


Research Article

Complexities in Financial Network Topological Dynamics: Modeling of Emerging and Developed Stock Markets

Yong Tang ^{1,2,3}, Jason Jie Xiong,⁴ Zi-Yang Jia,⁵ and Yi-Cheng Zhang^{3,6}

¹Center of Cyberspace and Security, School of Computer Science and Engineering,
University of Electronic Science and Technology of China, Chengdu 610054, China

²School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 610054, China

³Department of Physics, University of Fribourg, Chemin du Musée 3, CH-1700 Fribourg, Switzerland

⁴Department of Computer Information Systems and Supply Chain Management, Walker College of Business,
Appalachian State University, Boone, NC 28608, USA

⁵Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA

⁶Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China

Correspondence should be addressed to Yong Tang; tangyong@uestc.edu.cn

Received 17 May 2018; Accepted 16 September 2018; Published 1 November 2018

Guest Editor: Claudio Tessone

Copyright © 2018 Yong Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Policy makings and regulations of financial markets rely on a good understanding of the complexity of financial markets. There have been recent advances in applying data-driven science and network theory into the studies of social and financial systems. Financial assets and institutions are strongly connected and influence each other. It is essential to study how the topological structures of financial networks could potentially influence market behaviors. Network analysis is an innovative method to enhance data mining and knowledge discovery in financial data. With the help of complex network theory, the topological network structures of a market can be extracted to reveal hidden information and relationships among stocks. In this study, two major markets of the most influential economies, China and the United States, are systematically studied from the perspective of financial network analysis. Results suggest that the network properties and hierarchical structures are fundamentally different for the two stock markets. The patterns embedded in the price movements are revealed and shed light on the market dynamics. Financial policymakers and regulators can gain inspiration from these findings for applications in policy making, regulations design, portfolio management, risk management, and trading.

1. Introduction

The visualization of networks and research of hierarchy structures are essential to study complex systems like financial markets. Thanks to the significant development of complex network science [1], quantitative methods and models have been applied in the studies of financial markets network structure. In financial network analysis, entities like assets, stocks, markets, companies, and institutions are modeled as vertices while their mutual relationships are abstracted as edges. This approach empowers industrial professionals and researchers to reveal hidden information embedded in the topological structures of financial networks, such as the market dynamics, trading activities, and investment sentiment.

This information is essential to evaluate and monitor the financial market risks, contagions, distress propagation, as well as market mode shifts. Financial network analysis has been utilized in applications like portfolio management, trading, market regulation, stress testing, and risk management.

The USA and China are the top two dominating economies with influences over the global economies. The two economies are similar in market size. However, the US economy is well established and developed while the Chinese economy is emerging and still undertaking fast development. As the leading economic powerhouses, the health and stabilities of these two economies are vital for the prosperities of the world economy. During the past few years, both countries suffered a series of stock market disasters, such

as the 2008 US subprime crisis and the 2007 and 2015 Chinese stock market bubble bursts. These dramatic market crashes brought widespread and long-lasting negative impacts on economies and markets. The stock markets of both countries are also different regarding history, regulations, maturities, and scales. Thus, it is essential to understand the properties of these two markets by utilizing the data-driven science approach. Recently, various major markets in the world have been investigated using the financial network analysis approach. However, there is still a lack of systematical studies dedicating to compare the network structures and properties of the US and Chinese stock markets using the financial network analysis approach.

To understand how the two markets differ in the structures and topological properties, as well as the dynamics market properties, this research investigates the markets using a dataset spanning over nine years. In this research, the stock markets are modeled as multiple networks including hierarchical trees, minimum spanning trees, planar maximally filtered graphs, and assets graphs. Meanwhile, their detailed topological properties are analyzed and systematically compared. Through quantitative analysis and network visualization, results show the two markets are different in many ways. This provides insights for regulators on the structures and dynamics of stock markets from the perspective of network science.

This paper is organized as follows. First, Section 2 gives a background on the theory of complex networks, financial network analysis, and relevant complex network parameters are introduced. Then, in Section 3, the data and method used to construct networks are described. Section 4 presents the network properties. In Section 5, the detailed hierarchical structures of both stock networks are carefully investigated and compared. Finally, conclusions and discussions are presented in Section 6.

2. Literature Review of Financial Network Analysis

Network science has become an innovative tool widely used in studies of complex systems in a variety of engineering and scientific domains [2–4]. The network modeling methodologies and theoretical frameworks have revealed informative and useful empirical discoveries [5]. Studying the statistical properties such as degree distribution, average length, and clustering coefficient can help describe the networks topologies and the dynamics of network evolution. Furthermore, it is possible to study the information spreading, network stability, and phase changes and hopefully to predict and control the network dynamics [6].

Time series data can be modeled as networks [7, 8]. For price time series [9], calculation of correlation matrices for a group of assets is possible [10–12]. From the correlation matrices, financial network analysis could be applied to construct networks for further analysis and data mining [13–17]. In most existing literature, assets are treated as vertices, while the interconnectivity relationships are modeled as pairwise edges among assets. The correlation matrices are not only important for network analysis and topological visualization

but also serve as a bridge between financial network analysis and traditional finance theories. This is similar to modern portfolio theory (MPT) [18, 19], which is based on the correlation relationships among assets. Network-based portfolio selection has been proposed for optimization and empirically proofed workable [20].

Since the minimum spanning tree approach is first used in the study of stock market structure [21], financial network analysis has grown into an essential tool of financial big data. However, this fast-growing field is still at an early stage [22]. Financial network analysis provides an unprecedented perspective shedding new insights on evaluating the market stability, market risk, shock propagation, and contagion [23–25]. The connectedness among assets plays the critical role of market contagion phase transition [26] which is similar to other tolerance properties of other nonfinancial complex networks [27]. Further research reveals that intermediate level of risk diversification can enhance the market robustness [28]. The importance and risk contribution of companies can be identified through the network analysis [29, 30]. The systemic risks and stability can also be evaluated according to the topological properties of the financial network, and providing implications for market regulations [22, 31, 32]. Through investigating the clustering of assets, portfolio optimization can be achieved with better predicted over realized risks [33]. Overlapping of portfolios is revealed by network analysis as one primary factor for market contagion [34]. In another approach, risk spillover networks are constructed to study the behaviors of financial institutions [35]. Instead of a single layer approach, by building multiple-layer network, the banking system risk is analyzed and quantified [36]. Regression models can also take network structure into consideration as factors for resilience and robustness of the markets [37]. This financial network approach opens more interesting new possibilities and dramatically enriches regression models in finance studies. Recently, there has been a thread of studies on major players in financial networks, such as ‘too interconnected to fail’ institutions [38], ‘too central to fail’ [39], and ‘too big to fail’ [40]. The research demonstrated that financial network analysis brings new insights to finance studies and benefits to finance practices.

In the rise of quantitative trading, the causality and lead/lag relationships revealed by financial network analysis can be particularly interesting for trading strategy design [41, 42]. Many researches have revealed stylized evidence that the network structure has a profound influence on the asset returns [43]. Taking risks into consideration, it has been found that investing in peripheries of financial networks might generate better returns over risks [44]. Furthermore, industry professionals would be inspired by financial network analysis to seek price movement signals for potential predictions [45].

While most existing financial network analysis literatures focus on the stock markets or specific economy sectors [46–48], a variety of financial systems have been studied as financial network such as global financial institutions [39, 40], world trade web [49, 50], interbank markets [51–54], exchanges [55], monetary market [56], corporate networks [57, 58], global banking [59], CDS market [60], and credit market [61].

Many major individual financial markets around the world have also been studied in network approach, such as US [62, 63], China [64–66], Germany [67, 68], EU [69–71], Brazil [33, 72], Italian [53, 56], Korea [73], Russia [74], and Mexico [36, 75]. Furthermore, there is literature focusing on the cross-board global markets [76, 77]. Using the partial data, networks of global markets are reconstructed, and methods are compared [56]. Bayesian graphical models are applied to identify groups of countries which are major contributors to systemic risks according to banking behaviors in the global banks [78]. For the European markets, the risk and contagion channels are studied, and results show the EU markets are vulnerable to risks [71]. Global stock exchange network is investigated to evaluate the attractions for IPOs [79]. A recent study has demonstrated the approach which uses transfer entropy to study a selection of major individual stocks around the world and reveals that stocks are clustered according to their countries and industries [80]. By looking into the network structures of the global financial network, it is possible to give new insights into the international business cycle [81]. The diversification and participation are investigated for various economies [82].

Considering a large number of assets in financial markets, the initial networks have massive edges. By filtering the noises of the networks, the financial networks can be significantly simplified to enable advanced analysis such as principal component analysis (PCA) [76] and random matrix theory (RMT) [83, 84] to further extract hidden patterns. Hierarchical tree (HT) [21, 68], minimum spanning tree (MST) [85], planar maximally filtered graphs (PMFG) [86], asset graph (AG) [87, 88], and partial correlation network [15, 89, 90] are major approaches applied in filtering financial networks. Mantegna [21] first introduces the minimum spanning tree method into the study of hierarchical structures in financial markets. With the network, we can study the topological structure of a market or a portfolio. In this research, we adopt the frameworks to study the correlations and the corresponding networks of stock markets both in China and the United States to systematically study how the two markets behave differently.

3. Data and Research Methods

3.1. Indices of CSI300 and S&P500. We study the stock markets of China and United States: the former is a typical representative of emerging countries with fast-growing GDP rate and influence on global economies, while the latter is the most established and developed economy in the world. To study the major stocks of each market, we focus on the component stocks of the major indices of the two stock markets, i.e., China Securities Index 300 (CSI300) for the Chinese stock market and Standard & Poor's 500 (S&P500) for the US market. In our study, we cover a period of nine years starting from 04/01/2007 and ending on 06/11/2015 with 2149 trading days for CSI300 and 2228 trading days for S&P500. The reason why the two markets have different numbers of trading dates is that the two markets have different trading calendars. Index and all component stocks daily price data of CSI300 are retrieved from the CSMAR

Solution Database of Shenzhen GTA Education Tech. Ltd. We download the S&P500 index and component stocks daily prices data through Yahoo finance service. Since not all stocks are traded on each trading date, so we only select those CSI300 stocks with at least 2000 trading dates, and without continuous 100 nontrading dates, this selection results in a final set of 163 stocks. For S&P500, we select those stocks with at least 2100 trading dates, and in results, we get 468 stocks. After stocks selection, we take the prices on the available closest trading date to fill the nontrading dates. In Figures 1(a) and 1(b) we plot the daily close prices and the daily log returns for the index of CSI300 in the study period of 04/01/2007 and 06/11/2015 with 2149 trading days. In Figures 1(c) and 1(d), we plot the daily close prices and the daily log returns for the index of S&P500 in the same study period from 04/01/2007 to 06/11/2015 with 2228 trading days. From the figures, we see that the two markets show large fluctuations in the last nine years. CSI300 experienced huge market crashes in 2008 and 2015, while S&P500 kept climbing almost continuously after the 2008 financial crisis.

3.1.1. CSI300. China has two independent stock market exchanges, i.e., the Shanghai stock exchange and the Shenzhen stock exchange. Opened at the beginning of 1990s with only 25 years of trading history, the two markets have grown into important financial markets playing vital roles in China's financial markets and economy. Among the many stock market indices, the China Securities Index 300, or CSI300, was introduced by the China Securities Index Company, Ltd. in 2005 to a base of 1000 on 31/12/2004. In CSI300, a set of 300 stocks are included as the index components; all of them have the largest market values and are actively traded in Shanghai or Shenzhen stock exchanges. CSI300 has become a widely accepted benchmark to evaluate the whole stock markets behaviors in China as well as a good basis for other derivative products. Starting from 1000 points in the early of 2004, now CSI300 has reached 3793 points as of 06/11/2015 [91]. To give an image of the Chinese stock market, we plot the 2149 CSI300 index daily close prices and daily log returns in the study period between 04/01/2007 and 06/11/2015 in Figures 1(a) and 1(b). In the past nine years, CSI300 experienced two major market crashes in 2007–2008 and 2015, respectively, during which the market suffered huge losses and fluctuations. There are 163 stocks of CSI300 component stocks included in our dataset, as shown in Table 1, in which we summarize the numbers of these 163 stocks for all 20 industry sectors. As shown, all industry sectors from Agriculture to Comprehensive are represented. For convenience, we will refer to these 163 stocks as *CSI163* in the following parts.

3.1.2. S&P500. Compiled by Standard & Poor's in 1957, the S&P500 is an established American stock market index with more components, more risk diversification, and better reflection of the overall stock market performance than both the New York Stock Exchange (NYSE) and Nasdaq. All components are large stocks in capitalization with good liquidities and diversifications in different industry sectors. The S&P500 represents major parts of the market and is

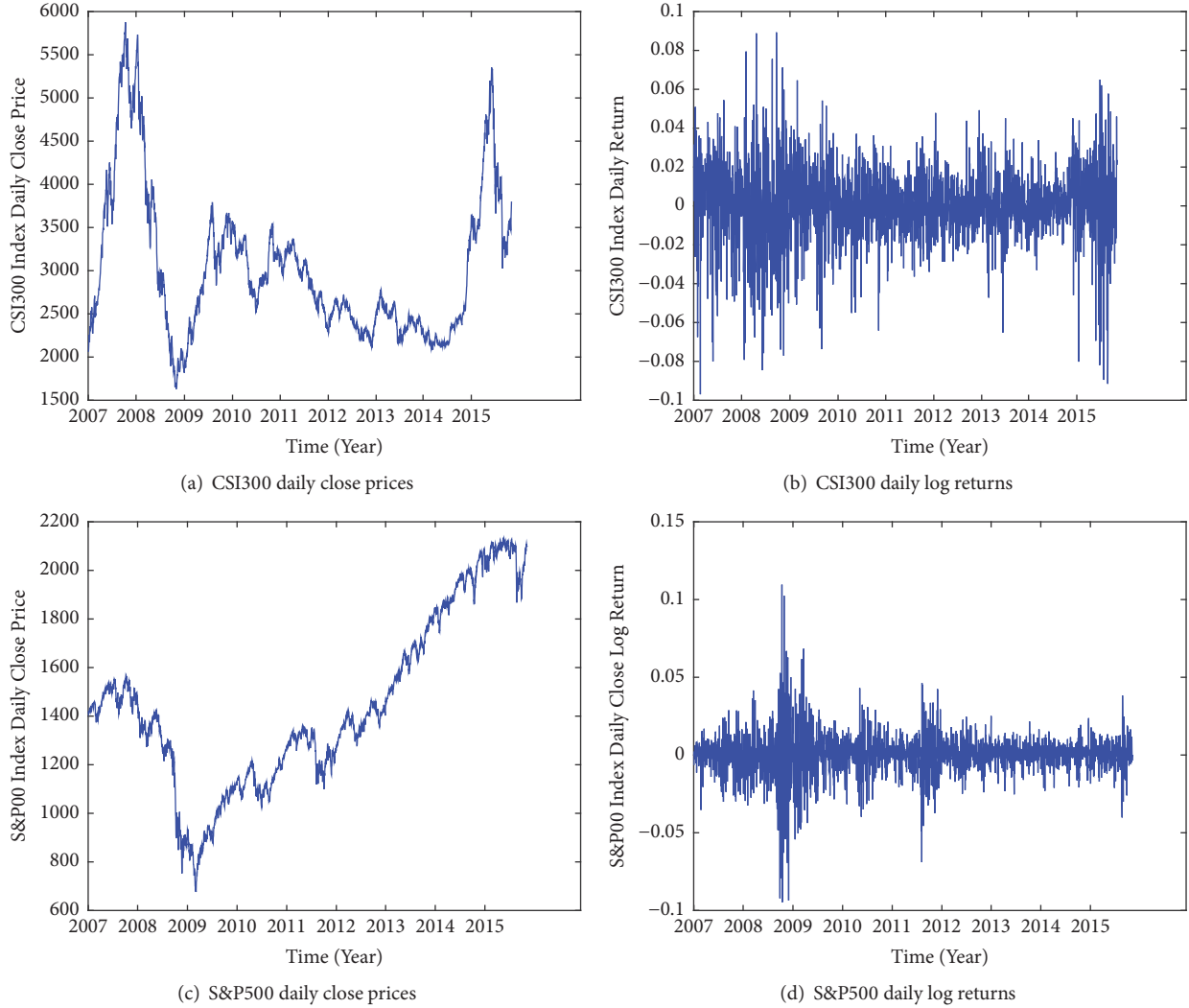


FIGURE 1: CSI300 index daily close prices (a) and log returns (b) in the study period between 04/01/2007 and 06/11/2015 with 2149 trading days. S&P500 index daily close prices (c) and log returns (d) in the study period between 04/01/2007 and 06/11/2015 with 2228 trading days.

considered as one of the best benchmarks for the US financial markets and economy. Starting from less than 100, after more than 50 years of development, the S&P500 reached 2099 on 06/11/2015 [92]. We plot 2228 daily close prices and log returns of the index of the S&P500 in our study period between 04/01/2007 and 06/11/2015 in Figures 1(c) and 1(d). We can observe that the S&P500 index suffered a major crash between 2008 and 2009 then recovered almost steadily with minor fluctuations. After the selection, there are 468 stocks of the S&P500 component stocks included in our dataset, as shown in Table 2. We summarize the numbers of these 468 stocks for all ten industry sectors. As shown, all industry sectors from Energy to Utilities are represented. For convenience, from now we will refer to these 468 stocks as $S\&P468$ in the following parts. Table 3 gives a summary of the two datasets of both CSI163 and S&P468; CSI163 has a larger standard deviation σ_r of the log returns, indicating larger fluctuations than S&P468. We use $\langle x \rangle$ to denote the average of variable x in this paper.

3.2. Construction of Stock Networks

3.2.1. Price Returns and Correlations. From the time-stamped price time series of a blanket of stocks, it is possible to calculate the correlations for any pair of stocks once a time window is given. $P_i(t)$ is the price at time t of stock s_i . It could be one of the daily prices of open, close, high, or low. Per most literature suggests, we choose the most used daily close price. To smooth the fluctuation without loss of generality, the logarithmic return for s_i in the period of $[t - \Delta t, t]$ is defined as

$$Y_i(t) = \ln P_i(t) - \ln P_i(t - \Delta t), \quad (1)$$

and usually used instead of $P_i(t)$ itself. In most cases, daily log returns are used where $\Delta t = 1$. For stock pair of s_i and s_j , we can extract the two price time series in a time window with a length or size of L , i.e., with L price values included in the window. The selection of L is expected to meet the requirement of $L/N > 1$. In a sliding window approach, we

TABLE 1: 163 component stocks of CSI300 are included in our dataset. In this table, we list the China Securities Regulatory Commission (CSRC) industry code, sector name, and numbers of stocks for each industry sector of these 163 stocks. All 20 industry sectors are represented.

Industry code	Industry Sector	Number of Stocks
A	Agriculture	1
B	Mining	6
C0	Food & Beverage	4
C1	Textiles & Apparel	4
C3	Paper & Printing	2
C4	Petrochemicals	9
C5	Electronics	7
C6	Metals & Non-metals	20
C7	Machinery	27
C8	Pharmaceuticals	15
D	Utilities	6
E	Construction	5
F	Transportation	10
G	IT	8
H	Wholesale & retail trade	10
I	Finance and insurance	10
J	Real estate	11
K	Social Services	3
L	Communication & Cultural Industry	2
M	Comprehensive	3

TABLE 2: 468 component stocks of S&P500 are included in the dataset. In this table, we list the Global Industry Classification Standard (GICS) code, sector name, and number of stocks for each industry sector in S&P500. All ten industry sectors are represented.

Industry code	Industry Sector	Number of Stocks
10	Energy	36
15	Materials	26
20	Industrials	63
25	Consumer Discretionary	78
30	Consumer Staples	33
35	Health Care	50
40	Financials	87
45	Information Technology	61
50	Telecommunication Services	5
55	Utilities	29

TABLE 3: Basic information of CSI163 and S&P468 datasets including the number of stocks, the number of trading days, the average log returns $\langle r \rangle$, and the standard deviation σ_r of the log returns is presented. Compared to the US market, the Chinese market has higher fluctuations in our study period between 04/01/2007 and 06/11/2015.

Dataset	Stocks	Days	$\langle r \rangle$	σ_r	$\langle r_{\min} \rangle$	$\langle r_{\max} \rangle$
CSI163	163	2149	1.4795e-04	0.0340	-0.4832	0.1002
S&P468	468	2228	1.4691e-04	0.0252	-0.3413	0.2168

can extract subsets of prices in a serial of sliding windows: $[1, L]$, $[2, L + 1]$, \dots . For a given window, with the two time

series of prices, it is possible to generate two log return time series using equation (1) for both stocks s_i and s_j . Thus, the correlation coefficient between two stocks can be calculated by using the Pearson correlation coefficient [21]

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}}, \quad (2)$$

where $\langle \dots \rangle$ stands for the average. The value of ρ_{ij} ranges between -1 and 1, where a negative value of $\rho_{ij} < 0$ indicates the two stocks fluctuate in a noncorrelated manner, i.e., one falls down while another one climbs up. For a positive value of $\rho_{ij} > 0$, the two stocks fluctuate in a positively correlated way. In this case, they move in the same direction. If $\rho_{ij} \approx 0$, then they are not correlated. If $|\rho_{ij}| \approx 1$, then the two stocks are perfectly correlated or noncorrelated. In a stock market, the stocks from the same industry are more likely to be correlated.

For a portfolio of N stocks s_1, s_2, \dots, s_N , we can calculate all $N \times N$ pairs of correlation coefficients ρ_{ij} for any s_i and s_j . These N^2 pairs of values can be expressed as a correlation coefficient matrix C with a size of $N \times N$.

Based on the correlation matrix C , we can define the distance d_{ij} between stock pair of s_i and s_j as

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}. \quad (3)$$

The values of d_{ij} form an adjacent symmetric matrix D , in which there are $N(N - 1)/2$ different elements. It is verified that this definition satisfies the three rules of Euclidean distance: (1) $d_{ij} = 0$ if and only if $i = j$; (2) $d_{ij} = d_{ji}$;

(3) $d_{ij} \leq d_{ik} + d_{kj}$ [21]. Since $-1 \leq \rho_{ij} \leq 1$, we have $0 \leq d_{ij} \leq 2$. With this definition, the distance for two stocks has a value of 2 when they are completely anticorrelated ($\rho_{ij} = -1$), and a small distance close to 0 when they are positively and completely correlated ($\rho_{ij} \rightarrow 1$). This makes it possible to compare the distances for any two pairs of stocks.

3.2.2. Network $N(V, E)$. With the adjacency matrix D , we can further construct the network $N(V, E)$ for the stocks, where stock s_i is represented as vertex $v_i \in V$, and $e_{ij} \in V$ represents the edge between v_i and v_j with a distance of d_{ij} . The network $N(V, E)$ is undirected in which only one edge exists between a pair of vertices. The size of the network is the number of stocks N . The possible maximum number of edges is $N(N - 1)/2$ for an undirected complete network in which all vertex pairs are connected. For a portfolio with a large number of stocks N , the number of edges is a huge number. Thus it is necessary to simplify the network by filtering less important edges. In a simple threshold approach, by introducing a threshold value θ , we can reduce the network by chopping those edges whose distance are greater than θ and keeping the remaining edges. In other words, we only retain the connections which are strong enough, i.e., those with small distances, and all other weak edges with distances larger than the threshold θ are filtered as the following equation:

$$e_{ij} = \begin{cases} 1 & \text{if } d_{ij} < \theta, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Or in an edge ranking approach, we only keep a certain number of top edges with the strongest relationships, say $N-1$ edges. With this approach, the remaining edges are more likely to form loops in strongly connected vertices and are referred to as an *asset graph* [93].

3.2.3. Network Filtering. By filtering edges in a threshold approach, we may get isolated vertices or loops in the filtered network. To avoid this, tree approaches including *minimum spanning tree* (MST) can be used to chop edges but still keep all vertices connected as a tree. MST is introduced to investigate the hierarchical structure of stock networks first by Mantegna [21]. Many studies also use this approach, such as Jang et al., to investigate the foreign exchange market using in the periods of currency crises finding that the values of correlation coefficients decrease but the normalized tree length increase in crises [94]. Matteo et al. find that the dynamical *planar maximally filtered graphs* (PMFGs) can preserve same hierarchical structure as the dynamical MST, and the financial sector dominates the central role in the network [47]. As an application of network analysis in portfolio management, Onnela et al. suggest the assets of the classic Markowitz portfolio are always located on the outer leaves of the tree [88], and Pozzi et al. further suggest that even it is better to invest in the peripheries of the MST of a market [44]. In [85], MST networks extracted from real correlation data are compared with those generated from artificial random models. Results reveal that the properties of MST from real data cannot be reproduced, showing the uniqueness of real stock networks.

Based on the network $N(V, E)$, we can extract a tree connecting all vertices with $N-1$ edges with a minimum total distance also known as minimum spanning tree (MST) of the stocks. By only using the $N-1$ edges out of the maximum $N(N-1)/2$ edges, the network is dramatically simplified or filtered while keeping the most important shortest edges. To extract the MST, Kruskal's Algorithm was applied in three steps: (1) we rank all edges according to the distances from the shortest to the longest; (2) in each round, we choose the shortest edge into the MST while avoiding loops; (3) we repeat round #2 until all vertices and all $N-1$ shortest edges are added [95]. Bonanno et al. review the MST approach in revealing information of markets [96].

Using the MST, we can construct the *hierarchical tree* (HT) in which the subdominant ultrametric distance $d_{ij}^<$ is defined as the maximum distance of an edge along the path between v_i and v_j . The HT satisfies the first two rules with a stronger third one:

$$d_{ij}^< \leq \max(d_{ik}^<, d_{kj}^<), \quad (5)$$

With this ultrametric inequality, we can construct a hierarchical tree based on a MST and present a unique topological structure of the stocks [21, 97].

By loosening the requirements of MST up to 4 vertices, but forbid crossings, as many as $3(N-2)$ edges containing the MST as the subgraph including all the top $N-1$ shortest edges can be gathered. This new network can be drawn on a planar surface without link crossings is called *planar maximally filtered graphs* (PMFG) [47, 86, 98–101]. This makes PMFG different from MST, which also shows richer structures of the network. In a similar construction to MST, to construct PMFG, we firstly rank the edges in ascending orders according to the distances of edge pairs. Then we add the shortest edges into the PMFG but keeping the genus $g = k$, where the g is the largest number of simple closed curves one can draw on a planar surface without separating it. For the case of $g = 0$, when all edges are considered, PMFG can be gathered [86]. It has also been proved that an MST is a subgraph of a PMFG and the number of 3- and 4-cliques in a PMFG is $3n-8$ and $3n-4$, respectively [102].

Since PMFG contains more edges and allows loops and cliques, there is more information embedded in PMFG than in MST. After the introducing of PMFG into the study of network structures of stocks, PMFG has been used in studies of many stock markets, and more recently, PMFG is applied in investment strategy design [44]. Based on PMFG, a clustering approach called *Directed bubble hierarchical tree* (DBHT) is proposed and show good performance compared with other algorithms and also been applied to study financial data [103]. It has been reported that, in a running window approach, the PMFG shows stronger stability in a long run compared with MST [104].

4. Stock Network Topological Properties

A network $N = (V, E)$ is a graph composed of a set of vertices V and a set of edges E . In a network model, the participants are represented as the vertices V , and the

TABLE 4: For the CSII63 network, the maximum possible number of edges of $|e|_{\max}$ for $N = 163$ vertices, the existing edge number $|e|$, the edge density $|e|_{density}$, the average degree $\langle d \rangle$, the average distance $\langle d_{ij} \rangle$, the minimum distance d_{ij}^{\min} , and the maximum distance d_{ij}^{\max} are presented for different θ from 0 to 1.5 in a step of 0.1.

θ	$ e _{\max}$	$ e $	$ e _{density}$	$\langle d \rangle$	$\langle d_{ij} \rangle$	d_{ij}^{\min}	d_{ij}^{\max}
0.1	13203	0	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	13203	0	0.0000	0.0000	0.0000	0.0000	0.0000
0.3	13203	2	0.0002	0.0245	0.0000	0.0049	0.0089
0.4	13203	30	0.0023	0.3681	0.0000	0.0010	0.0593
0.5	13203	91	0.0069	1.1166	0.0003	0.0003	0.2200
0.6	13203	276	0.0209	3.3865	0.0010	0.0003	0.3689
0.7	13203	1594	0.1207	19.5583	0.0038	0.0003	0.4291
0.8	13203	5620	0.4257	68.9571	0.0177	0.0004	0.5353
0.9	13203	10555	0.7994	129.5092	0.0666	0.0004	0.7098
1	13203	12816	0.9707	157.2515	0.1866	0.0005	0.7996
1.1	13203	13194	0.9993	161.8896	0.4052	0.0006	0.9443
1.2	13203	13203	1.0000	162.0000	0.6957	0.0902	1.0887
1.3	13203	13203	1.0000	162.0000	0.9637	0.3719	1.1910
1.4	13203	13203	1.0000	162.0000	1.0966	0.5515	1.2878
1.5	13203	13203	1.0000	162.0000	1.1141	0.5515	1.3230

relationship between any pair of two participants i and j is represented as the edge e_{ij} connecting the two vertices v_i and v_j . In this study, the following properties of financial markets are researched: (1) *Degree and Degree Distribution* which describes the connectivities of vertices; (2) *Clustering Coefficient* which is the indication of the transitivity and density of a network; (3) *Average Path Length*, which is a global property indicating how the network spans; (4) *Betweenness Centrality* which describe the global importance or centrality of vertices or edges; (5) *Components* which describe the grouping phenomena of substructures of the networks.

Based on how the correlations are calculated, there are two approaches, static or dynamic. In a static approach, the correlations are calculated over the whole period using all available prices. Thus we get a single static correlation matrix to describe the market regardless of the different market periods. When sliding windows are used in a dynamic approach, we get a sequence of correlation matrices. The static approach, which is the most used in literature, gives a static description of the structure of the market with details of different market periods like bear markets or bull markets. However, the dynamic approach can reveal the evolution of market structures and behaviors, which are especially useful for the comparisons of calm periods and crashes.

In this part, we present the topological properties of stock networks of the two markets, CSII63 and S&P468, in a dynamic approach. Considered to meet the requirement of $L/N > 1$, we set the sliding window size $L_{CSII63} = 170$ for CSII63 and $L_{S\&P468} = 500$ for S&P468. In total, there are 2149 windows for CSII63 and 2228 windows for S&P468. After calculating the log returns for both CSII63 and S&P468 by using equation (1), we calculate the correlation coefficient matrices over the period between 12/09/2007 and 06/11/2015 for CSII63, 26/12/2008 and 06/11/2015 for S&P468 using equation (2). Based on the correlation matrices, it is

straight to get stock networks. For CSII63, we have a network of 163 vertices, and for the S&P468, we have a network of 468 vertices. The edge connecting two stocks indicates how the two stocks behave correlatively or anticorrelatively. For a positive correlation coefficient value, the two prices move in the same direction, while for a negative value, the two prices move in opposite directions, so to normalize all correlation coefficients to positive values as edge distances, we adopt the definition of distance based on equation (3). Through this definition, all negative values are transformed into positive distance values, and the order of values is preserved. All vertices in the networks for both markets are fixed. However, the edges vary in each sliding window as the correlation coefficients change. In the following parts, the statistical properties of both networks evolved in our study period are investigated.

4.1. Degree and Degree Distribution. For a network of N stocks, there are $N \times N$ edges, which is a huge number for a large N . So we normally filter the weakest edges to simplify the network. In the threshold approach, a threshold θ can be used to chop the edges, if $d_{ij} > \theta$. For a given network, different θ can lead to different structures with same vertices but different sets of remaining edges. Based on the correlation matrices, we first investigate the stock networks with different θ for both CSII63 and S&P468. In the sliding window approach, using daily log return time series, we first calculate the correlation matrices of 163×163 for CSII63 and 468×468 for S&P468; then we average all the correlation matrices over the study periods. After that, we get the averaged correlation matrices, with which we can apply the edge filtering process for different θ based on the equation (4). Based on the result, small θ closes to 0 will filter most edges while larger θ close to the maximum value two will keep most edges. We use an θ interval of $[0.1 - 1.5]$ with a step of 0.1. We present the basic network properties in Table 4 for

TABLE 5: For the S&P468 network, the max possible edges $|e|_{\max}$ for $N = 468$ vertices, the existing edge number $|e|$, the edge density $|e|_{density}$, the average degree $\langle d \rangle$, the average distance $\langle d_{ij} \rangle$, the minimum distance d_{ij}^{\min} , and the maximum distance d_{ij}^{\max} are presented for different θ from 0 to 1.5 in a step of 0.1.

θ	$ e _{\max}$	$ e $	$ e _{density}$	$\langle d \rangle$	$\langle d_{ij} \rangle$	d_{ij}^{\min}	d_{ij}^{\max}
0.1	109278	0	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	109278	2	0.0000	0.0085	0.0000	0.0935	0.1094
0.3	109278	2	0.0000	0.0085	0.0000	0.1959	0.2033
0.4	109278	18	0.0002	0.0769	0.0000	0.0005	0.2033
0.5	109278	137	0.0013	0.5855	0.0001	0.0003	0.3265
0.6	109278	729	0.0067	3.1154	0.0007	0.0003	0.4980
0.7	109278	3571	0.0327	15.2607	0.0038	0.0004	0.6125
0.8	109278	16433	0.1504	70.2265	0.0213	0.0004	0.7255
0.9	109278	50364	0.4609	215.2308	0.0952	0.0005	0.8122
1	109278	88680	0.8115	378.9744	0.2739	0.0005	0.9264
1.1	109278	106179	0.9716	453.7564	0.5356	0.0006	1.0334
1.2	109278	108956	0.9971	465.6239	0.7994	0.0007	1.1486
1.3	109278	109277	1.0000	466.9957	0.9906	0.0044	1.2480
1.4	109278	109278	1.0000	467.0000	1.0711	0.1959	1.3465
1.5	109278	109278	1.0000	467.0000	1.0711	0.1959	1.3465

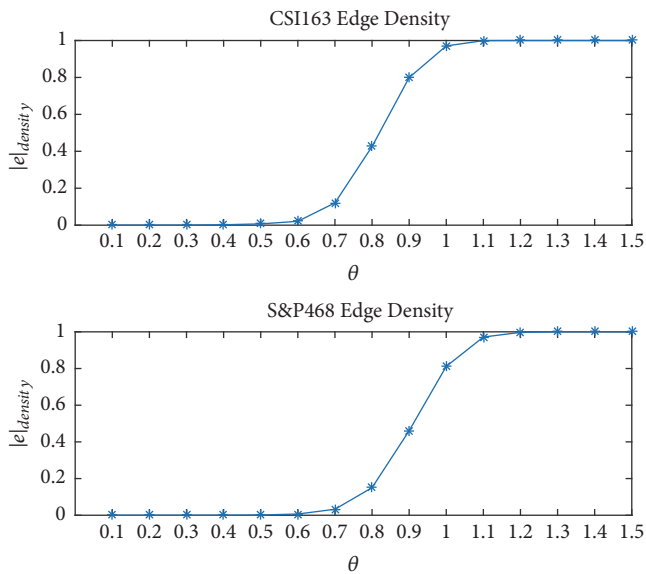


FIGURE 2: Edge densities of CSII63 and S&P468 for different thresholds θ from 0.1 to 1.5. It shows that the densities increase sharply from 0 to 1 in the θ interval of 0.6 and 1.

the CSII63 network and Table 5 with different θ between 0.1 and 1.5. The maximum possible edges are 13203 for CSII63 and 109278 for S&P468, respectively. For different distance thresholds θ , any edges whose distances are greater than the threshold are filtered. So with smaller θ , only a few edges remain in the network and this results a smaller edge density $|e|_{density}$, smaller average degree $\langle d \rangle$, average distance $\langle d_{ij} \rangle$, minimum distance d_{ij}^{\min} , and maximum distance d_{ij}^{\max} as well.

In Figure 2, we plot the edge densities of CSII63 and S&P468 for different thresholds θ from 0.1 to 1.5. In the

interval of 0.1 to 0.6, the densities for both networks are close to 0, meaning all edges are filtered. While in the interval of 1 to 1.5, the densities are close to 1, meaning that all edges are preserved. Between these two intervals, we see that the two curves have a similar shape with a slope when θ lies between 0.6 and 1. This indicates that most edges are within this interval. A similar edge density distribution is also reported in [65]. The study shows that stock networks also demonstrate a similar transforming interval.

We investigate the degree distributions of both networks with different θ . No matter if θ is too small or too large, the degree distributions are noisy, while in a narrow interval around 0.7, the distributions follow the power law. The regression fitting curve is a straight line in the plots of $\log P(k)$ against $\log(k)$, where the $\log P(k)$ is the \log_{10} probability for a vertex with k degrees and the $\log(k)$ is the \log_{10} degree. After running on the data, we plot the typical power law distributions in Figure 3(a) for CSII63 and Figure 3(b) for S&P468 respectively. For both distributions, we fit the $\log_{10} - \log_{10}$ distribution and get the power law exponents $\gamma = -0.9935$ for CSII63 and $\gamma = -1.2323$ for S&P468. In the plots, we use the same $bins = 20$ to calculate the probabilities for different degrees. As shown in Figure 3, we see that a large number of vertices have small degrees. Only a few vertices have large degrees. As the vertices are stocks, and the degrees are rooted in the correlations between stocks, for both CSII63 and S&P468 networks, only a few stocks are the highly correlated with the most parts of rest stocks. These stocks have a wider and larger influence over the whole networks, while other stocks with relatively smaller degrees are less correlated with other stocks. This presents limited influence over the network. The negative fitting slope value γ also indicates that both the CSII63 and S&P468 networks are scale-free networks in which a small portion of vertices have larger degrees while a large portion of

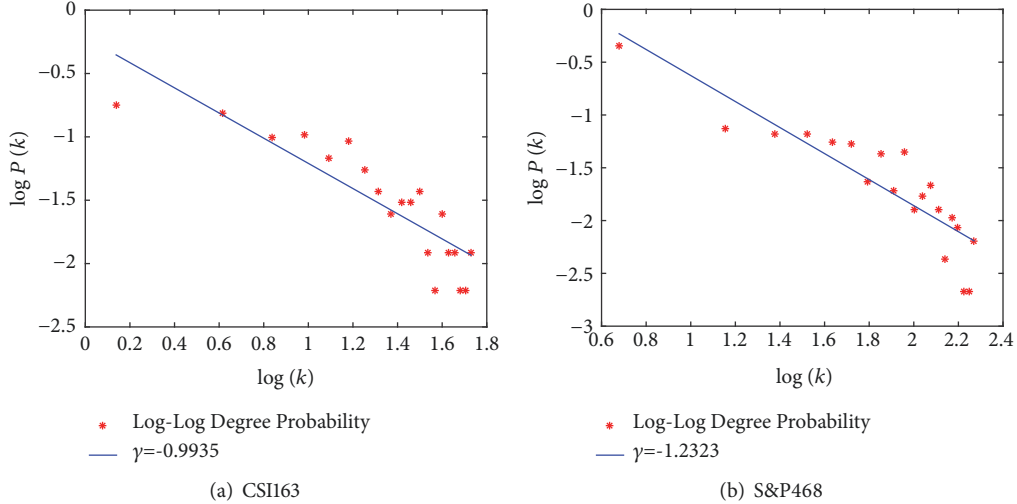


FIGURE 3: Log-Log degree distributions of CSI163 and S&P468 networks. By using different θ , we can filter out edges with larger distances. We find that not all the filtered networks can demonstrate the power law degree distributions. Only when θ falls within a narrow interval around 0.7, the filtered networks follow the power law degree distribution. In Figure 3(a), for the CSI163 network, we use $\theta = 0.68$ and we get a fitting line with a slope of $\gamma = -0.9935$. In Figure 3(b), for the S&P468 network, we use $\theta = 0.75$ and we get a fitting line with a slope of $\gamma = -1.2323$. This indicates that the degree distributions of both stock markets follow the power law in the form of $P(k) \sim k^{-\gamma}$, which also means the networks are scale free in which a small portion of vertices have larger degrees, while a large portion of vertices have small degrees.

vertices have smaller degrees this agree with previous studies [105].

4.2. Average Clustering Coefficient. Average clustering coefficient $\langle C \rangle$ is an average of all clustering coefficients $\langle C_i \rangle$ of all vertices. The clustering coefficient $\langle C_i \rangle$ indicates the transitivity for an individual vertex v_i , while the overall averaged clustering coefficient $\langle C \rangle$ is an indication of the transitivity and density of the whole network. In Figure 4, we present the average clustering coefficient $\langle C \rangle$ for both CSI163 and S&P468 networks comparing with random networks. The $\langle C \rangle$ gets larger with the θ when larger θ will preserve more edges, while it remains almost unchanged with a slight increase in both random networks. Comparing with random networks of same sizes of 163×163 for CSI163 and 468×268 for S&P468, $\langle C \rangle$ of both two stock networks are significantly larger than that of the corresponding random networks. For CSI163, $\langle C \rangle$ is 4.9574 times larger than that of the random networks on average with a maximum of 11.8903 times. For S&P468 the average multiple is 5.2305 times, and maximum multiple is 10.4180 times compared with the random networks. This shows that both stock networks are well connected with better transitivity. This result agrees with many other previous studies.

4.3. Average Path Length. Unlike clustering coefficient which is a local property, for any two vertices v_i and v_j in a network, the number of edges covering the shortest route linking the two vertices is defined as the characteristic path length, l_{ij} , in [2] which is a typical global property. By averaging the lengths of all possible pairs, we can calculate the average path length $\langle L \rangle$. As an indication of how the network is connected, many real networks have small $\langle L \rangle$ compared with random networks. In Figure 5, we plot the average path length $\langle L \rangle$

of both CSI163 and S&P468 networks with comparisons of random networks in same sizes. The two stock networks are significantly different from the corresponding random networks with the same sizes of 163×163 and 468×468 . While the flat curves of $\langle L \rangle$ of random networks stay almost unchanged with θ , this is a result of the universal homogeneous edge distribution on the whole network. There are peaks for the stock networks. On the left hand of the peak, there is a decline of $\langle L \rangle$ with the decrease of θ , since when θ gets too small, most edges are filtered, and the giant networks break into small parts and $\langle L \rangle$ in small parts are decreasing significantly. However, for the right hand of the peak, $\langle L \rangle$ gets smaller with the increase of θ due to the increasing connectivity when more and more edges are preserved. This shows the stock networks of both CSI163 and S&P468 are different from random networks.

4.4. Betweenness Centrality. The betweenness b_{v_i} of vertex v_i is defined as the number of shortest paths passing v_i , which is an indication of the importance of an individual vertex in the contribution to the global connectivity. By averaging over the betweenness of all vertices, we can compare the betweenness for any two vertices. Larger betweenness means great global influence of the stock networks. This is the same to the edges. The betweenness $b_{e_{ij}}$ is the number of shortest paths passing the edge e_{ij} indicating the importance of this edge for its contribution to the global connectivity. In this study, we focus on the vertex betweenness. In the calculation of the shortest paths, we can use the original distance d_{ij} defined in equation (3) or simplify the network as a binary network according to

$$e_{ij} = \begin{cases} 1 & d_{ij} > 0, \\ 0 & d_{ij} = 0. \end{cases} \quad (6)$$

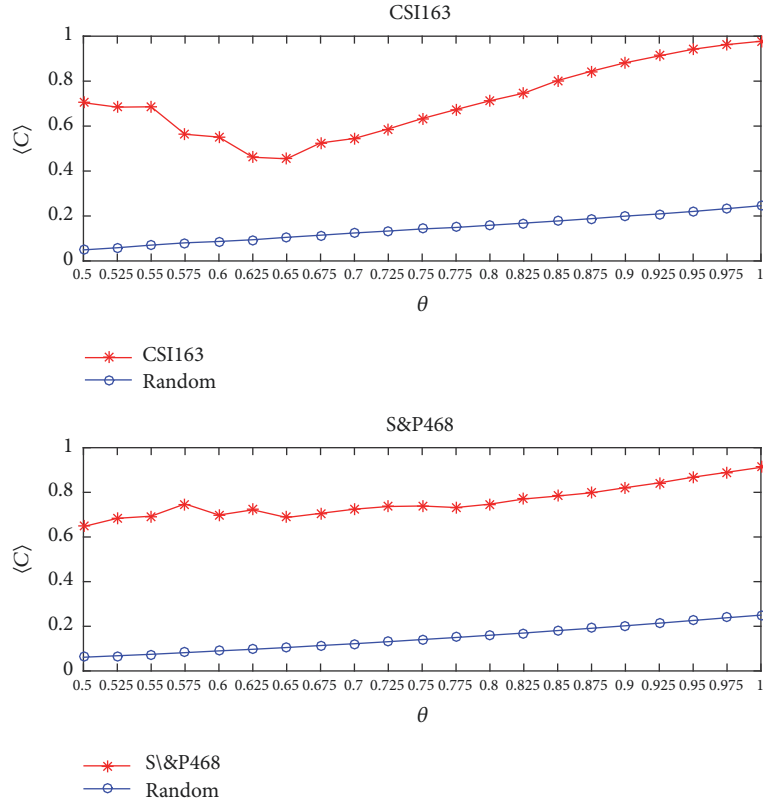


FIGURE 4: Average clustering coefficient $\langle C \rangle$ of CSI163 and S&P468 for different thresholds θ . It shows that $\langle C \rangle$ gets larger with θ . To compare with random networks, we plot the corresponding average clustering coefficients under the same interval of θ . As is shown, for both CSI163 and S&P468 networks, $\langle C \rangle$ values are significantly larger than the random networks of the same size. This indicates the stock markets are far from random and the stocks are comparatively clustered.

In the former, edges with different distances have different contributions to the paths, while, for the latter, all edges of nonzero distance are normalized as 1 and treated equally with great scarifying of original distance information. As shown in Figure 6, we plot the average betweenness $\langle B \rangle$ of CSI163 and S&P468 for both binary case and weighted case under different θ in Figures 6(a) and 6(b) respectively. For binary network case, all edges with positive distances are normalized as unit 1, while, in weighted networks, the original distances are directly used in the calculation of shortest paths. The shapes for binary and weighted networks are different. There are peaks for both stock networks in binary networks, while $\langle B \rangle$ gets larger with θ from almost zero to large numbers in weighted networks.

We visualize the stock networks of CSI163 and S&P468 with different θ of 0.6, 0.7, 0.8, 0.9 in Figures 7(a)–7(d) and 8(a)–8(d), respectively. It shows that the networks can be dramatically simplified using small values of θ and the edges are preserved in larger values of θ . As listed in Table 4, the edge density of CSI163 grows dramatically from 0.0209 ($\theta = 0.6$) to 0.7994 ($\theta = 0.9$), while, for S&P468 as shown in Table 5, the edge density of S&P468 grows also dramatically from 0.0067 ($\theta = 0.6$) to 0.4609 ($\theta = 0.9$). All networks in this paper are generated using the *Pajek* complex network software [106].

4.5. Components. A component *Com* is a subnetwork of the whole network with connected vertices. For a given network with a set of N vertices, the possible size of *Com* can range from 1 for an isolated vertex to N for all connected vertices. When an individual vertex v_i is disconnected from any other vertices, v_i itself forms the smallest component with a single vertex. When all the vertices are connected without any isolated vertices, the network is a single giant component. For a stock network, the stocks are correlated with each other, while stocks belonging to different components are not correlated. The component structures of stock networks have great implications for risk management of a portfolio. Since the stocks fall in the same component are correlated, so it is a bad idea to invest in most stocks from the same components. We should invest stocks from different components to diverse the risk of the whole portfolio. When θ is small, most edges are filtered leaving many vertices isolated. As a result, we see the emerging of a large number of small components. However, with the growth of θ , more and more edges are preserved. This allows the connectivity increase resulting in the appearing of larger components. In Figure 9, the properties of components of CSI163 and S&P468 networks with different θ is presented. For the two networks, the number of components $N_{components}$ (red), the max component size S_{max} (green), and the average component size $\langle S \rangle$ (blue)

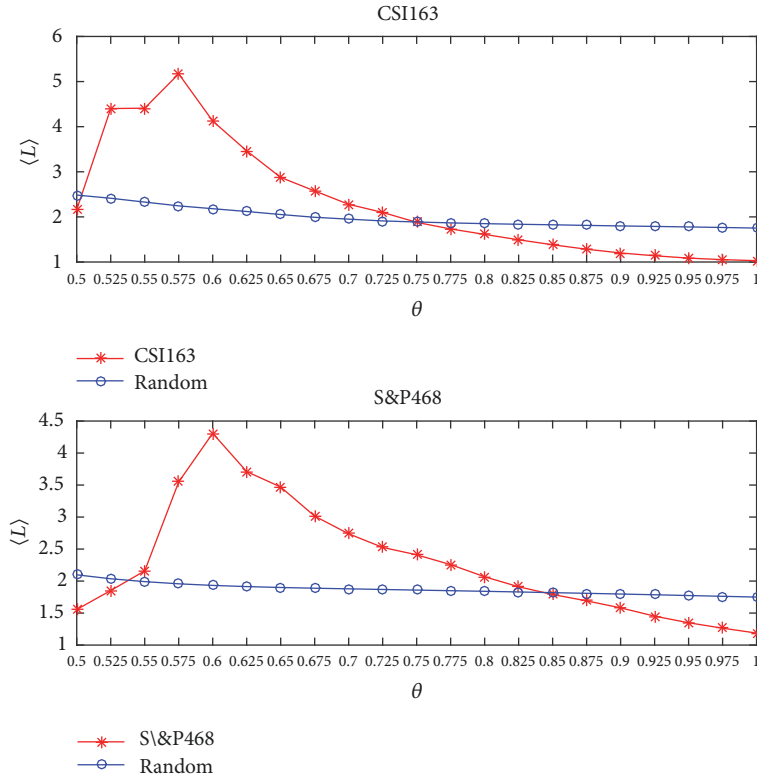


FIGURE 5: Average path lengths $\langle L \rangle$ for CSI163 and S&P468 under different θ compared with values for random networks in same sizes of 163×163 and 468×468 . It shows that for the two networks, there are peaks of $\langle L \rangle$ above the curve of the corresponding random networks. At first, θ is small, most edges are filtered, and the whole networks are broken into parts; thus the disconnected vertices and edges are also filtered. So on the left hand, starting from the peak, with the decrease of θ , we see that $\langle L \rangle$ decreases too, while starting from the peak, with the increase of θ , we see a constant decline of $\langle L \rangle$, for more edges remained resulting in decreasing of $\langle L \rangle$. For the random networks, $\langle L \rangle$ stay almost unchanged because of the homogeneous edge distributions across the whole networks.

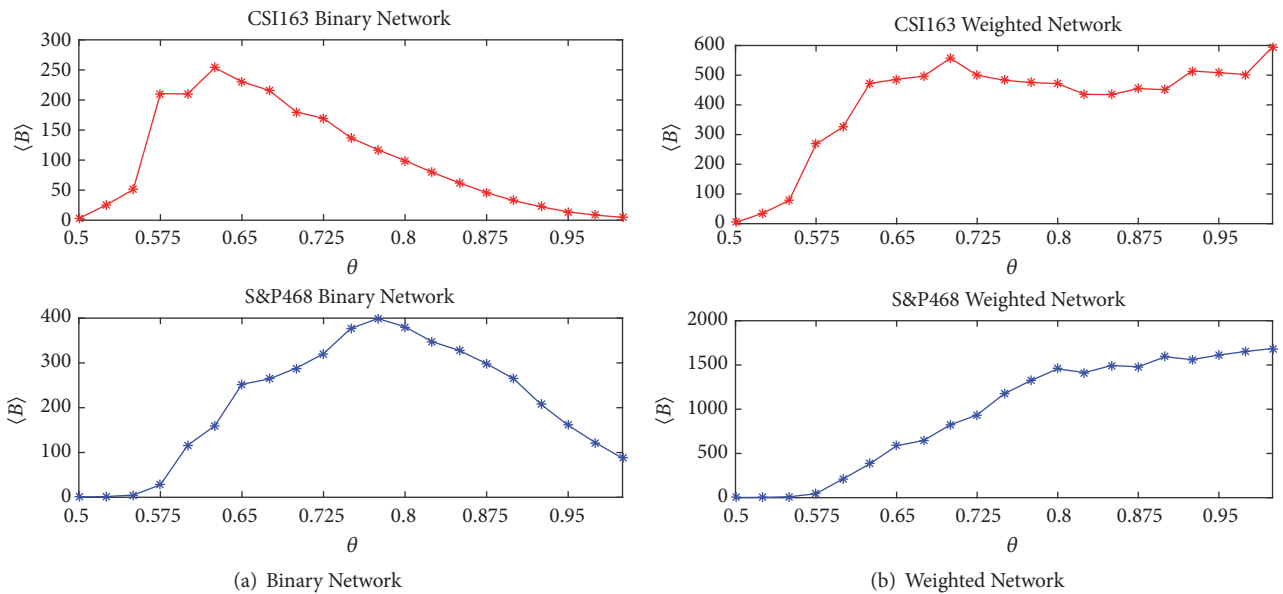


FIGURE 6: Average betweenness $\langle B \rangle$ of CSI163 and S&P468 is calculated for both the original weighted approach and binary simplification approach under different θ . The average betweenness for binary networks is different from the weighted network. For binary networks, the curves for CSI163 and S&P468 share a similar shape. On the left hand of the peak, the $\langle B \rangle$ gets larger with the increase of θ , for more edges and more vertices are preserved, and this leads to a growing number of paths, while on the right hand of the peak, large connected network emerges leading to a small value of averaged $\langle B \rangle$. In other words, the importance of a single individual vertex or edge is weakened in well-connected networks (large θ).

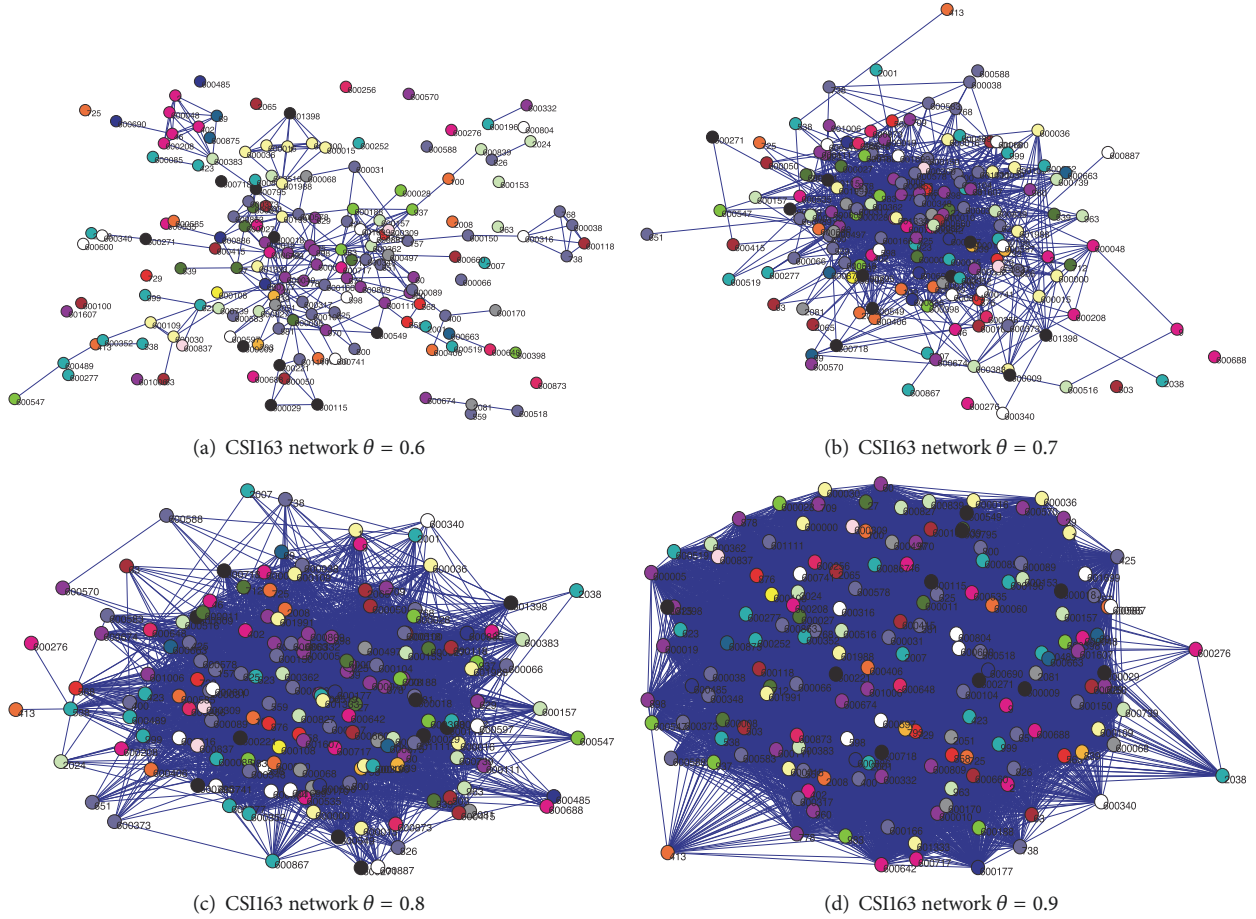


FIGURE 7: CSII63 networks with different θ of 0.6 (a), 0.7 (b), 0.8 (c), 0.9 (d). It shows that the network is relatively sparser with small θ while denser with large θ , for small θ greatly simplifies the network by filtering most edges with larger distance. As a result, the edge density when $\theta = 0.9$ is about 38.25 times to that when $\theta = 0.6$. Different vertex colors indicate different industry sectors.

shows similar pattern and changes with the values of θ . Critical changes are obvious for both networks in the θ interval about $[0.3 - 0.8]$ for CSII63 and in $[0.5 - 1.1]$ for S&P468, respectively. Before the transition interval, most vertices are isolated when the whole network breaks into small components, and both of the maximum and average component sizes are small. In the transition interval, the number of components decreases with both maximum and average component sizes. After the transition interval, the three properties stay unchanged when the giant connected component appear with maximum and average size equal to the number of total vertices. The similar component properties transition under different of θ phenomena is also observed in the study of a set of Chinese stocks [65] with a reported transition critical value about $\theta = 0.17$.

To investigate how industry sectors are connected in the stock network, we summarize the properties of both CSII63 and S&P468 networks with $\theta = 1.0$ listed in Tables 6 and 7, respectively. As it shows, in the CSII63 network, the industry sectors are all most the same in average degree $\langle d \rangle$ and average clustering coefficient $\langle C \rangle$, while with significantly different values of average betweenness coefficient $\langle B \rangle$. The

difference of average degree $\langle d \rangle$ and average clustering coefficient $\langle C \rangle$ are not significant among the industry sectors. This indicates that all sectors have similar degrees and clustering coefficients. The difference between the average betweenness coefficient $\langle B \rangle$ shows that the sectors contribute to the global connectivity differently. It is worth mentioning that the finance and insurance sector has the largest average clustering coefficient $\langle C \rangle$ of 0.9829 but with a relatively small value of average betweenness coefficient $\langle B \rangle$ which is only 118.4000. For the S&P468 network, as shown in Table 7, we observe that Financials sector has the largest value of $\langle d \rangle$ of 421.7471 and the 3rd largest value of $\langle B \rangle$ of 1297.8161, with a smaller value of the average clustering coefficient $\langle C \rangle$ of 0.8975, which are very different from the CSII63 network. Furthermore, the Energy and Industrials have the largest values of $\langle d \rangle$ and $\langle B \rangle$, while Consumer Staples and Telecommunication Services have the smallest $\langle d \rangle$ and $\langle B \rangle$. From this, we also observe that, for the S&P468 network, the sectors with larger $\langle B \rangle$ are likely to have smaller values of $\langle C \rangle$ and vice versa. The findings indicate that the US market is dominated by Financials while the finance and insurance in Chinese stock markets play relatively less influential roles.

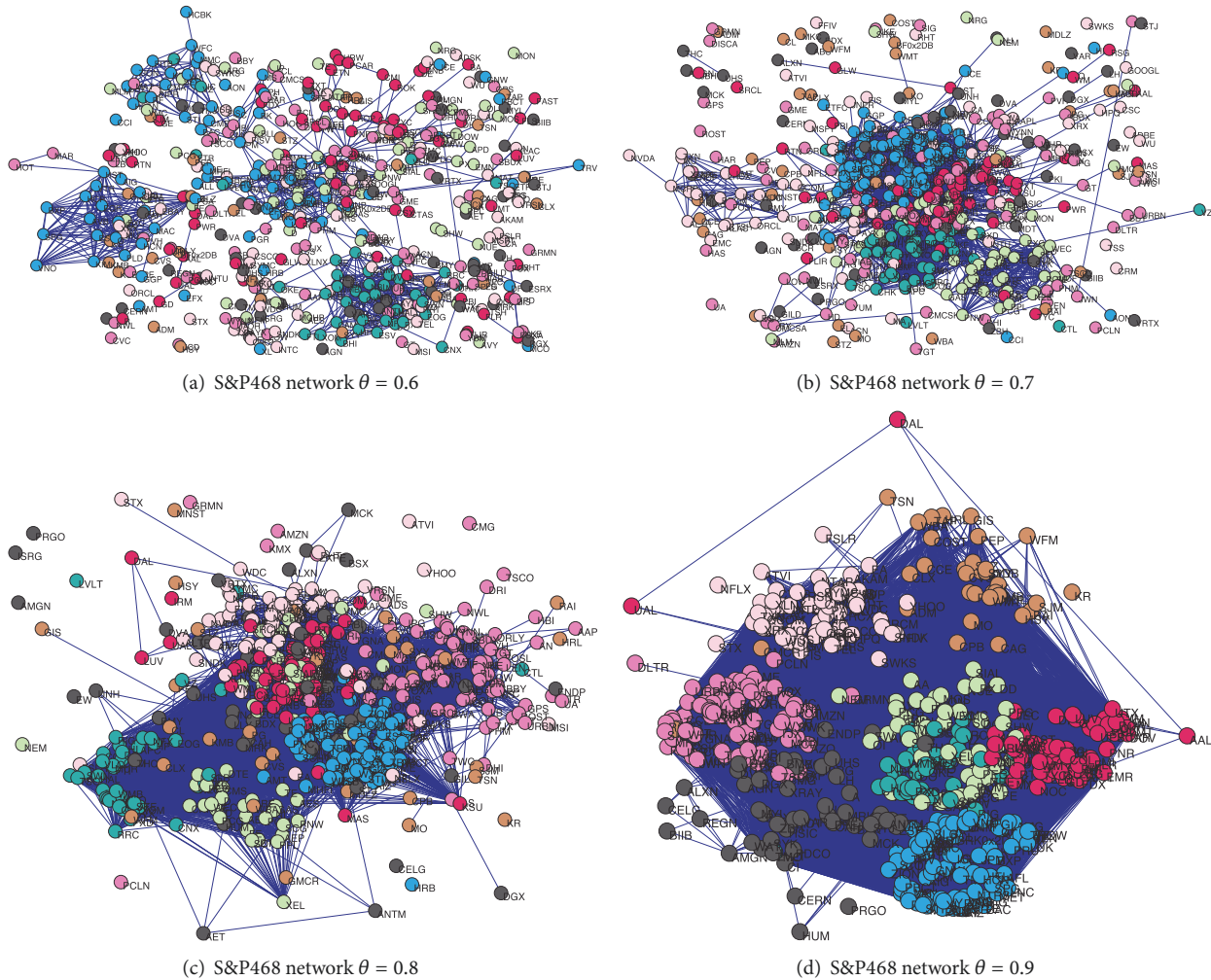


FIGURE 8: S&P468 networks with different θ of 0.6 (a), 0.7 (b), 0.8 (c), and 0.9 (d). It also shows that the edge densities with different θ change dramatically. We find that the edge density when $\theta = 0.9$ is 68.79 times to that when $\theta = 0.6$. Different vertex colors indicate different industry sectors.

By focusing on the top stocks, it is possible to look into the details of the networks. In Table 8, for the CSI163 network with $\theta = 0.8$, we present the top 10 stocks with the largest values of degree d_i and betweenness b_i ranked in descending order in the upper part and the lower part, respectively. The stock code, company name, industry name, and values of d_i , c_i , and b_i are listed. Younger Group, which is a leading fashion brand in China, has the largest degree of $d_i = 133$, and HuDong Heavy Machinery, which is a major machinery manufacturer in China, has the largest betweenness coefficient of $b_i = 5444$. While, in the S&P468 network, as shown in Table 9, T. Rowe Price Group has the largest degree of 266, and Loews Corp. has the largest betweenness coefficient value of 13202, both stocks are in the Financials sector. For both stock networks, the two lists based on d_i and b_i are similar. In other words, top stocks with largest degree values also appear as top stocks with largest betweenness coefficients b_i . It is worth noting that the list based on the ranking of clustering coefficients c_i are dramatically different those based on degrees or betweenness

coefficients. This indicates that degree d_i and the betweenness b_i are consistent in describing the importance of an individual vertex, since the higher degree a vertex has, the more likely it is on the shortest paths. As indicated in the two tables, the stocks with codes labeled in bold appear on both top 10 lists, and in fact, the rest of the stocks on one list also can be found appearing in a similar ranking position on another list. We can also observe that stocks belong to Industries of Metals & Nonmetals, Machinery, and Pharmaceuticals are dominant the two top 10 lists in the CSI163 network. However, for S&P468 network, Financials, Industrials, and Materials are major stocks in the two lists. As an emerging market, Industrials sector has great influence in Chinese stock market, while the Financials sector has greater influence in US stock market which agrees with [10]. The significant difference between the two stock markets confirms the previous studies with similar results indicating that Industrials is the most influential sector among all industry sectors, while the financial sector has weaker influence [107]. This is consistent with our previous results revealed Tables 6 and 7.

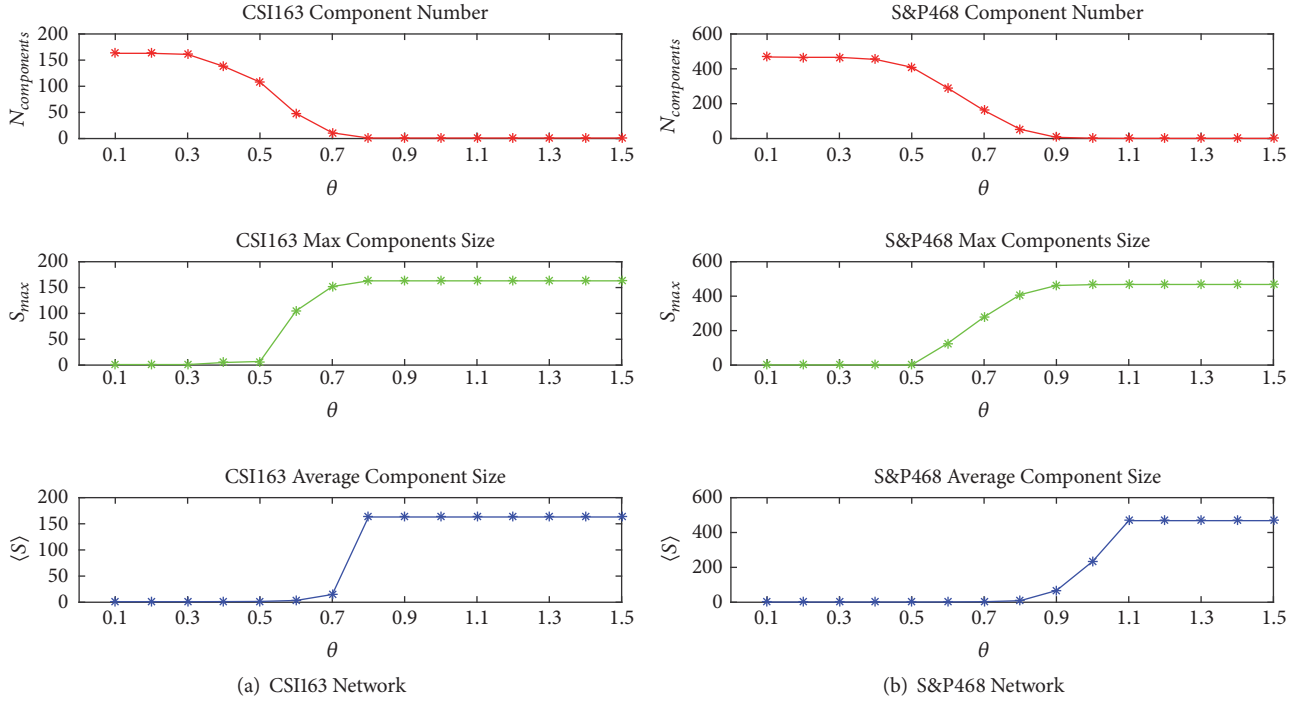


FIGURE 9: The component properties of the components number $N_{components}$ (red), the max component size S_{max} (green), and the average component size $\langle S \rangle$ (blue) are plotted for CSI163 and S&P468 with different thresholds θ in (a) and (b), respectively. Both networks show similar transitions when the networks transform from a large number of small isolated components into a connected giant network. Before the transition interval, edges are filtered leaving isolated vertices are not correlated. After the transition interval, edges are preserved making most vertices connected to form a single giant network in which all vertices are correlated.

TABLE 6: In this table, we list the China Securities Regulatory Commission (CSRC) industry code, the sector name and the numbers of stocks, the average degree $\langle d \rangle$, the average clustering coefficient $\langle C \rangle$, and the average betweenness coefficient $\langle B \rangle$ for each industry sector of these 163 stocks. The values are calculated from the CSI163 network with $\theta = 1.0$.

Code	Industry Sector	Number	$\langle d \rangle$	$\langle C \rangle$	$\langle B \rangle$
A	Agriculture	1	162.0000	0.9703	754.0000
B	Mining	6	154.3333	0.9803	974.6667
C0	Food & Beverage	4	160.2500	0.9738	770.0000
C1	Textiles & Apparel	4	159.2500	0.9764	437.5000
C3	Paper & Printing	2	159.0000	0.9757	343.0000
C4	Petrochemicals	9	157.4444	0.9776	417.5556
C5	Electronics	7	157.1429	0.9783	176.8571
C6	Metals & Non-metals	20	159.9500	0.9756	808.0000
C7	Machinery	27	157.5926	0.9772	664.7407
C8	Pharmaceuticals	15	151.6000	0.9764	324.2667
D	Utilities	6	159.6667	0.9770	199.6667
E	Construction	5	160.2000	0.9768	945.2000
F	Transportation	10	157.9000	0.9792	426.2000
G	IT	8	156.3750	0.9775	444.0000
H	Wholesale & retail trade	10	158.8000	0.9775	341.2000
I	Finance and insurance	10	155.8000	0.9829	118.4000
J	Real estate	11	153.2727	0.9791	302.3636
K	Social Services	3	161.0000	0.9726	109.3333
L	Communication	2	160.0000	0.9775	821.0000
M	Comprehensive	3	159.6667	0.9777	582.6667

TABLE 7: In this table, we list the industry code, the sector name, the numbers of stocks, the average degree $\langle d \rangle$, the average clustering coefficient $\langle C \rangle$, and the average betweenness coefficient $\langle B \rangle$ for each industry sector of S&P468 stocks. The values are calculated for the S&P468 network with $\theta = 1.0$.

Code	Industry Sector	Number	$\langle d \rangle$	$\langle C \rangle$	$\langle B \rangle$
10	Energy	36	419.6944	0.8948	1264.6111
15	Materials	26	404.2692	0.8948	1717.8462
20	Industrials	63	415.8571	0.8894	1937.0476
25	Consumer Discretionary	78	376.0256	0.9194	786.8205
30	Consumer Staples	33	269.0303	0.9194	37.8182
35	Health Care	50	309.6000	0.9135	213.8000
40	Financials	87	421.7471	0.8975	1297.8161
45	Information Technology	61	373.9344	0.9245	363.6721
50	Telecommunication	5	293.2000	0.9456	124.8000
55	Utilities	29	375.3448	0.9142	579.2414

TABLE 8: Top stocks with highest values of degree d_i , and betweenness b_i ranked in descending order of d_i and b_i when the $\theta = 0.8$ for CSI163 network. Stock codes in bold indicate the stocks appear at both top 10 stocks.

Code	Name	Industry	$d_i \downarrow$	c_i	b_i
600177	Youngor Group	Textiles & Apparel	133	0.5416	2154
600642	Shanghai Wanye Enterprises	Real estate	131	0.5732	2362
39	China International Marine Containers (Group)	Metals & Non-metals	127	0.5929	2276
600010	Inner Mongolia Baotou Steel Union	Metals & Non-metals	125	0.5738	1112
600166	Beiqi Foton Motor	Machinery	123	0.6047	72
825	Shanxi Taigang Stainless Steel	Metals & Non-metals	122	0.6005	2644
623	Jilin Aodong Medicine Industry (Groups)	Pharmaceuticals	121	0.6058	60
600362	Beijinghualian Hypermarket	Wholesale and retail trade	121	0.6127	1552
600717	Qinhuangdao Yaohua Glass	Real estate	119	0.6095	1354
600005	Wuhan Iron And Steel	Metals & Non-metals	118	0.6009	532
Code	Name	Industry	d_i	c_i	$b_i \downarrow$
600150	HuDong Heavy Machinery	Machinery	110	0.6155	5444
898	Angang Steel	Metals & Non-metals	114	0.6367	4560
600398	Anyuan Industrial	Mining	95	0.6710	4140
2051	China CAMC Engineering	Construction	114	0.6226	4136
601607	Aluminum Corporation of China Limited	Metals & Non-metals	116	0.6304	2766
825	Shanxi Taigang Stainless Steel	Metals & Non-metals	122	0.6005	2644
600031	Sany Heavy Industry	Machinery	85	0.7300	2584
600642	Shanghai Wanye Enterprises	Real estate	131	0.5732	2362
39	China International Marine Containers (Group)	Metals & Non-metals	127	0.5929	2276
600177	Youngor Group	Textiles & Apparel	133	0.5416	2154

5. Hierarchical Structures of Stock Networks

Mantegna introduced the minimum spanning tree and hierarchical clustering methods into the study of financial networks [21], in which a distance matrix D is built from the correlation matrix for all stocks. Based on the distance matrix, the minimum spanning tree is extracted. Since the minimum spanning tree contains the information of edges connecting all vertices in a single spanning tree with the shortest total length, it is also possible to extract the hierarchical clustering tree from the minimum spanning tree, where the distance for vertex v_i and v_j is subdominant *ultrametric* distance $d^{\mathcal{L}}(i, j)$ as the maximum value of distance along the shortest

path between the two vertices v_i and v_j [108]. However, the subdominant ultrametric distance approach will lose much edge distance information, for two separated vertices which are indirectly connected on the minimum spanning tree with a specific larger subdominant ultrametric distance might be directly linked in the original distance matrix. Vertices which should be clustered together might be separated in a hierarchical clustering tree based on ultrametric distance. To preserve the hierarchical structure of the minimum spanning tree as well as more information allowing loops and cliques, Planar Maximally Filtered Graph (PMFG) is proposed in [86]. Based on PMFG, the influence of different sectors of CSI300 is studied revealing that the industrial sector is

TABLE 9: Top stocks with highest values of degree d_i , and betweenness b_i ranked in descending order of d_i and b_i when the $\theta = 0.8$ for S&P468 network. Stock codes in bold indicate the stocks appear at both top 10 stocks.

Code	Name	Industry	$d_i \downarrow$	c_i	b_i
TROW	T. Rowe Price Group	Financials	266	0.3882	4536
L	Loews Corp.	Financials	263	0.4069	13202
SNA	Snap-On Inc.	Consumer Discretionary	263	0.3894	8350
IVZ	Invesco Ltd.	Financials	261	0.4016	2868
BEN	Franklin Resources	Financials	259	0.4074	7836
HON	Honeywell Int'l Inc.	Industrials	256	0.4151	1790
AMG	Affiliated Managers Group Inc	Financials	255	0.4128	2042
DD	Du Pont (E.I.)	Materials	255	0.4163	4966
SIAL	Sigma-Aldrich	Materials	254	0.4186	7020
ROP	Roper Industries	Industrials	247	0.4326	8626
Code	Name	Industry	d_i	c_i	$b_i \downarrow$
L	Loews Corp.	Financials	263	0.4069	13202
HST	Host Hotels & Resorts	Financials	223	0.4938	12510
OKE	ONEOK	Energy	174	0.5329	9978
ROP	Roper Industries	Industrials	247	0.4326	8626
SNA	Snap-On Inc.	Consumer Discretionary	263	0.3894	8350
BEN	Franklin Resources	Financials	259	0.4074	7836
FLS	Flowserve Corporation	Industrials	202	0.5271	7726
JPM	JPMorgan Chase & Co.	Financials	124	0.7409	7614
UTX	United Technologies	Industrials	217	0.4746	7590
SIAL	Sigma-Aldrich	Materials	254	0.4186	7020

the dominant part of the whole market [107]. In [109], the hierarchical tree structure of multiple industry indices in China are investigated before and after a crisis showing the structure changes around the crisis period. A similar study of global financial crisis impact on stock market shows that the Turkish market is less influenced [110]. Authors propose to use *Kullback-Leibler* distance for the filtering procedures in [11]. International real estate market networks in different countries are studied in [99] revealing that markets are clustered according to geographical locations. Instead of using the methods of [21], a typical approach is applied to extract the hierarchical structure of the German stock market in [68]. Using the industry classification information as the benchmark, authors compared the methods used to extract the clusters in financial networks [103]. In [111], *Neighbor-Net* approach is applied in which more distance information is used in the construction of the tree compared to the hierarchical clustering.

Since a sliding window approach with a window size of L is utilized, in a study period of total P trading days, we can get a sequence of $P - L + 1$ trading windows. We have $P_{csi163} = 2149$ for CSI163 and $P_{s\&p468} = 2228$ for S&P468 trading dates in our study period between 04/01/2007 to 06/11/2015, respectively. As we adapted in previous parts, we set the sliding window size as $L_{csi163} = 170$ for CSI163 and $L_{S\&P468} = 500$ for S&P468. We have $W_{csi163} = 1980$ windows for CSI163 and $W_{S\&P468} = 1729$ for S&P468 respectively. For each sliding window at time t , we can get the distance matrices $D_{csi163}(t)$ and $D_{S\&P468}(t)$ where $t = 1, \dots, W$. To investigate the structures of the two markets taking the whole

study period as a whole, we calculate the averaged distance matrix by averaging all elements overall sliding windows as

$$\langle D \rangle = \frac{1}{W} \sum_t D_t. \quad (7)$$

With this averaged distance matrix, we construct the hierarchical trees, minimum spanning trees, and asset graphs and study the evolvement of the properties of minimum spanning trees and asset graphs for both CSI163 and S&P468.

5.1. Hierarchical Tree. In the study of the stock market or a portfolio, a set of individual stocks belonging to different economic sectors behavior correlated together. Based on the prices information, the correlation matrix can be formed. Based on that, a distance matrix can be derived. Using the distance matrix, clustering algorithms can be further applied to extract the clustering structures of the stocks. For the stocks falling in the same cluster, they behave similar sharing correlated price fluctuations, while, for the stocks coming from different clusters, they are less similar than the ones of the same clusters. The main objective of clustering algorithms is to minimize the dissimilarity for stocks in the same cluster and maximize the dissimilarity for stocks in different clusters meanwhile. Since the dissimilarity is naturally measured by the distance, the selection of definition of distance between clusters is important for clustering algorithms. Four distance definitions as shown in equations (8)-(11) are used in extracting of hierarchical clustering trees. The distance between two clusters, c_m and c_n , is defined as the minimum distance for

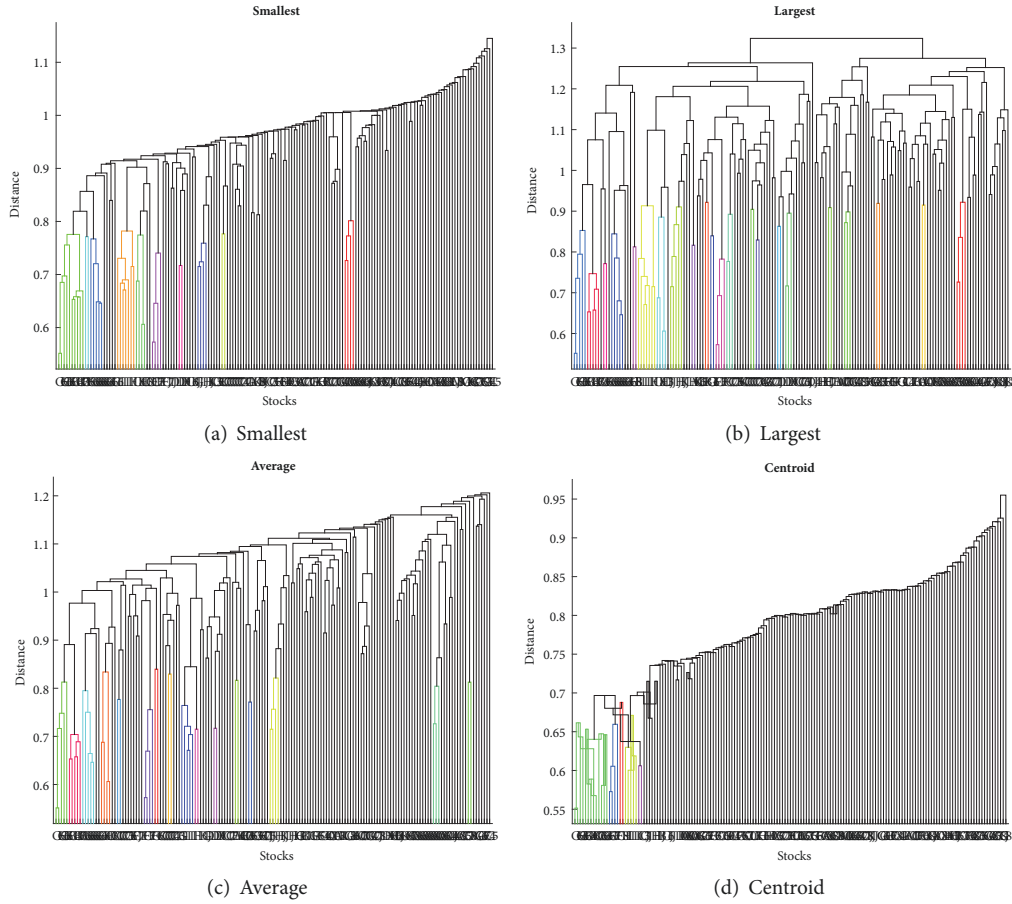


FIGURE 10: CSI163 dendrogram hierarchical cluster trees extracted with different distance definitions in (a) smallest distance for stock pairs, equation (8); (b) largest distance for stock pairs, equation (9); (c) average distance for stock pairs, equation (10); (d) distance between centroids for clusters, equation (11). The color threshold is 0.7. All stocks whose linkage values are less than this threshold would be colored with a unique color. As shown in the figures, different distance definitions extract different dendrogram hierarchical cluster trees whereas the same color threshold generates different results. We see that the largest distance definition reveals more details of network.

all pairs as in equation (8), the maximum distance for all pairs as in equation (9), the average distance for all pairs as in equation (10), and the distance between average centroids of the two clusters as in equation (11), respectively.

$$d_{m,n} = \min(d_{o_m, o_n^i}) \quad (8)$$

$$d_{m,n} = \max(d_{o_m, o_n^i}) \quad (9)$$

$$d_{m,n} = \frac{1}{N_m N_n} \sum_i \sum_j d_{o_m^i, o_n^j} \quad (10)$$

$$d_{m,n} = d(\bar{o}_m, \bar{o}_n) \quad (11)$$

In our study, we use all these four definitions of cluster distance. For CSI163 network, we present the dendrogram hierarchical cluster trees in Figures 10(a)–10(d). For S&P468 network, we present the trees in Figures 11(a)–11(d). In these trees, the leaf nodes are individual stocks, and the height of two merged branches indicates the distance or dissimilarity between two clusters or stocks. The higher they merge, the larger the distance is. For similar clusters or stocks, they merge in a lower value of height. To color the similar stocks,

a color threshold of $0.7 \times \max(d_{o_m^i, o_n^j})$ is used. Thus all similar clusters or stocks are colored with the same colors. By adjusting this color threshold, we can get the clusters from the dendrogram hierarchical cluster trees. As shown in the figures, using different definitions, we can get different hierarchical cluster trees and it is obvious that Figures 10(b) and 10(c) reveal more details of the structures, in which the distance between clusters is the largest of all pairs and the average distance of all pairs, respectively. The similar effect is also observed in Figures 11(b) and 11(c) for S&P468 networks. These clustering results are found to agree with the classifications of stocks very well. The colored clusters are composed of stocks mostly from the same economy sectors. Though there are exceptions that some stocks from different sectors are clustered together, or stocks from the same sector are clustered in different clusters. It is still astonishing to see that stocks can be clustered which agree with the economy sectors classification only from the prices information. These results indicate that hierarchical cluster trees constructed from price correlation matrix can reveal economy sectors and this has potential applications in portfolio selection and risk management.

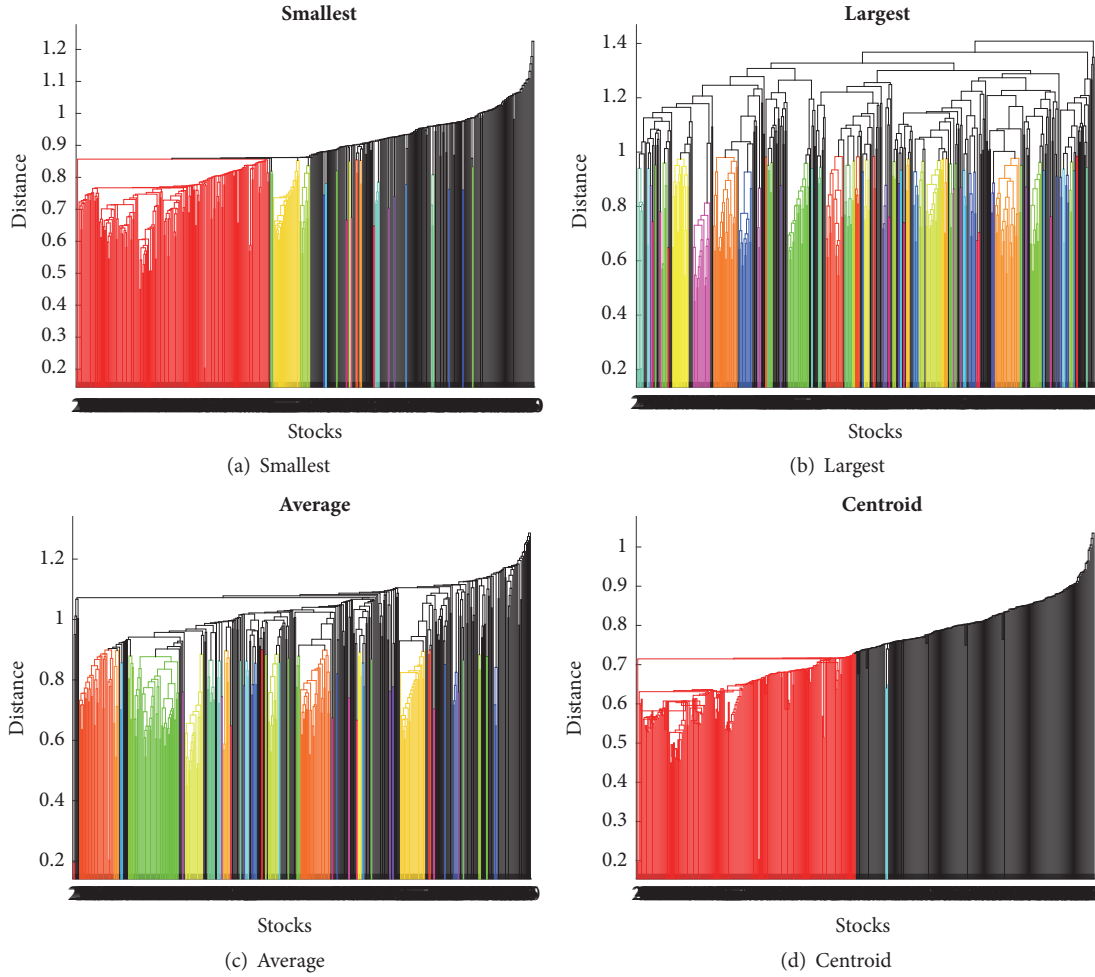


FIGURE 11: S&P468 dendrogram hierarchical cluster trees extracted with different distance definitions in (a) smallest distance for stock pairs, equation (8); (b) largest distance for stock pairs, equation (9); (c) average distance for stock pairs, equation (10); (d) distance between centroids for clusters, equation (11). The color threshold is 0.7. All stocks whose linkage values are less than this threshold would be colored with a unique color. As shown in the figures, different distance definitions extract different dendrogram hierarchical cluster trees whereas the same color threshold generates different results. Again, the largest distance definition reveals more details of network.

5.2. *Minimum Spanning Tree.* For a given undirected weighted network with N vertices, we can simplify the network by extracting the backbone of the network connecting all vertices, but with a minimum total length of edges, this backbone is called minimum spanning tree, or MST for short. Since loops or circles are not allowed to connect vertices, a MST has a topological structure of tree with $N - 1$ edges which is dramatically simplified from the original network which might have a maximum of $N(N - 1)/2$ edges. This brings huge advantages to the study of networks of stocks by reducing noises and simplifying the computation as well.

To construct a minimum spanning tree from a given network, it is easy to be achieved by using *Kruskal's Algorithm* [95], in which all edges are ranked in ascending order. Starting from the shortest edge on the edges ranking list, we add edges to the tree by keeping the tree in spanning form without introducing circles. When all edges are considered, we get a final minimum spanning tree comprising all connected N vertices with a minimum total length of $N - 1$ edges. For

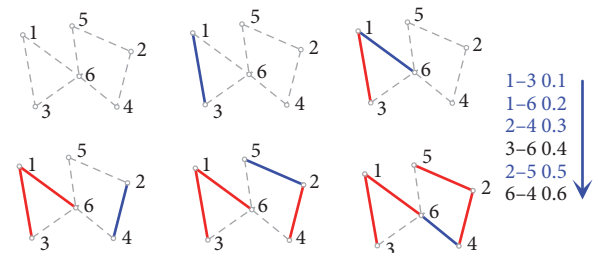


FIGURE 12: The extraction of a minimum spanning tree from a six vertices network using Kruskal's MST Algorithm. We rank all edges in descending order according to the edge lengths. Starting from the shortest edge and add the edges into the tree but avoiding loops or circles, after considering all edges, we get a final tree connecting all vertices with the minimum total edge lengths. In our example, after adding $e_{1,3}$, $e_{1,6}$, $e_{2,4}$, $e_{2,5}$, and $e_{6,4}$, we finally extract a tree of $T_{3,1,6,4,2,5}$ with a total length of 1.1.

a network in which all edges are with distinct lengths, the extracted MST is unique. In Figure 12, we demonstrate the

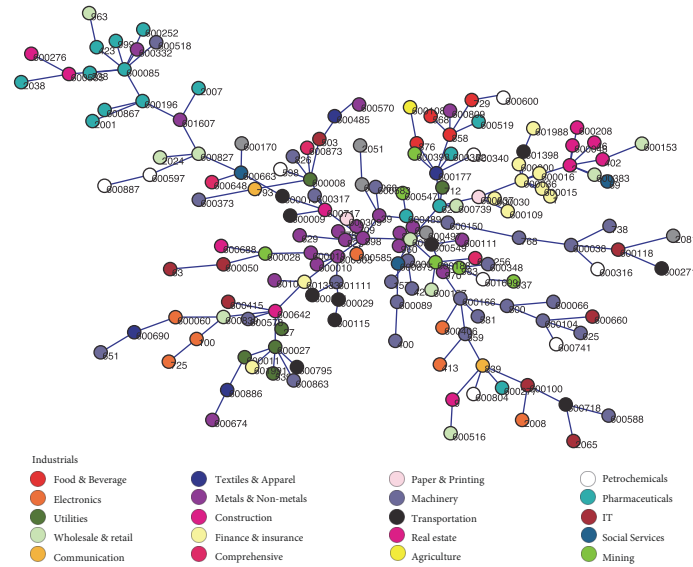


FIGURE 13: Minimum spanning tree of CSI163. Vertices are colored to indicate different industry sectors.

TABLE 10: In this table, we present the top 10 stocks with the largest degrees in the MST of CSI163. As is shown, the top 10 stocks are diverse in industry sectors which represents 1 Wholesale & retail stock, 1 Metals & Non-metals stock, 2 Pharmaceuticals stocks, 2 Real estate stocks, 1 Finance & insurance stock, 2 Utilities stocks, and 1 Textiles & Apparel stock.

Degree	Code	Name	Industry
11	600362	Beijingshualian Hypermarket	Wholesale & retail
9	898	Angang Steel	Metals & Non-metals
8	600085	Beijing Tongrentang	Pharmaceuticals
7	2	China Vanke	Real estate
7	600036	China Merchants Bank	Finance & insurance
6	600008	Beijing Capital	Utilities
6	600027	Huadian Power	Utilities
6	600177	Youngor Group	Textiles & Apparel
6	600642	Shanghai Wanye Enterprises	Real estate
5	623	Jilin Aodong Medicine Industry	Pharmaceuticals

process of extracting the minimum spanning tree from a six vertices network following Kruskal’s MST Algorithm. The edges are ranked in descending order, and we start from the shortest ones and add them into the tree but omit the edges which might introduce loops; after considering all edges, we get a minimum spanning tree with a minimum total length. In this example, edge (3,6) and (6,4) are omitted because that $e_{3,6}$ might bring a loop of (3,6,1) and $e_{6,4}$ might bring a loop of (6,4,2,5). Another widely used algorithm is *Prim’s Algorithm* [112] which begins with a starting vertex and adds the shortest one to the existing tree from all edges connected to the tree. By repeating this greedily, we can extract the minimum spanning tree of the given network. In this research, we apply Kruskal’s Algorithm to analyze the network structures of CSI163 and S&P468.

To extract the minimum spanning trees of the stock networks of CSI163 and S&P468, we average all correlation matrices over the investigated time windows, presented in

Figures 13 and 14 for CSI163 and S&P468, respectively. We see that the stocks of the same industry sectors are clustered in the MSTs for both CSI163 and S&P468, and this clustering effect is much more obvious for S&P468 in which stocks are well clustered according to the industry sectors of S&P500.

We further look into the connectivities of MSTs for both CSI163 and S&P468; we find that after the edge filtering process, some stocks are still well connected with other stocks. These stocks are the key stocks in the contribution of connectivities of the MSTs, while most stocks are poorly connected with the degree of only one or two. In Tables 10 and 11, we present the top 10 stocks according to their degrees in MST of CSI163 and S&P468, respectively. We find that the most connected stocks of CSI163 are diverse, while, for the MST of S&P468, 3 Financials stocks appear in the top 10. This agrees with other analysis that the Chinese stock market is much more diverse and Financials stocks play important roles in the US market.

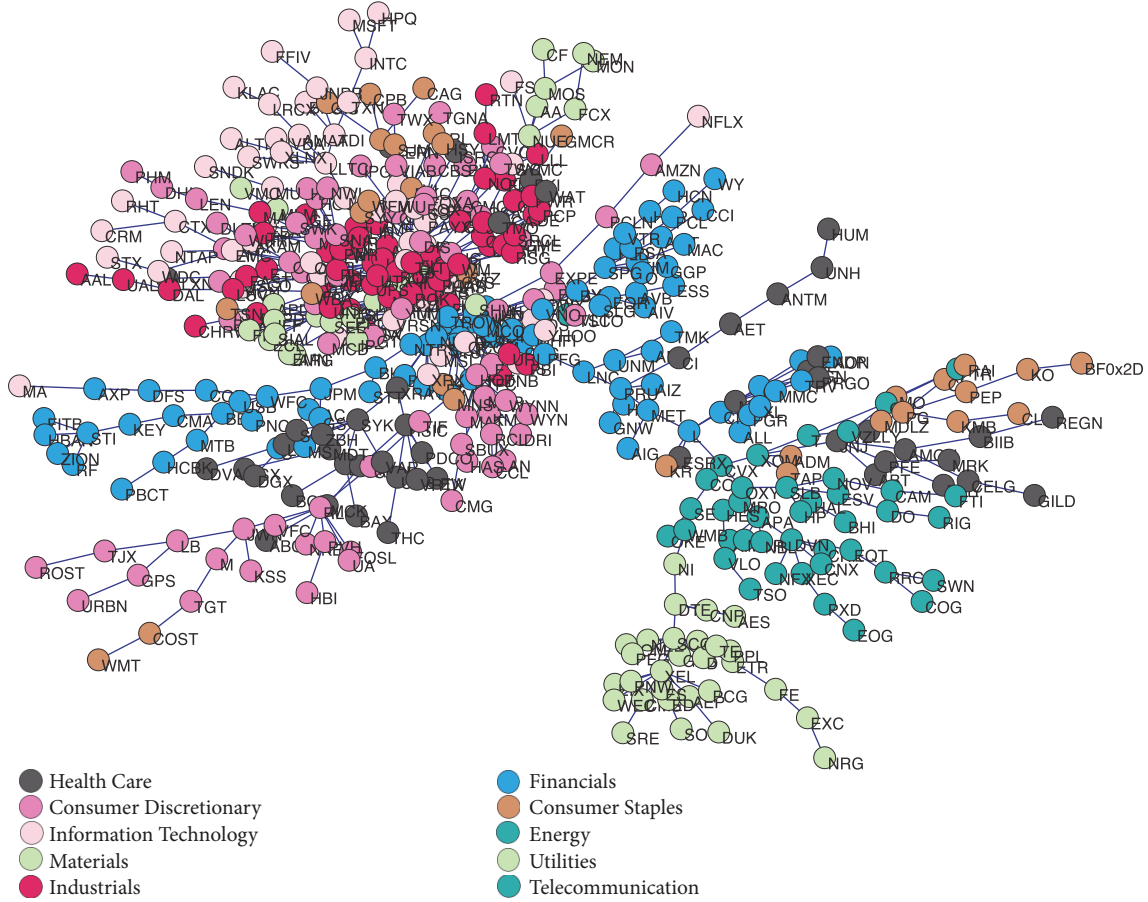


FIGURE 14: Minimum spanning tree of S&P468. Vertices are colored to indicate different industry sectors.

TABLE 11: In this table, we present the top 10 stocks with the largest degrees in the MST of S&P468. As is shown, Honeywell Intl. is the most connected stock in the MST with a degree of 38. The top 10 stocks are composed of 2 Industrials stocks, 3 Financials stocks, 2 Utilities stocks, 1 Health Care stock, and 1 Consumer Discretionary stock.

Degree	Tick	Stock Name	Industry
38	HON	Honeywell Intl.	Industrials
18	TROW	T. Rowe Price Group	Financials
11	SCG	SCANA Corp	Utilities
10	ITW	Illinois Tool Works	Industrials
9	ADP	Automatic Data Processing	IT
9	XEL	Xcel Energy Inc	Utilities
8	JNJ	Johnson & Johnson	Health Care
8	SPG	Simon Property	Financials
8	SNA	Snap-On Inc.	Consumer Discretionary
7	AMG	Affiliated Managers	Financials

5.3. Planar Maximally Filtered Graph. Like the minimum spanning tree (MST) approach, planar maximally filtered graphs, PMFG, is also an edge filtering method but the allowance of cliques up to 4 vertices shows much richer structure information of a network rather than a single tree. Based on the correlation matrix, the PMFG spans on a planar surface without crossing of edges but with loops and holes. It is believed that a PMFG might reveal more details of the

networks. After the introduction of PMFG by Tumminello et al. in the study of 100 stocks of NYSE [86], PMFG has been applied in many studies of financial networks. In [67], the authors study the PMFG networks of DAX 30 stocks. Instead of using the correlation matrix, a p -values matrix of EngleGranger cointegration test is used to extract the PMFG for Chinese stocks in [98]. The stability and robustness of PMFG for 300 NYSE stocks are compared with MST in a

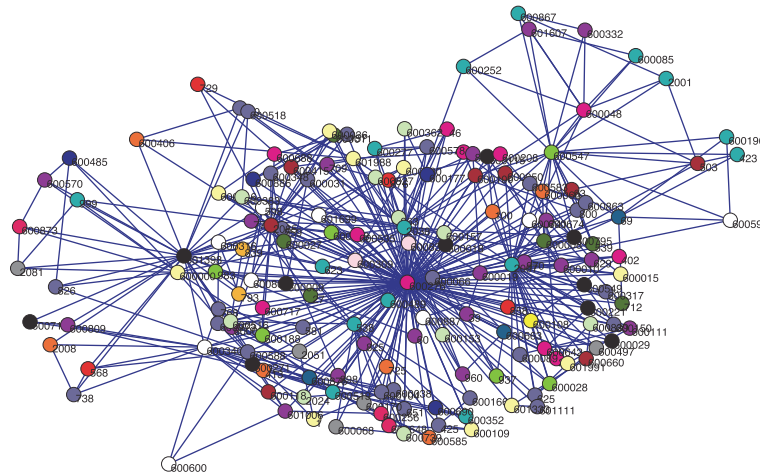


FIGURE 15: PMFG of CSII63.

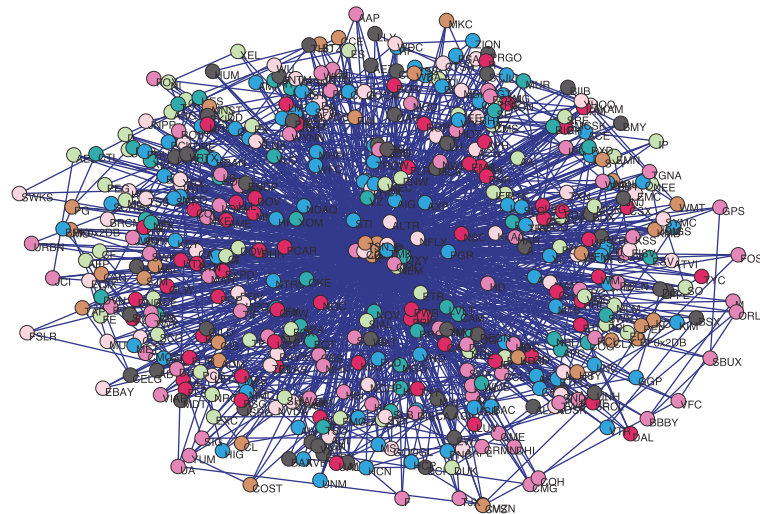


FIGURE 16: PMFG of S&P468.

running window approach, and the results reveal that PMFG is stabler than MST [104]. In [47], the same authors of [104] confirm that PMFG provides stronger robustness and stability in revealing network structures of stock markets. It has also been proven that the PMFG always contains an MST for the same distance matrix [86].

The PMFGs of CSII63 and S&P468 networks are plotted in Figures 15 and 16, respectively. We see that PMFGs have much more edges compared to MSTs. Further, we use another layout to plot the two PMFGs in Figures 17 and 18, from which, we find that PMFGs also produce good clusters for stocks of different industry sectors.

5.4. Asset Graph. In the minimum spanning tree (MST), a connected tree structure connecting all vertices with a minimum total length of edges is extracted. The selection process of adding edges in generating an MST out of a distance matrix is presented in Figure 12; an MST is always a connected single tree without disconnected parts. By connecting the N vertices, a total of $N - 1$ edges are needed,

where N is the number of vertices in the original network. It is obvious that an MST does not guarantee to be with the possible minimum total lengths of the $N - 1$ edges. By changing the strategy of how edges are selected and allowing disconnected parts, the asset graph (AG) approach is proposed in [87, 88]. Similar to MST, to generate an AG, we start the distance matrix containing all pairwise distances information of the network; we first rank all edges in ascending order from the shortest to the longest. Without considering the requirement of keeping a tree connected, we choose the top $N - 1$ edges to form an AG. It has been found that AG extracts similar structures as MST can do with smaller normalized length and with better stable structure over time. In this section, we show the AG networks for both markets. In Figures 19 and 20, we present AG structures for CSII63 and S&P468 networks, respectively. Compared with Figures 13 and 14 of the minimum spanning trees of CSII63 and S&P468 networks, we see that AG structures are more complex than MST and there are many isolated vertices in AG. The connected cliques in AG are the most correlated

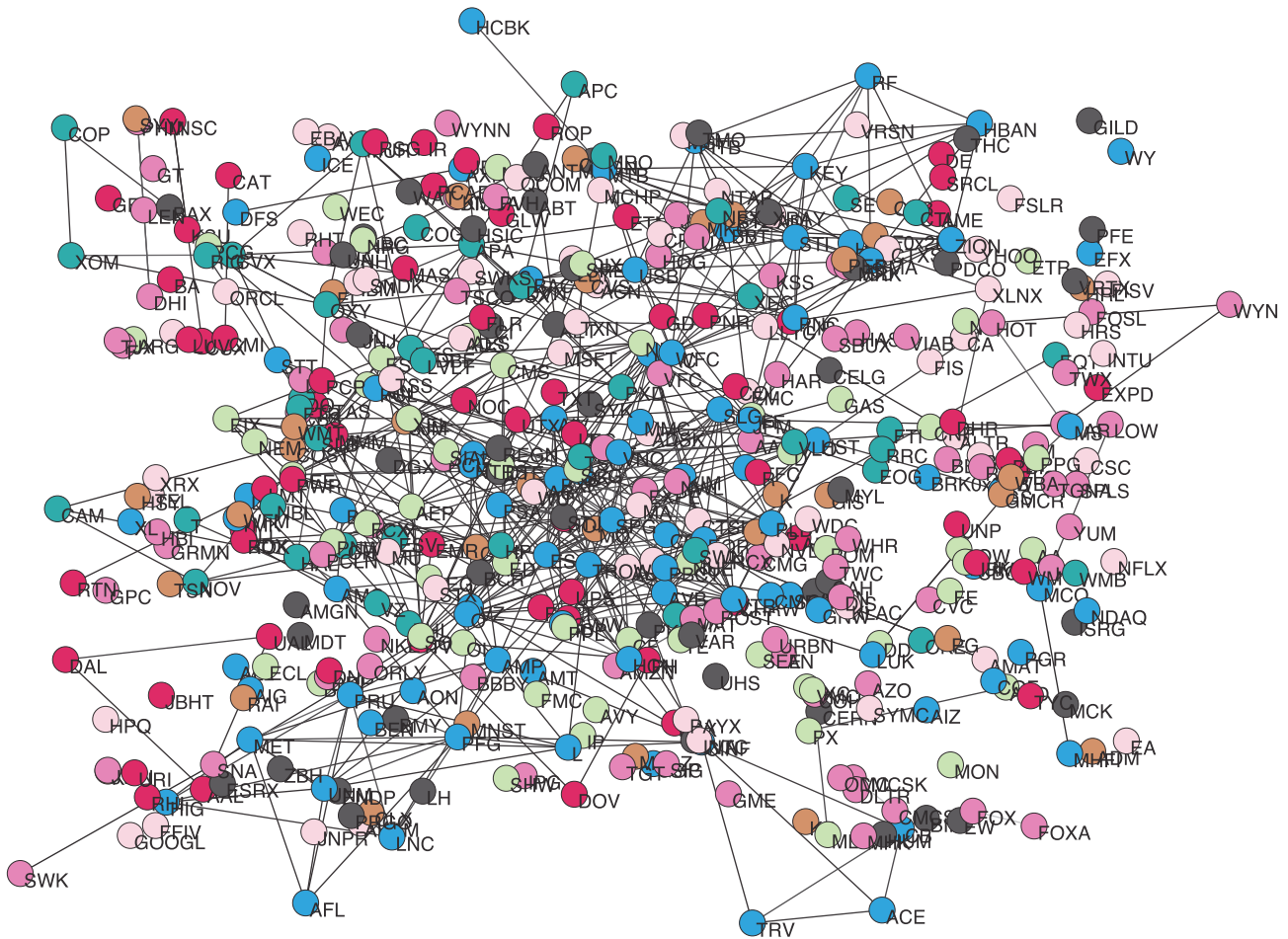


FIGURE 20: Asset graph of S&P468 with $N - 1$ shortest edges.

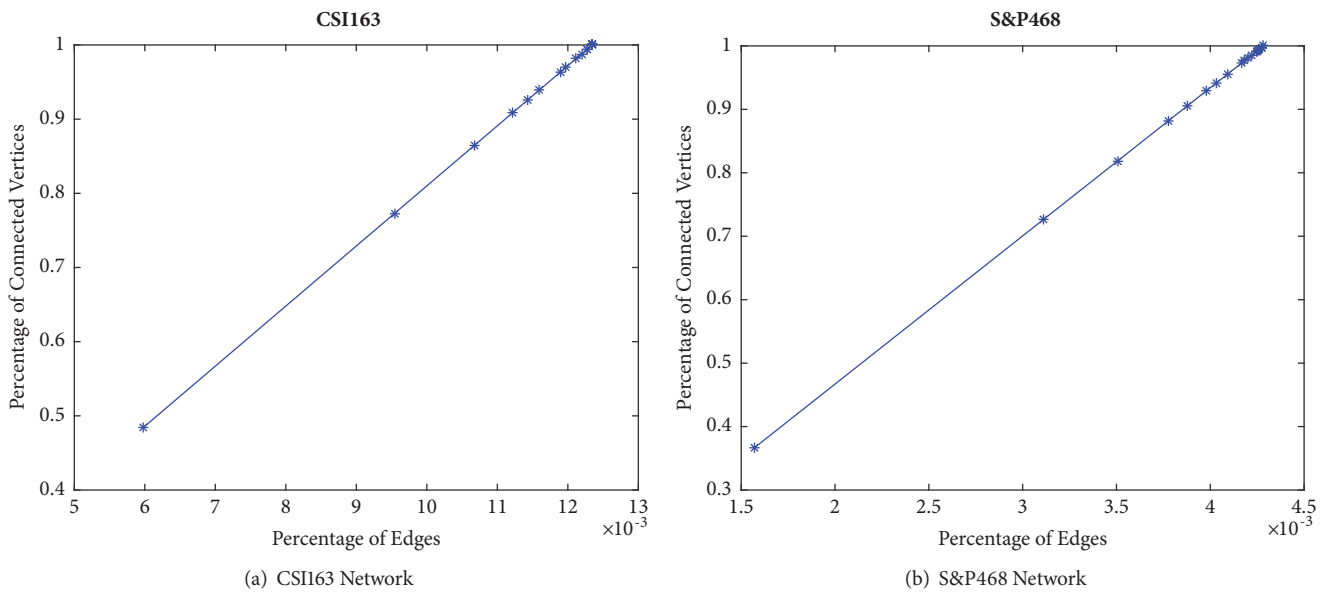


FIGURE 21: Percentages of connected vertices of AG against edge densities for CSII163 and S&P468 networks.

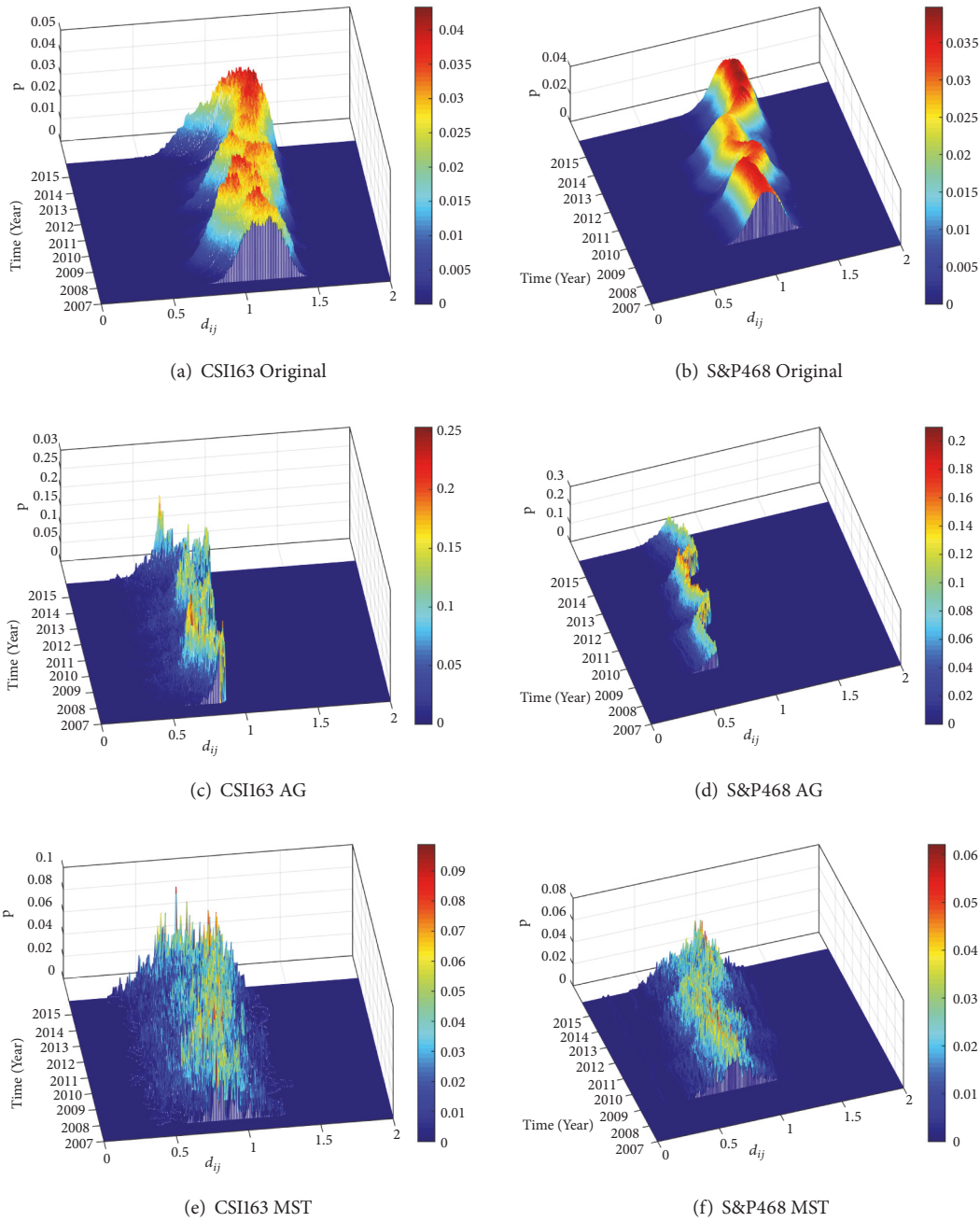


FIGURE 22: Probability distributions of all distances d_{ij} of N_t , AG_t , and MST_t of CSI163 and S&P468 over the years in our study period between 04/01/2007 and 06/11/2015. The total number of edges is $N(N-1)/2$ for N_t , and $N-1$ for AG_t and MST_t , respectively. Since the sliding window size $L_{CSI163} = 170$ and $L_{S\&P468} = 500$, so the data only starts after a period of L .

In the same way, we can calculate the total distances for AG_t and MST_t using equation (12), but considering the edge number for AG_t and MST_t is $N-1$, we normalize the average distance for them as

$$\langle d_{ij} \rangle = \frac{1}{N-1} \sum_{i,j} d_{ij}. \quad (14)$$

To investigate the tightness of the network, the total distance d_{total} and average distance $\langle d_{ij} \rangle$ for the original

network N_t , AG_t , and MST_t in our study periods for both networks are investigated. We plot the results in Figure 23 for d_{total} and Figure 24 for $\langle d_{ij} \rangle$, respectively. For each stock market, total distance d_{total} and average distance $\langle d_{ij} \rangle$ show similar shapes. For both stock markets, the values are in this order: $N_t > MST_t > AG_t$; i.e., the original networks have the largest values of d_{total} and $\langle d_{ij} \rangle$ compared to MST_t and AG_t , while AG_t has the smallest values.

By comparing the total distance d_{total} plotted in Figure 23(a) and the average distance $\langle d_{ij} \rangle$ plotted in

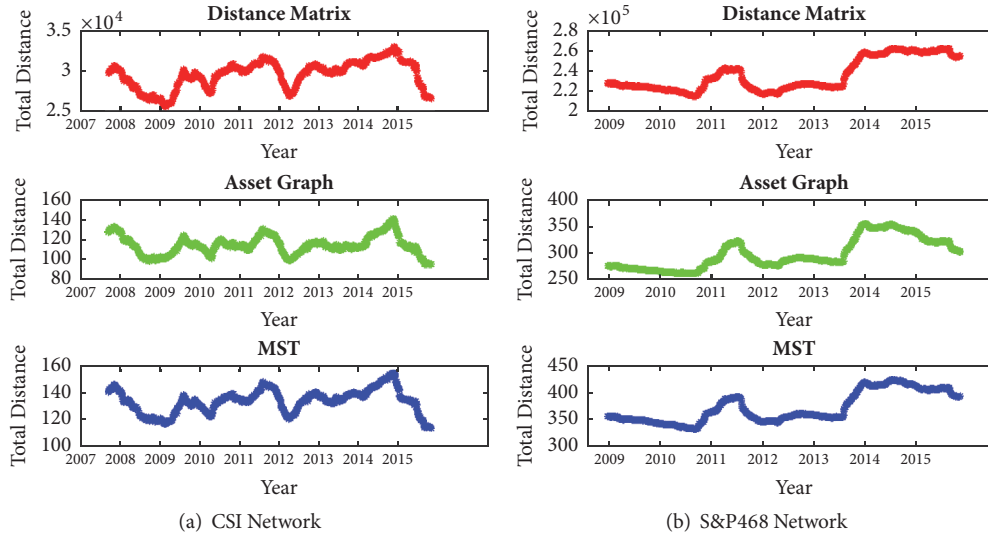


FIGURE 23: The evolving of total distances d_{total} of original network N_t , asset graph AG_t , and minimum spanning tree MST_t for CSI163 and S&P468 over time in the study period.

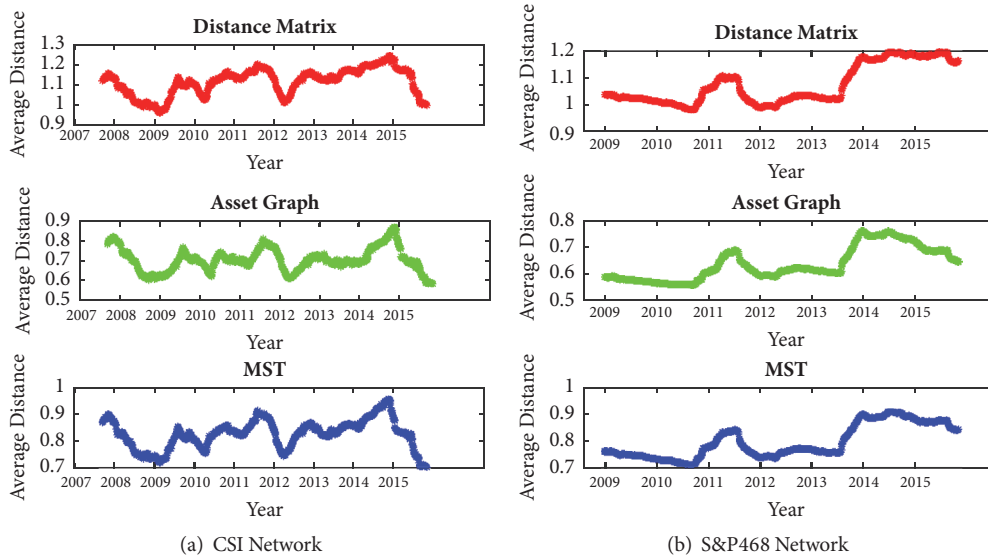


FIGURE 24: The evolving of average distances $\langle d_{ij} \rangle$ of original network N_t , asset graph AG_t , and minimum spanning tree MST_t for CSI163 and S&P468 over time in the study period.

Figure 24(a) for CSI163 over the study period, we find the six plots share similar shapes. The same similarities are also observed in Figures 23(b) and 24(b) for the S&P468 network. This indicates that the AG and MST are both good backbones of the whole original market networks, and this tracking stays robust over time. For both networks, we also find that the curve of MST_t is above AG_t which means the total and average distances are slightly larger in MST than in AG. Our findings agree with the results reported in [87]. Since the two stock markets datasets have different stock numbers, we compare the average distance between the two markets, and as shown in Figure 24, we see that the CSI163 is slightly sparser than S&P468, which indicates that the CSI163 which is a developing market is more diversified than

S&P468 which is a developed market; this also agrees with many previous research.

In Table 12, we summary the average $\langle\langle d_{ij} \rangle\rangle$, minimum $\langle d_{ij} \rangle_{\min}$, maximum $\langle d_{ij} \rangle_{\max}$, and standard deviation $\langle d_{ij} \rangle_{\sigma}$ of CSI163 and S&P468 networks for three kinds networks: original, AG, and MST. We can see that the values are in the order of $N > MST > AG$ for average, minimum, and maximum. Also the three networks have similar standard deviations. We find that the values of $\langle\langle d_{ij} \rangle\rangle$ and minimum $\langle d_{ij} \rangle_{\min}$ for CSI163 are slightly larger than S&P468 which indicates stocks in CSI163 are less likely to correlated than in S&P468. To visualize the distributions of these three kinds of networks, we plot the probability density function (PDF) for the original network, AG, and MST for CSI163 and S&P468

TABLE 12: Average $\langle\langle d_{ij} \rangle\rangle$, minimum $\langle d_{ij} \rangle_{\min}$, maximum $\langle d_{ij} \rangle_{\max}$, and standard deviation $\langle d_{ij} \rangle_{\sigma}$ of CSI163 and S&P468 networks. The values of first row belong to the original network N , those of the second row belong to the AG, and those of the third row belong to MST.

CSI163				S&P468			
$\langle\langle d_{ij} \rangle\rangle$	$\langle d_{ij} \rangle_{\min}$	$\langle d_{ij} \rangle_{\max}$	$\langle d_{ij} \rangle_{\sigma}$	$\langle\langle d_{ij} \rangle\rangle$	$\langle d_{ij} \rangle_{\min}$	$\langle d_{ij} \rangle_{\max}$	$\langle d_{ij} \rangle_{\sigma}$
1.1145	1.2445	0.9619	0.0650	1.0754	1.1968	0.9816	0.0731
0.7049	0.8708	0.5868	0.0578	0.6377	0.7607	0.5579	0.0623
0.8251	0.9565	0.7017	0.0540	0.7971	0.9092	0.7113	0.0615

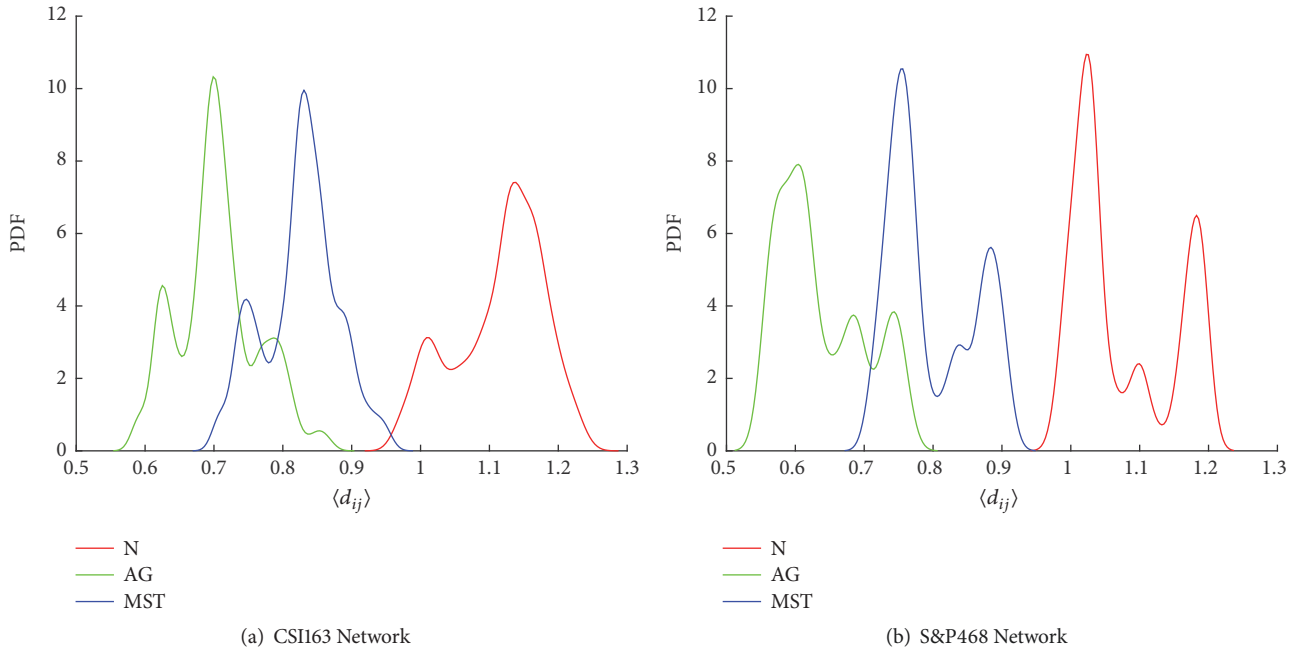


FIGURE 25: Probability density function (PDF) of average distance $\langle d_{ij} \rangle$ of original network N , asset graph AG, and minimum spanning tree MST for CSI163 and S&P468.

in Figures 25(a) and 25(b), respectively. We see that the distributions of all three networks share similar shapes but with different mean centers; as shown in the figures, the AG locates on the left, MST locates in the center, and original locates on the right.

6. Conclusion and Discussion

In this research, we investigated the properties and models of the complex network theory and its applications from data science perspective. Using the daily close prices of two sets of stocks from CSI300 and S&P500, we constructed the correlation matrices for both the whole study periods and all sliding windows. Based on these correlation matrices, we build the networks with stocks as the vertices and correlation relationships as the edges. We systematically applied network filtering methods like hierarchical tree, minimum spanning tree, planar maximally filtered graph, and asset graph to simplify the networks. For each filtered network, the network properties are discussed. Financial markets are complex systems, and it is important to extract useful information from the noise background by applying methods like complex networks. We find that, for the stock markets, CSI300

and S&P500, the former is an emerging market while the latter is a mature well-developed market. They share similar properties in many ways and also vary in many aspects. The revealed properties and robustness might provide sights of the structures and dynamics of the two stock markets for practitioners and regulators. Furthermore, it is interesting to develop trading strategies with the information revealed from the topological networks of stocks or indices. For instance, the pair trading [113–116] is a basic and market neutral strategy considering the movement of a correlated stock pair, in which if the spread widens, then traders can short one and long another one to gain the spread. One might use the information of the networks to identify the pairs and evaluate the reliabilities. Also, considering pairs between groups of stocks rather than only two stocks, we might use the component or cluster information revealed in the networks to build the trading groups. Furthermore, the directed networks built with Granger causalities or lagged correlations might give more lead/lag details of stock pairs on the time factors asynchronously. Second, with the help of network edge filtering, we can significantly simplify the networks, but most studies focus on the topological simplification without concerns of the original portfolio returns. What if

we consider the returns with the topology of the networks to optimize the portfolio selection? The topological structure can give us information on how diverse the portfolio is but this is not enough to design the portfolio without return information. A possible way is to adjust the portfolio selection by considering measurements like the ratio of returns over total distances of a portfolio or other approaches combining both topological and return information. Also, the techniques of network modeling and analysis can enhance the ability in policy modeling and decision making. We hope this work can inspire policymakers and researchers in applying network theories in wider applications.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

An online appendix file containing all tick information is available on the publisher website. (*Supplementary Materials*)

References

- [1] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [2] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [3] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [4] B. Albert-László, *Linked: The New Science of Networks*, Perseus, 2002.
- [5] L. D. F. Costa, O. N. Oliveira Jr., G. Travieso et al., "Analyzing and modeling real-world phenomena with complex networks: a survey of applications," *Advances in Physics*, vol. 60, no. 3, pp. 329–412, 2011.
- [6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. W. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [7] Z. Gao, M. Small, and J. Kurths, "Complex network analysis of time series," *EPL (Europhysics Letters)*, vol. 116, no. 5, Article ID 50001, 2016.
- [8] M. Barigozzi and M. Hallin, "A network analysis of the volatility of high dimensional financial series," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 66, no. 3, pp. 581–605, 2017.
- [9] S. An, X. Gao, M. Jiang, and X. Sun, "Multivariate financial time series in the light of complex network analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 503, pp. 1241–1255, 2018.
- [10] C. K. Tse, J. Liu, and F. C. M. Lau, "A network perspective of the stock market," *Journal of Empirical Finance*, vol. 17, no. 4, pp. 659–667, 2010.
- [11] M. Tumminello, F. Lillo, and R. N. Mantegna, "Correlation, hierarchies, and networks in financial markets," *Journal of Economic Behavior & Organization*, vol. 75, no. 1, pp. 40–58, 2010.
- [12] M. Lin, G. Wang, C. Xie, and H. E. Stanley, "Cross-correlations and influence in world gold markets," *Physica A: Statistical Mechanics and its Applications*, vol. 490, pp. 504–512, 2018.
- [13] V. Boginski, S. Butenko, and P. M. Pardalos, "Statistical analysis of financial networks," *Computational Statistics & Data Analysis*, vol. 48, no. 2, pp. 431–443, 2005.
- [14] V. Boginski, S. Butenko, and P. M. Pardalos, "Mining market data: A network approach," *Computers & Operations Research*, vol. 33, part Special Issue: Operations Research and Data Mining, no. 11, pp. 3171–3184, 2006.
- [15] G.-J. Wang, C. Xie, and H. E. Stanley, "Correlation Structure and Evolution of World Stock Markets: Evidence from Pearson and Partial Correlation-Based Networks," *Computational Economics*, vol. 51, no. 3, pp. 607–635, 2018.
- [16] G.-J. Wang, C. Xie, and S. Chen, "Multiscale correlation networks analysis of the US stock market: a wavelet analysis," *Journal of Economic Interaction and Coordination*, vol. 12, no. 3, pp. 561–594, 2017.
- [17] Nicolò Musmeci, Vincenzo Nicosia, Tomaso Aste, Tiziana Di Matteo, and Vito Latora, "The Multiplex Dependency Structure of Financial Markets," *Complexity*, vol. 2017, Article ID 9586064, 13 pages, 2017.
- [18] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.
- [19] V. Kalyagin, A. Koldanov, P. Koldanov, and V. Zamaraev, "Market graph and Markowitz model," in *Optimization in Science And Engineering*, T. M. Rassias, C. A. Floudas, and S. Butenko, Eds., pp. 293–306, Springer, New York, USA, 2014.
- [20] V. Boginski, S. Butenko, O. Shirokikh, S. Trukhanov, and J. Gil Lafuente, "A network-based data mining approach to portfolio selection via weighted clique relaxations," *Annals of Operations Research*, vol. 216, no. 1, pp. 23–34, 2014.
- [21] R. N. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 11, no. 1, pp. 193–197, 1999.
- [22] S. Battiston, J. D. Farmer, A. Flache et al., "Complexity theory and financial regulation," *Science*, vol. 351, no. 6275, pp. 818–819, 2016.
- [23] X.-G. Yan, C. Xie, and G.-J. Wang, "Stock market network's topological stability: evidence from planar maximally filtered graph and minimal spanning tree," *International Journal of Modern Physics B*, vol. 29, no. 22, Article ID 1550161, 2015.
- [24] X.-G. Yan, C. Xie, and G.-J. Wang, "The stability of financial market networks," *EPL (Europhysics Letters)*, vol. 107, no. 4, Article ID 48002, 2014.
- [25] P. Glasserman and H. P. Young, "How likely is contagion in financial networks?" *Journal of Banking & Finance*, vol. 50, Supplement C, pp. 383–399, 2015.
- [26] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi, "Systemic risk and stability in financial networks," *American Economic Review*, vol. 105, no. 2, pp. 564–608, 2015.
- [27] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.
- [28] S. Battiston, D. Delli Gatti, M. Gallegati, B. Greenwald, and J. E. Stiglitz, "Liaisons dangereuses: increasing connectivity, risk sharing, and systemic risk," *Journal of Economic Dynamics and Control*, vol. 36, no. 8, pp. 1121–1141, 2012.

- [29] N. Hautsch, J. Schaumburg, and M. Schienle, “Financial network systemic risk contributions,” *Review of Finance*, vol. 19, no. 2, pp. 685–738, 2015.
- [30] G. Cimini, T. Squartini, D. Garlaschelli, and A. Gabrielli, “Systemic Risk Analysis on Reconstructed Economic and Financial Networks,” *Scientific Reports*, vol. 5, Article ID 15758, 2015.
- [31] E. Nier, J. Yang, T. Yorulmazer, and A. Alentorn, “Network models and financial stability,” *Journal of Economic Dynamics and Control (JEDC)*, vol. 31, no. 6, pp. 2033–2060, 2007.
- [32] K. Anand, P. Gai, S. Kapadia, S. Brennan, and M. Willison, “A network model of financial system resilience,” *Journal of Economic Behavior & Organization*, vol. 85, Supplement C, pp. 219–235, 2013.
- [33] B. M. Tabak, T. R. Serra, and D. O. Cajueiro, “Topological properties of stock market networks: the case of Brazil,” *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 16, pp. 3240–3249, 2010.
- [34] F. Caccioli, M. Shrestha, C. Moore, and J. D. Farmer, “Stability analysis of financial contagion due to overlapping portfolios,” *Journal of Banking & Finance*, vol. 46, Supplement C, pp. 233–245, 2014.
- [35] G.-J. Wang, C. Xie, K. He, and H. E. Stanley, “Extreme risk spillover network: application to financial institutions,” *Quantitative Finance*, vol. 17, no. 9, pp. 1417–1433, 2017.
- [36] S. Poledna, J. L. Molina-Borboa, S. Martínez-Jaramillo, M. van der Leij, and S. Thurner, “The multi-layer network nature of systemic risk and its implications for the costs of financial crises,” *Journal of Financial Stability*, vol. 20, Supplement C, pp. 70–81, 2015.
- [37] T. Kauê Dal’Maso Peron, L. da Fontoura Costa, and F. A. Rodrigues, “The structure and resilience of financial market networks,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 1, 2012.
- [38] M. Gofman, “Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions,” *Journal of Financial Economics*, vol. 124, no. 1, pp. 113–146, 2017.
- [39] S. Battiston, M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli, “DebtRank: too central to fail? financial networks, the FED and systemic risk,” *Scientific Reports*, vol. 2, article no. 541, 2012.
- [40] C. Zhou, “Are Banks Too Big to Fail? Measuring Systemic Importance of Financial Institutions,” *International Journal of Central Banking*, vol. 6, no. 34, pp. 205–250, 2010, <https://ideas.repec.org/a/ijc/ijcjou/y2010q4a10.html>.
- [41] C. Brooks, A. G. Rew, and S. Ritson, “A trading strategy based on the lead-lag relationship between the spot index and futures contract for the FTSE 100,” *International Journal of Forecasting*, vol. 17, no. 1, pp. 31–44, 2001.
- [42] C. Curme, M. Tumminello, R. N. Mantegna, H. E. Stanley, and D. Y. Kenett, “Emergence of statistically validated financial intraday lead-lag relationships,” *Quantitative Finance*, vol. 15, no. 8, pp. 1375–1386, 2015.
- [43] K. Chen, P. Luo, B. Sun, and H. Wang, “Which stocks are profitable? A network method to investigate the effects of network structure on stock returns,” *Physica A: Statistical Mechanics and its Applications*, vol. 436, pp. 224–235, 2015.
- [44] F. Pozzi, T. Di Matteo, and T. Aste, “Spread of risk across financial markets: better to invest in the peripheries,” *Scientific Reports*, vol. 3, Article ID 1665, 2013, <http://dx.doi.org/10.1038/srep01665>.
- [45] W. Zhang, C. Li, Y. Ye, W. Li, and E. W. T. Ngai, “Dynamic Business Network Analysis for Correlated Stock Price Movement Prediction,” *IEEE Intelligent Systems*, vol. 30, no. 2, pp. 26–33, 2015.
- [46] M. Billo, M. Getmansky, A. W. Lo, and L. Pelizzon, “Econometric measures of connectedness and systemic risk in the finance and insurance sectors,” *Journal of Financial Economics*, vol. 104, no. 3, pp. 535–559, 2012.
- [47] T. Di Matteo, F. Pozzi, and T. Aste, “The use of dynamical networks to detect the hierarchical organization of financial market sectors,” *The European Physical Journal B*, vol. 73, no. 1, pp. 3–11, 2010.
- [48] X. F. Jiang, T. T. Chen, and B. Zheng, “Structure of local interactions in complex financial dynamics,” *Scientific Reports*, vol. 4, Article ID 5321, 2014, <http://dx.doi.org/10.1038/srep05321>.
- [49] L. De Benedictis and L. Tajoli, “The World Trade Network,” *The World Economy*, vol. 34, no. 8, pp. 1417–1454, 2011.
- [50] J. He and M. W. Deem, “Structure and response in the world trade network,” *Physical Review Letters*, vol. 105, Article ID 198701, 2010.
- [51] D. in’t Veld and I. van Lelyveld, “Finding the core: Network structure in interbank markets,” *Journal of Banking & Finance*, vol. 49, pp. 27–40, 2014.
- [52] Y. Luo, J. Xiong, L. G. Dong, and Y. Tang, “Statistical correlation properties of the SHIBOR interbank lending market,” *China Finance Review International*, vol. 5, no. 2, pp. 91–102, 2015.
- [53] M. Affinito and A. F. Pozzolo, “The interbank network across the global financial crisis: Evidence from Italy,” *Journal of Banking & Finance*, vol. 80, no. Supplement C, pp. 90–107, 2017.
- [54] C.-P. Georg, “The effect of the interbank network structure on contagion and common shocks,” *Journal of Banking & Finance*, vol. 37, no. 7, pp. 2216–2228, 2013.
- [55] D. J. Fenn, M. A. Porter, P. J. Mucha et al., “Dynamical clustering of exchange rates,” *Quantitative Finance*, vol. 12, no. 10, pp. 1493–1520, 2012.
- [56] G. Iori, G. De Masi, O. V. Precup, G. Gabbi, and G. Caldarelli, “A network analysis of the Italian overnight money market,” *Journal of Economic Dynamics and Control (JEDC)*, vol. 32, no. 1, pp. 259–278, 2008.
- [57] C. Piccardi, L. Calatroni, and F. Bertoni, “Communities in Italian corporate networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 22, pp. 5247–5258, 2010.
- [58] S. Vitali, J. B. Glattfelder, and S. Battiston, “The network of global corporate control,” *PLoS ONE*, vol. 6, no. 10, Article ID 0025995, pp. 1–6, 2011.
- [59] C. Minoiu and J. A. Reyes, “A network analysis of global banking: 1978–2010,” *Journal of Financial Stability*, vol. 9, no. 2, pp. 168–184, 2013.
- [60] T. A. Peltonen, M. Scheicher, and G. Vuillemeij, “The network structure of the CDS market and its determinants,” *Journal of Financial Stability*, vol. 13, Supplement C, pp. 118–133, 2014.
- [61] L. Marotta, S. Micciché, Y. Fujiwara et al., “Backbone of credit relationships in the Japanese credit market,” *EPJ Data Science*, vol. 5, no. 1, 2016.
- [62] F. X. Diebold and K. Yilmaz, “On the network topology of variance decompositions: measuring the connectedness of financial firms,” *Journal of Econometrics*, vol. 182, no. 1, pp. 119–134, 2014.
- [63] S. Saavedra, L. J. Gilarranz, R. P. Rohr, M. Schnabel, B. Uzzi, and J. Bascompte, “Stock fluctuations are correlated and amplified across networks of interlocking directorates,” *EPJ Data Science*, vol. 3, no. 1, Article ID 30, 2014.

- [64] H. Chen, Y. Mai, and S.-P. Li, "Analysis of network clustering behavior of the Chinese stock market," *Physica A: Statistical Mechanics and its Applications*, vol. 414, pp. 360–367, 2014.
- [65] W.-Q. Huang, X.-T. Zhuang, and S. Yao, "A network analysis of the Chinese stock market," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 14, pp. 2956–2964, 2009.
- [66] F. Ren and W.-X. Zhou, "Dynamic evolution of cross-correlations in the Chinese stock market," *PLoS ONE*, vol. 9, no. 5, Article ID e97711, 2014.
- [67] J. Birch, A. A. Pantelous, and K. Soramäki, "Analysis of Correlation Based Networks Representing DAX 30 Stock Price Returns," *Computational Economics*, vol. 47, no. 4, pp. 501–525, 2016.
- [68] J. G. Brida and W. A. Risso, "Hierarchical structure of the German stock market," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3846–3852, 2010.
- [69] P. Caraiiani, "Characterizing emerging European stock markets through complex networks: from local properties to self-similar characteristics," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 13, pp. 3629–3637, 2012.
- [70] K. Kosmidou, D. Kousenidis, A. Ladas, and C. Negkakis, "Determinants of risk in the banking sector during the European Financial Crisis," *Journal of Financial Stability*, vol. 33, pp. 285–296, 2017.
- [71] N. Paltalidis, D. Gounopoulos, R. Kizys, and Y. Koutelidakis, "Transmission channels of systemic risk and contagion in the European financial network," *Journal of Banking & Finance*, vol. 61, Supplement 1, pp. S36–S52, 2015.
- [72] M. F. Da Silva, É. J. De Area Leão Pereira, A. M. Da Silva Filho, A. P. N. De Castro, J. Garcia Vivas Miranda, and G. F. Zebende, "Quantifying cross-correlation between Ibovespa and Brazilian blue-chips: The DCCA approach," *Physica A: Statistical Mechanics and its Applications*, vol. 424, pp. 124–129, 2015.
- [73] J. Lee, J. Youn, and W. Chang, "Intraday volatility and network topological properties in the Korean stock market," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 4, pp. 1354–1360, 2012.
- [74] A. Vizgunov, B. Goldengorin, V. Kalyagin, A. Koldanov, P. Koldanov, and P. M. Pardalos, "Network approach for the Russian stock market," *Computational Management Science*, vol. 11, no. 1-2, pp. 45–55, 2014.
- [75] S. Martinez-Jaramillo, B. Alexandrova-Kabadjova, B. Bravo-Benitez, and J. P. Solórzano-Margain, "An empirical study of the Mexican banking system's network and its implications for systemic risk," *Journal of Economic Dynamics and Control (JEDC)*, vol. 40, Supplement C, pp. 242–265, 2014.
- [76] E. Jondeau, E. Jurczenko, and M. Rockinger, "Moment component analysis: An illustration with international stock markets," *Journal of Business & Economic Statistics*, pp. 1–23, 2017, <https://doi.org/10.1080/07350015.2016.1216851>.
- [77] R. Kali and J. Reyes, "The architecture of globalization: a network approach to international economic integration," *Journal of International Business Studies*, vol. 38, no. 4, pp. 595–620, 2007.
- [78] P. Giudici and A. Spelta, "Graphical network models for international financial flows," *Journal of Business & Economic Statistics*, vol. 34, no. 1, pp. 128–138, 2016.
- [79] N. Cetorelli and S. Peristiani, "Prestigious stock exchanges: A network analysis of international financial centers," *Journal of Banking & Finance*, vol. 37, no. 5, pp. 1543–1551, 2013.
- [80] L. Sandoval, "Structure of a Global Network of financial companies based on transfer entropy," *Entropy*, vol. 16, no. 8, pp. 4443–4482, 2014.
- [81] P. Caraiiani, "Using complex networks to characterize international business cycles," *PLoS ONE*, vol. 8, no. 3, Article ID e58109, 2013.
- [82] T. C. Silva, S. R. S. de Souza, and B. M. Tabak, "Structure and dynamics of the global financial network," *Chaos, Solitons & Fractals*, vol. 88, Supplement C, pp. 218–234, 2016.
- [83] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud, "Random matrix theory and financial correlations," *International Journal of Theoretical and Applied Finance*, vol. 03, no. 03, pp. 391–397, 2000.
- [84] I. I. Dimov, P. N. Kolm, L. Maclin, and D. Y. C. Shiber, "Hidden noise structure and random matrix models of stock correlations," *Quantitative Finance*, vol. 12, no. 4, pp. 567–572, 2012.
- [85] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, "Topology of correlation-based minimal spanning trees in real and model markets," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 4, Article ID 046130, 2003.
- [86] M. Tumminello, T. Aste, T. di Matteo, and R. N. Mantegna, "A tool for filtering information in complex systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 30, pp. 10421–10426, 2005.
- [87] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, "Asset Trees and Asset Graphs in Financial Markets," *Physica Scripta*, vol. 2003, no. T106, p. 48, 2003.
- [88] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, "Dynamics of market correlations: Taxonomy and portfolio analysis," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 5, Article ID 056110, 2003.
- [89] D. Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershoren, R. N. Mantegna, and E. Ben-Jacob, "Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market," *PLoS ONE*, vol. 5, no. 12, Article ID e15032, 2010.
- [90] D. Y. Kenett, X. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, "Partial correlation analysis: applications for financial markets," *Quantitative Finance*, vol. 15, no. 4, pp. 569–578, 2015.
- [91] L. China Securities Index Company, *China Securities Index 300 Index*, 2017, http://www.csindex.com.cn/sseportal_en/csiportal/zs/jbxx/report.do?code=000300.
- [92] "S&P 500 index," 2017, <http://us.spindices.com/indices/equity/sp-500>.
- [93] J.-P. Onnela, *Complex Networks in The Study of Financial And Social Systems*, Helsinki University of Technology, 2006.
- [94] W. Jang, J. Lee, and W. Chang, "Currency crises and the evolution of foreign exchange market: evidence from minimum spanning tree," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 4, pp. 707–718, 2011.
- [95] J. B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 48–50, 1956.
- [96] G. Bonanno, G. Caldarelli, F. Lillo, S. Micciché, N. Vandewalle, and R. N. Mantegna, "Networks of equities in financial markets," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 363–371, 2004.
- [97] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, New York, NY, USA, 2000.

- [98] C. Tu, "Cointegration-based financial networks study in Chinese stock market," *Physica A: Statistical Mechanics and its Applications*, vol. 402, no. 3, pp. 245–254, 2014.
- [99] G.-J. Wang and C. Xie, "Correlation structure and dynamics of international real estate securities markets: A network perspective," *Physica A: Statistical Mechanics and Its Applications*, vol. 424, pp. 176–193, 2015.
- [100] M. Eryiğit and R. Eryiğit, "Network structure of cross-correlations among the world market indices," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 17, pp. 3551–3562, 2009.
- [101] T. Di Matteo and T. Aste, "Extracting the correlation structure by means of planar embedding," in *Proceedings of the SPIE*, vol. 6039, p. 60390P, Brisbane, Australia, 2005.
- [102] J. Birch, A. A. Pantelous, and K. Zuev, "The maximum number of 3- and 4-cliques within a planar maximally filtered graph," *Physica A: Statistical Mechanics and its Applications*, vol. 417, pp. 221–229, 2015.
- [103] N. Musmeci, T. Aste, and T. Di Matteo, "Relation between financial market structure and the real economy: Comparison between clustering methods," *PLoS ONE*, vol. 10, no. 3, 2015, <http://dx.doi.org/10.1371>.
- [104] F. Pozzi, T. Aste, G. Rotundo et al., "Dynamical correlations in financial systems," in *Proceedings of the SPIE*, vol. 6802, p. 68021E, Canberra, ACT, Australia, 2007.
- [105] J. Liu, C. K. Tse, and K. He, "Fierce stock market fluctuation disrupts scalefree distribution," *Quantitative Finance*, vol. 11, no. 6, pp. 817–823, 2011.
- [106] V. Batagelj and A. Mrvar, "Pajek-program for large network analysis," *Connections*, vol. 21, no. 2, pp. 47–57, 1998.
- [107] Y. Mai, H. Chen, and L. Meng, "An analysis of the sectorial influence of CSI300 stocks within the directed network," *Physica A: Statistical Mechanics and its Applications*, vol. 396, pp. 235–241, 2014.
- [108] R. Rammal, G. Toulouse, and M. A. Virasoro, "Ultrametricity for physicists," *Reviews of Modern Physics*, vol. 58, no. 3, pp. 765–788, 1986.
- [109] R. Yang, X. Li, and T. Zhang, "Analysis of linkage effects among industry sectors in China's stock market before and after the financial crisis," *Physica A: Statistical Mechanics and its Applications*, vol. 411, pp. 12–20, 2014.
- [110] E. Kantar, M. Keskin, and B. Deviren, "Analysis of the effects of the global financial crisis on the Turkish economy, using hierarchical methods," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 7, pp. 2342–2352, 2012.
- [111] A. Rea and W. Rea, "Visualization of a stock market correlation matrix," *Physica A: Statistical Mechanics and its Applications*, vol. 400, pp. 109–123, 2014.
- [112] R. C. Prim, "Shortest connection networks and some generalizations," *Bell System Technical Journal*, vol. 36, no. 6, pp. 1389–1401, 1957.
- [113] R. J. Elliott, J. van der Hoek, and W. P. Malcolm, "Pairs trading," *Quantitative Finance*, vol. 5, no. 3, pp. 271–276, 2005.
- [114] E. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst, "Pairs trading: performance of a relative-value arbitrage rule," *Review of Financial Studies*, vol. 19, no. 3, pp. 797–827, 2006.
- [115] S. Mudchanatongsuk, J. A. Primbs, and W. Wong, "Optimal pairs trading: A stochastic control approach," in *Proceedings of the 2008 American Control Conference, ACC*, pp. 1035–1039, USA, June 2008.
- [116] J. Wang, C. Rostoker, and A. Wagner, "A high performance pair trading application," in *Proceedings of the 2009 IEEE International Symposium on Parallel & Distributed Processing (IPDPS)*, pp. 1–8, Rome, Italy, May 2009.

