see that the greater a component number is, the smaller the maximum component size will be.

The inset figure of Fig. 7 clearly demonstrates that the maximum component size is 1080 when $\theta \leq 0.17$ and the network is full connected. This means any a stock's price fluctuation will directly or indirectly influence all other stocks. If $\theta > 0.17$, there are some isolated stocks whose price fluctuations do nothing with other stocks. So $\theta = 0.17$ is the critical point whether the network is full connected. The network has a unique giant component whose size is larger than 100 if $\theta \leq 0.6$. If $0.6 < \theta < 0.75$, the network has several small components whose size is between 10 and 100. If $\theta > 0.75$, the network consists of large number of isolated stocks, the largest component size being smaller than 10. In the last case, there are almost no correlations among stocks and the network's connectivity is low.

As we know, the stock correlation network is scale-free when $\theta \in [0.55, 0.69]$. Meanwhile, as the threshold is at this interval, the maximum component size is between 10 and 506 and the component number is between 559 and 1030. As a result, though the component structures are quite different, the networks can present a similar degree distribution. From Fig. 3 we know that the smaller θ is, the smaller the degree distribution parameter is and the more vertices that are highly connected will have. This can explain why the maximum component size gradually increases as θ decreases.

3.2.4. Clique and independent set

A clique in the stock correlation network represents a stock portfolio in which the stocks have similar price fluctuation characters. When θ is a relative large positive number, the price fluctuations of stocks in a clique are significantly positively correlated. In actual investments, a portfolio in which stocks are negatively or slightly positively correlated can help to reduce risks. When $\theta \leq 0$ (or a small positive number), the independent set in the stock correlation network is just the stock portfolio that can satisfy aforesaid conditions. It equals the clique in the complementary network.

The maximum clique and the maximum independent set problems are NP-hard [21]. Traditional deterministic arithmetic cannot solve these problems because of the increase of the vertex number and edge density. In general, software Ucinet can be used to solve the maximum clique of a network. Unfortunately, this software is based on deterministic arithmetic, so it cannot solve the aforesaid problems in the stock correlation network due to its large number of vertices and large edge density. Therefore, we use a greedy heuristic for finding a lower bound of the clique number, and a special preprocessing technique which reduces a problem's size. This greedy heuristic was introduced in Ref. [16]. After this preprocessing, we transform the original stock correlation network $G = (V, E)$ into $G' = (V', E')$ with a smaller size. As a result, the maximum clique (maximum independent set) which was solved in $G' = (V', E')$ by Ucinet is just that of the original network.

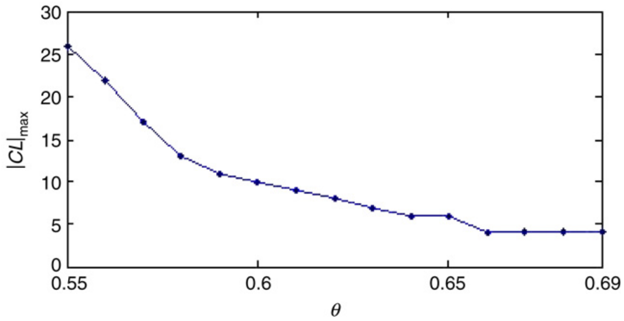

Fig. 8. The maximum clique size of scale-free stock correlation networks. When $\theta \in [0.55, 0.69]$, the networks are both highly clustered and scale-free, so we analyze the clique at this interval.

Table 1
The maximum independent set size of a stock correlation network at certain correlation thresholds.

θ[a]	0	0.05	0.1	0.15	0.2
$\|CL\|_{max}$	0	4	7	12	23

[a] Means the complementary network, that is if $c_{ij} < \theta$, $e_{ij} = 1$; or else $e_{ij} = 0$.

To find a large clique, we apply the following greedy algorithm (this algorithm is used to solve the maximum clique problem, and the maximum independent set can be indirectly obtained by solving the maximum clique in the complementary network). Starting with an empty set G_e, we recursively add to the clique a vertex from the neighborhood of the clique adjacent to the most vertices in the neighborhood of the clique. If we denote by $N(i) = \{j|(i, j) \in E\}$ the set of neighbors of i in $G = (V, E)$, then the neighborhood of a clique G_e is $\bigcap_{i \in G_e} N(i)$, and we obtain the following algorithm:

$$G_e = \phi, G_0 = G = (V, E);$$

do

$$G_0 = \bigcap_{i \in G_e} N(i) \setminus G_e;$$

$$G_e = G_e \cup j, \quad \text{where } j \text{ is a vertex of largest degree in } G_0.$$

Until $G_0 = \phi$.

After running this algorithm, we applied the following preprocessing procedure. We recursively remove from the graph all of the vertices which are not in G_e and whose degree is less than $|G_e|$, where G_e is the clique found by the above algorithm. This simple procedure enabled us to significantly reduce the size of the maximum clique search space. Let us denote by $G' = (V', E')$ the graph induced by remaining vertices. The maximum clique in $G' = (V', E')$ is just that of original network $G = (V, E)$. It is demonstrated that the network size is largely reduced after this preprocessing. Then we can efficiently solve the maximum clique problem by Ucinet.

If $\theta \in [0.55, 0.69]$, the stock correlation networks have large clustering coefficients and scale-free degree distribution. Fig. 8 represents the maximum clique size of the networks when the thresholds are at this interval. We find that the maximum clique size becomes smaller when the threshold increases. It is demonstrated that the networks do not have cliques as $\theta \geq 0.72$. Table 1shows the maximum independent set size when $\theta \leq 0.2$. From this table, we can see that the maximum independent set size becomes smaller with a decreasing threshold. Especially, the networks do not have cliques as $\theta \leq 0$. It tells us that we cannot find a risk diversified investment portfolio in which the price fluctuation correlation coefficient between every two stocks is negative in a strict sense. So the price behaviors of stocks do not demonstrate enough variety.

3.3. *Topological stability of stock correlation networks*

The topological stability measurement methods of stock correlation networks mainly include vertex attack and edge attack. If the structural properties such as connectivity have not been essentially changed after attacks, we call the network robust against attack. Considering the characteristics of a stock correlation network, we apply the vertex attack method, namely remove some vertices and all the edges that are connected with them from the network. This kind of attack corresponds to some extreme risks of listed companies such as delisting or bankruptcy. Then we use stochastic and selective removal methods respectively. Stochastic removal means removing some vertices from the network in a stochastic manner. Meanwhile, selective removal means removing the vertices in a certain order, namely from the vertices with the largest degree and so on.

As we know, the maximum connected component size reflects the connectivity of a stock correlation network. So we measure the stability of networks according to the relative changes of the maximum connected component size before and after vertex removal. We denote by $|CO|_{max}$ the maximum component size of a stock correlation network under certain

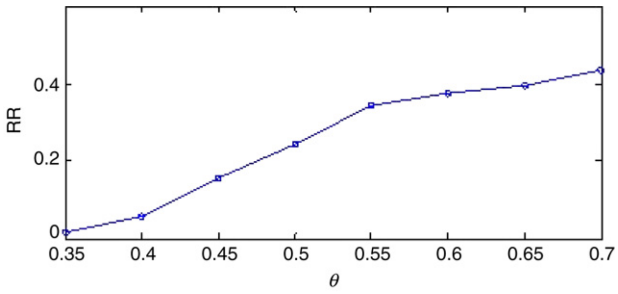

Fig. 9. The vertex removal effect comparison of stochastic removal and selective removal at different values of correlation threshold. We only analyze the threshold interval [0.3, 0.7], because under other circumstances, the networks are too dense or too sparse and are not suitable for analyses.

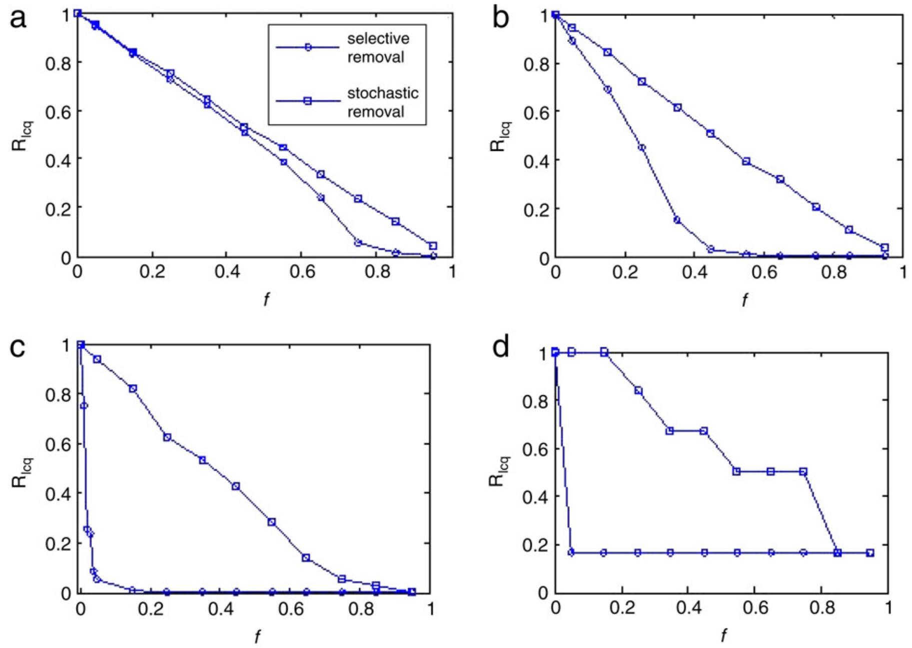

Fig. 10. The topological stability of stock correlation networks when (a) $\theta = 0.4$; (b) $\theta = 0.5$; (c) $\theta = 0.6$; (d) $\theta = 0.7$. With an increase of correlation threshold, we find that the network becomes more and more fragile to selective removal and is always robust against stochastic removal. If the threshold is equal to other values in interval [0.3, 0.7], the general rules of topological stability are not changed. So we only demonstrate the aforesaid four cases.

threshold θ. After stochastic or selective removal of $f * N$ (N is the network vertex number, f is the removed proportion) vertices, the ratio of the maximum component size of the new network to that of the original network is $R_{|CO|}$ ($0 < R_{|CO|} \leq 1$).

We introduce index *RR* to compare the influence of stochastic and selective removal on the network connectivity. In the same network, the influence of vertex removal on network connectivity will change with removal proportion f and removal method (stochastic or selective). We denote by r and d the stochastic and selective removal manner respectively, defining index *RR* as follows

$$RR = \frac{\sum_{f}(R_{|CO|}(r,f) - R_{|CO|}(d,f))}{m} \tag{6}$$

where $R_{|CO|}(r,f)$ and $R_{|CO|}(d,f)$ represent the influencing effects of stochastic and selective removal respectively when the removal proportion is f, m is the number of different removal proportions that we choose. Here we let $f = 0.05, 0.1, 0.15, \ldots 0.95$. From formula (6), we find that *RR* represents the average influence difference between the two vertex removal methods. The larger *RR* is, the larger the average influence difference will be. Fig. 9 shows values of index *RR* of a stock correlation network under different thresholds. When $\theta < 0.3$ or $\theta > 0.7$ the network is too dense or too sparse, so we only consider cases of $\theta \in [0.3, 0.7]$. From the figure we can see that the value of *RR* increases with threshold θ. This result indicates that a stochastic and a selective removal make no difference to removal effects. With an increase of threshold, the influence effect differences of such two removal methods become larger and larger.

Fig. 10(a–d) show the stability of stock correlation network when $\theta = 0.4, 0.5, 0.6, 0.7$. If $\theta = 0.4$, the removal effects of stochastic and selective removal are comparable. The influence of selective removal on connectivity is larger than that of stochastic removal when $\theta = 0.5$. If $\theta = 0.6, 0.7$, a minor removal proportion can make the network connectivity drop rapidly, while it needs to remove a large number of vertices to realize the same connectivity reduction extent under stochastic removal. For example, in Fig. 10(c), the selective removal of 5% vertices can make the maximum connected component size reduced to 5.82% of the original size, while we need to stochastically remove 75% vertices to realize the same effects.

In conclusion, the connectivity of a stock correlation network becomes more and more fragile to selective removal with an increase in threshold, and is always robust against stochastic removal. We can explain that networks become sparser and the degree distribution gets farther away from uniformity as θ increases. At this time, the vertices that have a large degree are always "key" members of large components. The removal of such "key" vertices will delete vast correlation relationships among vertices in a component and make the component size reduce fast.

On one hand, the robustness against stochastic removal indicates that some stochastic events of listed companies such as delisting or bankruptcy will not have an essential influence on the whole price fluctuation correlation of the networks. On the other hand, the frangibility to selective removal demonstrates the stocks which have a large degree play very important roles in the whole market price fluctuation correlation. So we must pay more attention to some fateful events of listed companies which have a large degree in the corresponding stock correlation network. This attention will help investors understand the dynamic price fluctuation correlation patterns among stocks, and such an understanding can be used for references in portfolio investments and risk managements.

4. Conclusions

In this paper, we presented a detailed study of the properties of the stock correlation network. Finding components, cliques and independent sets in the network gives us a new tool for the analysis of the market structure by classifying the stocks into different groups. It would be helpful to investors for making decisions regarding their portfolios. Therefore, this technique is useful from both theoretical and practical points of view.

The statistical analysis of the degree distribution of the network has shown that the power-law model is valid in financial networks. It confirms an amazing observation that many real-life massive networks arising in diverse applications have a similar power-law structure, which indicates that the global organization and evolution of massive datasets arising in various spheres of life follow similar laws and patterns.

Stock correlation networks display a topological robustness against random node failures, but they are also fragile to intentional attacks. The result indicates that certain stocks having large degree are important for the price fluctuation correlation patterns of stocks. Such a network stability property would be also useful for portfolio investments and risk management.

Although we addressed many issues in our analysis of the stock correlation network, there are still a lot of open problems. For instance, we can study other types of stock correlation network based on the data of the liquidity of financial instruments. It would be very interesting to study the properties of this network and compare it with the network considered in this paper. Such a comparison would be useful for studying the relationship between return and liquidity, which is one of the fundamental problems of the modern finance theory. Therefore, this research direction is very promising and important for a deeper understanding of the market's behavior.

References

[1] R. Albert, A.-L. Barabasi, Reviews of Modern Physics 74 (2002) 47.
[2] M.E.J. Newman, SIAM Review 45 (2003) 167.
[3] T. Zhou, W.J. Bai, B.H. Wang, et al., Physics 34 (2005) 31.
[4] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[5] A.-L. Barabasi, R. Albert, Science 286 (1999) 509.
[6] M.E.J. Newman, D.J. Watts, Physics Letters A 263 (1999) 341.
[7] A. Barrat, M. Weigt, The European Physical Journal B 13 (2000) 547.
[8] G. Bianconi, A.-L. Barabasi, Physical Review Letters 65 (2001) 5632.
[9] X. Li, G. Chen, Physica A 328 (2003) 274.
[10] S.W. Chen, Journal of Management Science in China 2 (1999) 1.
[11] R.N. Mantegna, The European Physical Journal B 11 (1999) 193.
[12] H.J. Kim, Y. Lee, B. Kahng, I.M. Kim, Journal of Physical Society of Japan 71 (2002) 2133.
[13] J.P. Onnela, K. Kaski K, J. Kertesz, The European Physical Journal B 38 (2004) 353.
[14] M. Tumminello, T. Aste, T.D. Matteo, R.N. Mantegna, Proceedings of National Academy of Sciences 102 (2005) 10421.
[15] M. Tumminello, T.D. Matteo, T. Aste, R.N. Mantegna, The European Physical Journal B 55 (2007) 209.
[16] V. Boginski, S. Butenko, P.M. Pardalos, Computational Statistics & Data Analysis. 48 (2005) 431.
[17] A.E. Motter, T. Nishikawa, Y.C. Lai, Physical Review E 66 (2002) 065103.
[18] R. Albert, H. Jeong, A.-L. Barabasi, Nature 406 (2000) 378.
[19] P. Erdos, A. Renyi, Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5 (1960) 17.
[20] W. Aiello, F. Chung, L. Lu, Experimental Mathematica 10 (2001) 53.
[21] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-completeness, Freeman, New York, 1979.