Topological Structure of Stock Market Networks during Financial Turbulence: Non-Linear Approach

Arash Sioofy Khoojine*, Han Dong

School of mathematics and statistics, Shanghai Jiao Tong University, China

Received November 01, 2019; Revised December 09, 2019; Accepted December 17, 2019

Copyright ©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract In this paper, researchers utilize mutual information and distance covariance to establish the minimum spanning tree of the financial network of log-returns and trading volumes of the top 96 companies of the United States stock market listed on S&P 100 index. Researchers analyze the United States stock market's turbulence during 2015-2016, employing the data from January 2012 to July 2018. For investigating the turbulence, researchers construct three minimum spanning trees of the pre-turbulence, turbulence and post-turbulence. The findings represent that the degree distribution follows the power law and the minimum spanning tree of pre-turbulence contains a notable difference in topological characteristics and network's measures such degree ratio, betweenness, closeness, eigenvector centrality, node eccentricity, node strength, node domination compared with turbulence and post-turbulence minimum spanning trees. Moreover, the minimum spanning trees constructed by two methods of mutual information and distance covariance are different in topological characteristics and the network's behavior. Besides, the pre-turbulence and post-turbulence networks are robust against nodes attack, and turbulence network is tenuous against it.

Keywords Statistical Network Analysis, Mutual Information, Distance Covariance, Symbolic Time Series, Financial Turbulence

1 Introduction

1.1 Brief Literature Review

In the real world, most of the complex systems have been represented by complex networks [19]. Intrinsically, the stock market has been explained as a complex system. There exist an intricate relationship between stocks, which causes price oscillation. During the last two decades, researchers have scrutinized stock markets by forming the stock correlation networks, of which the nodes represent stocks and edges between nodes

are price oscillation relationships of stocks [18].

Scrutinizing financial systems, particularly stock price markets, using the complex networks perspective has become one of the most widespread fields within econophysics. Additionally, a similar tendency is nowadays coming into sight within the econometrics and finance community researchers.

Classifying methodologically, various methods have been implemented to numerous stock markets, Heimo et al., and Gan et al. use Pearson's correlation coefficient to analyze the New York stock exchange [12, 10], Huang et al. use a threshold method to construct China's stock correlation network and analyze the network's structural properties and topological stabilities [15]. Tabak et al. investigate the topological properties of the Brazilian stock market networks by building the minimum spanning tree. Their results suggest stocks tend to cluster by sector [23]. Galazka studies network structure of the Polish Stock Market using Minimum Spanning Tree (MST) and Weighted Random Graph (WRG) and compares them [9]. Coronello et al. apply random matrix theory and hierarchical clustering techniques to a portfolio of stocks at the London Stock Exchange. Results have shown that the application of just a distinct method is not enough to extract all the economic information in the correlation coefficient matrix of a stock portfolio [3]. Zhuang et al. build a minimum spanning tree of CSI 300 index of the Shanghai and Shenzhen Stock markets and analyze its structure [25].

A few articles have been published about the behavior of networks during financial crises. Majapa and Gossel, look over the topological characteristics of financial networks before and after the 2008 financial crisis of South African stock market. They use correlations networks of the daily closing prices of the South African Top 100 companies from June 2003 to June 2013 to create a minimum spanning tree of before, during and after the financial crisis. The findings reveal that the network shrinks during the crisis period and expands afterwards [17]. Nobi et al., analyze the effects of the 2008 global financial crisis on financial networks of the Korean financial market around the crisis period. They contemplate the prices of stocks belonging to KOSPI 200 (Korea Composite Stock Price Index 200) for three periods, before, during and after the crisis. Threshold

networks are built from fully connected cross-correlation networks, and thresholds of cross-correlation coefficients are appointed to acquire threshold networks. Findings show threshold networks are fatter during the crisis than before or after the crisis [20]. Heiberger establishes a longitudinal network of S&P 500 companies and their correlations between 2000 and 2012 the stock network. The results show that stock network changes its composition and transforms into a more centralized topology [11].

1.2 Motivation

Computing correlations between different stocks is an attractive topic not only for grasping the statistical aspects of this type of complex systems but also for practical reasons like asset management and portfolio-risk estimation. Inspecting characteristics of financial networks and the behavior of them in time of crisis or market turbulence can assist us to consider a more effective strategy to deal with critical conditions and risk management of the stock investment.

Many researchers apply Pearson's correlation coefficient, which is linear inherently; therefore, it prevents us from capturing the nonlinear behavior of financial networks [14]. For this reason, Fiedor seeks mutual information-based approaches to Warsaw stock exchange and New York stock exchange, [6, 7], also, Tao et al. deploy information-theoretic distances to analyze the structure of Shanghai stock exchange [24]. Moreover, a well-accepted principle in the financial area obtain the significant role of traded volume on stock price movement; therefore, we decided to use traded volume alongside log-returns.

In this paper, we consider a network analysis of the turbulence of the United States stock market from 2015 to 2016. The United States stock market turbulence began on 18 August 2015 and ceased in July 2016. For tackling the non-linearity of log-returns, we apply information theoretic approach and distance covariance for constructing networks. Trading volumes are used at the side of the log-returns using the symbolic time series. This study is the first network analysis of the United States stock market's turbulence, using trading volume and log-return and exploiting information theory and distance covariance. For preventing a massive amount of linkage, we use the minimum spanning tree to represent the anatomy of networks.

In this paper, section one displays the introduction and data. The second part introduces the methodology and construction of networks based on mutual information distances and distance covariance. The third section indicates the result of these methods regarding minimum spanning trees. The conclusion section summarizes the paper's contributions and proposes further research ideas.

1.3 Data

The data employed in this paper are extracted from Yahoo! Finance and cross-checked with the Bloomberg database; therefore, the provided data is more completed than other available resources. Data includes 96 quoted companies of S&P 100 Index maintained by Standard & Poor's. This index represents a subset of S&P 500, contains 102 leading U.S. stocks. S&P

100 represents virtually 51% of the market capitalization of the United States equity market. Six companies are excluded from the database because of missing data.

We extracted closing prices, volumes from January of 2012 to the end of July 2018, totally 3309 working days of the stock exchange. The period of data is determined long because of inhibiting the effect of short-term volatility during financial turbulence.

We used the industry sectors taken from the macro-classification of [5]. Sectors are consumer discretionary, consumer staples, energy, financial, healthcare, industrial, IT and material. The data set is divided into three non-overlapping windows regarding the volatility of S&P 100 Index; pre-turbulence, turbulence, and post-turbulence. The division of these windows is based on the volatility of log-returns of the stock prices within these years. Three windows of considered periods are revealed in table 1.

Table 1. Non-overlapping windows of the dataset

	Start date	End date	Name
Windows 1	2012-01-01	2015-07-31	pre-turbulence
Windows 2	2015-08-01	2016-07-31	turbulence
Windows 3	2016-08-01	2018-08-01	post-turbulence

2 Methodology

In this section, to start with we introduce a symbolization of time series to reduce the dimension of our data, after that, we use mutual information and distance covariance-based distances to establish the financial networks. Moreover, we define some tools to analyze the networks which we create in the subsequent section.

2.1 Symbolization of Time Series

We transform the time series of closing prices of 96 stocks in the S&P 100 to the log ratios between sequential daily closing prices. Log-returns are as follows:

$$r_{t,i} = \ln(c_{t,i}) - \ln(c_{t-1,i}) \quad (1)$$

Where $c_{t,i}$ is the stock closing price of the company i at time t ; also, the traded money at day t is computed as:

$$\rho_{t,i} = c_{t,i} \cdot v_{t,i} \quad (2)$$

where $v_{t,i}$ is the trading volume of the company i at day t .

In this paper, we use the symbolization of time series analysis method for reducing the dimension of our data. The time series of each company is transformed into a set of natural numbers; therefore, a new sequence of numbers is initiated for each company. This process is called symbolization. Gabriel et al. apply the symbolization method to analyze currency markets [8]. For more discussion about symbolization see [4], and for selecting an appropriate partition see [21, 13].

For applying symbolization method, to begin with, we split an ordered sequence of log-returns into three parts; each part

contains a third of the whole time series of a company, and creates the 3-quantiles of each company, where r_{iT_1} is the first 3-quantile, and r_{iT_2} is the second 3-quantile of each company i . Time series of log-return and traded money are transformed into a sequence of numbers $s_{t,i}$ as follows:

$$s_{t,i} = \begin{cases} 1 & r_{t,i} \leq r_{iT_1} & \text{and} & \rho_{t,i} \leq \bar{\rho}_i \\ 2 & r_{t,i} < r_{iT_1} & \text{and} & \rho_{t,i} > \bar{\rho}_i \\ 3 & r_{iT_1} < r_{t,i} \leq r_{iT_2} & \text{and} & \rho_{t,i} \leq \bar{\rho}_i \\ 4 & r_{iT_1} \leq r_{t,i} < r_{iT_2} & \text{and} & \rho_{t,i} > \bar{\rho}_i \\ 5 & r_{t,i} \geq r_{iT_1} & \text{and} & \rho_{t,i} \leq \bar{\rho}_i \\ 6 & r_{t,i} > r_{iT_2} & \text{and} & \rho_{t,i} \geq \bar{\rho}_i \end{cases} \quad (3)$$

where $\bar{\rho}_i = 1/96 \sum_t \rho_{t,i}$ is the average of traded money of each company i . We transform log-returns and traded money of each company to a sequence of numbers, defined above and arrange it for each window defined at the table. 1 and extract three new time series of pre-turbulence, turbulence, and post-turbulence sequences.

2.2 Mutual Information Based Network

Pearson's correlation coefficient only seizes the linear interdependencies of the networks. In this section, to address this issue, we propose mutual information networks to find non-linear dependencies in financial markets. Mutual information is based on Shannon's entropy. Mathematically, Mutual information can be defined for two discrete random variables, X , and Y , which are the symbolized data created in the previous section, as:

$$I(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (4)$$

where $p(x, y)$ is the joint probability distribution function of X and Y , $p(x)$ and $p(y)$ are the marginal probability distributions.

Mutual information can be demonstrates as follows as well:

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad (5)$$

where $H(X)$ is Shannon's entropy, which is defined as:

$$H(X) = - \sum_i p(x_i) \log p(x_i) \quad (6)$$

and $H(X, Y)$ is the joint entropy associated with both variables.

In this paper, we use Schürmann-Grassberger estimator which is a Bayesian parametric estimator to estimate Shannon's entropy. This estimator is derived by Dirichlet probability distribution. Its density function describes as follows:

$$p(x; \theta) = \frac{\prod_{i \in \{1, 2, \dots, |\chi|\}} \Gamma(\theta_i)}{\Gamma(\sum_{i \in \{1, 2, \dots, |\chi|\}} \theta_i)} \prod_{i \in \{1, 2, \dots, |\chi|\}} x_i^{\theta_i - 1} \quad (7)$$

where θ_i is the prior probability of an event x_i which is the i^{th} element of the set χ and $\Gamma(\cdot)$ is the gamma function. The entropy of a Dirichlet distribution can be calculated with the following equation:

$$\hat{H}^{dir}(X) = \frac{1}{m + |\chi|N} \sum_{x \in \chi} (D_x + N)(\psi(m + |\chi|N + 1) - \psi(D_x + N + 1)) \quad (8)$$

where D_x is the number of data points having value x , $|\chi|$ is the number of bins from the discretization step, m is the sample size, and $\psi(z) = \frac{d}{dz} \ln(\Gamma(z))$ is the digamma function. The Schürmann-Grassberger estimator assumes $N = \frac{1}{|\chi|}$ as a prior parameter.

Subsequently, we transfer mutual information to a Euclidean metric which attains the axioms of a metric distance as follows:

$$\delta(X, Y) = H(X, Y) - I(X, Y) \quad (9)$$

2.3 Distance Covariance Based Network

As mentioned in the previous section, Pearson's correlation coefficient is principally vulnerable to discover the linear relationship between two variables and could not capture the non-linear dependencies, therefore; we efficiently utilize distance covariance, initiated by Székely and Rizzo to measure non-linear statistical interdependence between two random vectors reliably [22].

Distance covariance between the column vector of p random variables $\mathbf{X}$ and a column vector of q random variables $\mathbf{Y}$, which are the symbolized data created at section 2.1., defined as follows:

$$\nu^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{c_q c_p} \int_{\mathbb{R}^{p+q}} \frac{|\mathcal{F}_{\mathbf{XY}}(t, s) - \mathcal{F}_{\mathbf{X}}(t) \mathcal{F}_{\mathbf{Y}}(s)|^2}{|t|^{1+p} |s|^{1+q}} dt ds \quad (10)$$

where $F_{\mathbf{X}}$, $F_{\mathbf{Y}}$ and $F_{\mathbf{XY}}$ are the characteristic functions of $\mathbf{X}$, $\mathbf{Y}$ and joint characteristic function of $\mathbf{X}$ and $\mathbf{Y}$ respectively, $t \in \mathbb{R}^p$ and $s \in \mathbb{R}^q$. Also c_p or c_q are defined as follows:

$$c_d = \pi^{\frac{1+d}{2}} / \Gamma(\frac{1+d}{2}) \quad (11)$$

where $\Gamma(\cdot)$ is the Gamma function. We use a standardized version of the distance covariance, distance correlation, which is defined as follows:

$$R^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \frac{\nu^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\nu^2(\mathbf{X})\nu^2(\mathbf{Y})}} & \nu^2(\mathbf{X})\nu^2(\mathbf{Y}) > 0 \\ 0 & \nu^2(\mathbf{X})\nu^2(\mathbf{Y}) = 0 \end{cases} \quad (12)$$

2.4 Tools of the Network Analysis

To analyze the financial networks constructed in previous sections, we need some convenient tools to illuminate the topological characteristics of them. These measures are borrowed from graph theory which empowers us to obtain a more thorough understanding of the structure of financial complex systems.

1. We construct two correlation matrices based on the measure created by (9) and (12), we calculate these measures for each pair of stocks and put the diagonal of these matrices equals to zero.

2. Minimum spanning tree (MST) is an appropriate implement to purify negligible information of the network to contemplate it favorably. Properly using the distance matrices constructed by (9) and (12), we can build a minimum spanning tree (MST) with Kruskal's algorithm [16].
3. Suppose $A = [a_{ij}]_{N \times N}$ is the adjacency matrix of correlations between the stocks, the degree of a company i which is the number of edges incident to a company in the financial network can be calculated as $d_i = \sum_{j=1}^N a_{ij}$. Degree ratios of a financial network defined as follow:

$$p_k = \frac{1}{N} \sum_{j=1}^N I_k(d_j) \quad (13)$$

where $k = 0, \dots, N - 1$ and $I_k(\cdot)$ is the function with $I_k(x) = 1$ if $x = k$, otherwise $I_k(x) = 0$.

4. Node eigenvector measures the centrality of a company and its neighbor companies with calculating the eigenvector of the principal eigenvalues of the adjacency matrix [1, 2].
5. Node betweenness measures the role of a company as a link among other paired companies of a network. Betweenness centrality of a node i , $\beta(i)$ is defined as:

$$\beta(i) = \sum_{i \neq j \neq k} \frac{\mathbf{l}_{jk}(i)}{\mathbf{l}_{jk}} \quad (14)$$

where $\mathbf{l}_{jk}(i)$ is the number of paths from j to k that pass through i and $\mathbf{l}_{jk}$ is the number of paths between nodes j and k .

6. Node eccentricity measures the maximum number of edges in the shortest paths between a company and any other company in a financial network.
7. Node strength typically measures the sum of the weights of the edges connected to a company, mathematically for a company i it is $\sum_{i \neq j} \delta_{ij}$ where δ_{ij} s are the distances defined at (9) and (12).
8. Closeness centrality measures the inverse of the sum of all distances to other companies in a financial network i.e. $1 / \sum_{i \neq j} d_{ij}$.
9. Domination strength measures which one of the companies or sectors most characteristically affect the network. It is computed as follows:

$$\Delta(i) = \sum_{j=1}^n (\lambda(i, j) / \tau(j)) \quad (15)$$

Where $\lambda(i, j)$ indicates the weight of joints linking the node i , while $\tau(j)$ is the power density of the node j defined as follows:

$$\tau(j) = \sum_{i=1}^n (\lambda(i, j)) \quad (16)$$

3 Results

This section exhibits the behavior of the minimum spanning trees regulated on the top 96 companies by market capitalization on the S&P 100 daily closing prices and traded volumes between January 2012, and July 2018 calculated by logarithmic returns. The analysis in this section is based on the result of the correlation and distance matrices built by (9) and (12), as well as the network analysis tools defined in section 2.4.

Before, during and after the financial turbulence, the data set is split into three subdivisions: A pre-turbulence subdivision running from the beginning of January 2012 to 31st July 2015; a turbulence subdivision running from first August 2015 to 31st July 2016; and a post-turbulence subdivision, running from early August 2016 to July 31st, 2018. Names, sectors and related color for each sector of a company are shown in Appendix 2.

3.1 Results of Mutual Information Based MSTs

Analyzing the topological structure of the MST of pre-turbulence depicts in Fig. 22 (Appendix 1) shows that the companies, 3M Company (MMM), Honeywell(HON) and Citigroup Inc (C) carry out a pivotal role in the MST of pre-turbulence and establish a hub. In this MST, most of the sectors are connected or in a similar region. Fig. 23 (Appendix 1) shows that the companies MasterCard Inc (MA), 3M Company (MMM) and Citigroup Inc (C) comprise the most leading companies in turbulence MST and create a hub; also, the majority of companies are grouped by sectors. As seen in Fig. 24 (Appendix 1) Microsoft (MSFT), 3M Company (MMM), Bank of America Corp (BAC) and Pfizer Inc (PFE) are the most influential companies in post-turbulence MST. As seen in these three MSTs, the shapes of them are similar, but the pivotal companies are changed during various periods.

Figure 1. Node degree ratio of MSTs based on MI, pre-turbulence 1a, turbulence 1b, post-turbulence 1c

Fig. 21 shows that the degree ratios in three periods compose a direct line against the node degree; therefore, they hold the power law and possess the scale-free structure. The majority of companies have the degree of one, two and three and slightly change during distinct periods.

Figure 2. Eigenvector centrality of MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

Figure 3. Average eigenvector centrality of sectors for MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

A high eigenvector score of a company demonstrates it is joined to the companies that themselves have high eigenvector scores. As shown in Fig. 2 the distributions of eigenvectors centralities of pre-turbulence, turbulence and post-turbulence are approximately indistinguishable. The highest eigenvector centrality of pre-turbulence, turbulence and post-turbulence periods be in the hands of 3M Company Energy (MMM), MasterCard Inc (MA) and Microsoft (MSFT) respectively.

Analyzing the performance of sectors during three periods encourages us to obtain an adequate understanding of the interconnectedness of the constructed MSTs. Fig. 3 shows that the various sectors maintain peculiar behavior during three periods. It shows pre-turbulence period has more sectors with the high eigenvector average; therefore, during pre-turbulence companies with the high eigenvector centrality tend to connect with companies with high eigenvector centrality.

Figure 4. Closeness centrality of MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

Figure 5. Average closeness centrality of sectors for MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

Closeness centrality score in a complex network represents that the more pivotal a company is, the more adjacent to all other companies. Fig. 4 shows that closeness scores of the companies during pre-turbulence are higher than the closeness of two various periods, in other words, the companies are nearer to each other during the pre-turbulence period. Honeywell (HON) Morgan Stanley (MS) and 3M Company (MMM) obtain the highest score of closeness. Fig. 5 shows that the closeness score of the sectors during pre-turbulence is less than the other two periods.

Figure 6. Betweenness centrality MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

Figure 7. Average betweenness centrality of sectors for MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

As shown in Fig. 6, betweenness centrality of MST during the turbulence and post turbulence are alike and marginally different from the pre-turbulence period. The betweenness centrality of 50 companies is zero during pre-turbulence this amount decreases to 47 and 44 companies during turbulence and post-turbulence periods respectively. Zero betweenness centrality score denotes that no shortest path between any two companies in the network passes through those companies. The highest score of betweenness belongs to Honeywell (HON) MasterCard Inc (MA) and 3M Company (MMM). Fig. 7 shows that average betweenness during the turbulence period is high in most of the sectors compared with other periods.

Figure 8. Node strength centrality of MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

Figure 9. Average Node strength centrality of sectors for MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

As seen in Fig. 8, the strength of nodes during the pre-turbulence period is approximately higher than turbulence and post-turbulence periods. Furthermore, Fig. 9 shows that majority of sectors have higher average node's strength during the pre-turbulence period than other periods. shows that average betweenness during the turbulence period is high in most of the sectors compared with other periods.

Figure 11. Domination centrality of MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

Figure 10. Node eccentricity of MSTs based on MI, during pre-turbulence, turbulence and post-turbulence.

As depicts in Fig. 10 pre-turbulence and post-turbulence domination strengths are approximately alike, but they have an impalpable difference with the turbulence period. The most valuable companies during pre-turbulence, turbulence, and post-turbulence are 3M Company (MMM), Alphabet Inc (GOOGL) and Microsoft (MSFT) respectively.

Fig. 11 shows that the eccentricity of the companies during pre-turbulence and post-turbulence is smaller than the MST of turbulence, therefore, the maximum number of edges in a short path of the companies in turbulence is higher than the maximum number of edges in a short path of the companies in the other two periods.

3.2 Results of Distance Covariance Based MSTs

In this section, we scrutinize the results of MSTs which are constructed by (12) with helping the tools defined in section 2.4. As shown in Fig. 25 (Appendix 1) companies gathered around a few dominant companies, also the sectors scattered around the MST. Fig. 26 (Appendix 1) shows that Southern Company (SO) is the most pivotal company in this MST, and the majority of companies try to connect to it. As seen in Fig. 27 (Appendix 1) the shape of the MST of post-turbulence is different from the MST of the turbulence period and similar to the pre-turbulence MST.

(a)

(b)

(c)

Figure 12. Node degree ratio of distance covariance based MSTs, pre-turbulence 12a, turbulence 12b, post-turbulence 12c

As seen in Fig. 12, most of the companies have a small node degree of one; also, the considerable number of companies with the degree one decreases during the turbulence and post-turbulence period. Netflix (NFLX), has the highest node degree in the pre-turbulence period and Southern Company (SO) has the highest node degree in turbulence and post-turbulence periods.

Figure 13. Eigenvector centrality of MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence.

Figure 14. Average eigenvector centrality of sectors for MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence.

Fig. 13 and Fig. 14 depict that the post-turbulence period is different from other periods and the average eigenvector of the sectors during pre-turbulence is greater than other periods. It means that companies with a high eigenvector score try to connect with the companies with a higher eigenvector score during the pre-turbulence period. Netflix (NFLX) has the highest score in the pre-turbulence period, and Southern Company (SO) has the highest score during the turbulence and the post-turbulence periods.

Figure 15. Closeness centrality of MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence.

Figure 16. Average closeness centrality of sectors for MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence.

As seen in Fig. 15 and Fig. 16, post-turbulence closeness

scores are smaller than other periods; also the average closeness of most of the sectors in the turbulence period is greater than others; therefore, companies during the turbulence period are close together. Netflix (NFLX) has the highest closeness score in the pre-turbulence period, and Southern Company (SO) has the highest closeness score during the turbulence and post-turbulence periods.

Figure 17. Betweenness centrality of MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence.

Figure 18. Average betweenness centrality of sectors for MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence.

As shown in Fig. 17 and Fig. 18, betweenness score of companies in post-turbulence period is greater than other periods; also, the companies in the utilities sector have a greater average among other sectors in all periods. Netflix (NFLX) has the highest betweenness score in the pre-turbulence period, and Southern Company (SO) has the highest betweenness score during turbulence, and post-turbulence periods.

Figure 19. Node strength of MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence.

Figure 20. Average Node strength of sectors for MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence.

As seen in Fig. 19 the strength of nodes during all periods are approximately similar, but Fig. 20 gives us a better view of MSTs, which confirms that more than 50% of sectors in turbulence period has a higher average compared with other periods.

Figure 21. Node eccentricity and domination centrality of MSTs based on distance covariance, during pre-turbulence, turbulence and post-turbulence. Node domination centrality 21a, Node eccentricity 21b

As depicted in Fig. 21a, domination strengths of all periods are approximately similar; also, Southern Company (SO) has the highest score among all companies in all the periods. Fig. 21b shows that the eccentricity of the companies during post-turbulence is higher than the MST of other periods; therefore, the maximum number of edges in a short path of the companies in post-turbulence is smaller than the maximum number of edges in a short path of the companies in the other two periods.

4 Topological stability of MSTs

Topological stability of financial networks typically is measured by removing vertices and edges from the networks. If the characteristics of a financial network are evolved drastically against vertex attack and edge attack, then the network is not robust against those attacks. Randomly, a different percentage of vertices is removed to evaluate the vertex attack in characteristics of the constructed financial network. The similar procedure to evaluate the topological stability of a network can be found at [15].

Definition (Maximal Connected Component): A financial network F is connected if there is a path from any stock to any other stocks in the financial network. A disconnected network can be broke down into a variety of connected sub-networks, such as F' , which is referred to as connected components of F . If F' is a maximal connected component of F then F'' is a sub-network of F and if F'' is also a connected sub-network of F , then $F'' = F'$.

In this paper, we deploy the size of the maximal connected component of a financial network to discover the connectivity of the financial network. Consequently, we evaluate the stability of networks with deploying the changes of the maximum connected component size before and after the vertex attack during pre-turbulence, turbulence and post-turbulence.

Results in Table. 2 show that the size of the maximal connected component of the MSTs of the pre-turbulence and post-turbulence with the mutual information and distance covariance methods have a slight change after removing a different percentage of stocks compared with turbulence period. Additionally, after removing 40% of companies, this size decreases significantly at the MSTs of pre-turbulence and post-turbulence. In closing, the topological stability of pre-turbulence and post-turbulence MST with MI and DC methods are more than turbulence period, and their MSTs are robust against vertex attacks.

5 Conclusions

In this paper, the structure of financial turbulence of the United States stock market occurred during 2015 to 2016 is analyzed with constructing the minimum spanning the tree of the top 96 companies of S&P 100 index for the years of 2012 to July 2018. Time-series of log-returns and trading volume of each company in this index is transformed into symbolized data than the two methods of mutual information, and distance covariance is used to construct the financial networks. Both of the derived measures from mutual information and distance

Table 2. Maximally component size of MSTs in the absence of percentage of degrees, during pre-turbulence, turbulence and post-turbulence for Mutual Information (MI) and Distance Covariance (DC).

Percentage	Size of Maximally Connected Component					
	Pre-Turbulence		Turbulence		Post-Turbulence	
	DC	MI	DC	MI	DC	MI
10	76	81	52	57	77	82
20	82	86	61	66	83	86
30	81	84	59	67	80	83
40	83	85	58	62	82	83
50	67	72	46	50	68	73
60	68	73	45	52	69	75
70	68	74	48	51	66	73
80	69	76	49	52	68	77

covariance follow the metric distance axioms. After constructing the correlation matrices of these two distances, the MSTs of pre-turbulence, turbulence, and post-turbulence are created.

Different network measures derived from MSTs for each method, and the results show that the degree distribution follows power law also there are significant differences between the topological structure and the network measures of the pre-turbulence period such as degree ratio, betweenness, closeness, eigenvector centrality, node eccentricity, node strength, node domination compared with the measures of the other two periods. However, the more essential companies are obtained for each period.

The behavior of sectors observed with average eigenvector, average betweenness, average closeness, and average node strength. These measures helped us to have a better understanding of the inter-connectedness relationship of MSTs.

Comparing the results derived from sections 3.1 and 3.2 shows that the two methods of building the MSTs, MI and distance covariance have a significant difference to analyze the complex financial networks. The MSTs built by two ways have a completely different structure and shape; moreover, the pivotal companies are different in each method. However, in both modes, the pre-turbulence MSTs have different results compared with the MSTs of the other two periods. Moreover, removing random companies shows that correlation networks of pre-turbulence and post-turbulence display the topological robustness, but turbulence is fragile against randomly node's failure.

Further research could look into choosing the best method to analyze the financial networks non-linearly. However, different symbolization methods of time series can be applied and be compared.

Acknowledgements

This research was Supported by National Basic Research Program of China (973 Program, 2105CB856004) and National Natural Science Foundation of China (11531001).

REFERENCES

- [1] P. Bonacich. Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, 2(1):113–120, 1972.
- [2] P. Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29:555–564, 2007.
- [3] C. Coronello, M. Tumminello, F. Lillo, S. Miccichè, and R. N. Mantegna. Sector identification in a set of stock return time series traded at the London Stock Exchange. *Acta Physica Polonica B*, 36(9):2653–2679, 2005.
- [4] C. S. Daw and C. E. a. Finney. A review of symbolic analysis of experimental data. *Review of Scientific Instruments*, 74(2):915–930, 2003.
- [5] E. F. Fama and K. R. French. Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193, feb 1997.
- [6] P. Fiedor. Networks in financial markets based on the mutual information rate. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 89(5):1–7, 2014.
- [7] P. Fiedor. Mutual information-based hierarchies on Warsaw Stock Exchange. *Acta Physica Polonica A*, 127(3):A33–A37, 2015.
- [8] J. Gabriel, D. Matesanz, and W. Adrián. Symbolic hierarchical analysis in currency markets : An application to contagion in currency crises q. *Expert Systems with Applications*, 36:7721–7728, 2009.
- [9] M. Galazka. Characteristics of the Polish Stock Market correlations. *International Review of Financial Analysis*, 20(1):1–5, 2011.
- [10] S. L. Gan and M. A. Djauhari. New York Stock exchange performance: Evidence from the forest of multidimensional minimum spanning trees. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(12):P12005, 2015.
- [11] R. H. Heiberger. Stock network stability in times of crisis. *Physica A: Statistical Mechanics and its Applications*, 393:376–381, 2014.
- [12] T. Heimo, J. Saramäki, J. P. Onnela, and K. Kaski. Spectral and network methods in the analysis of correlation matrices of stock returns. *Physica A: Statistical Mechanics and its Applications*, 383(1 SPEC. ISS.):147–151, 2007.
- [13] Y. Hirata, K. Judd, and D. Kilminster. Estimating a generating partition from observed time series: Symbolic shadowing. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 70(1):11, 2004.
- [14] D. A. Hsieh. Chaos and Nonlinear Dynamics: Application to Financial Markets. *The Journal of Finance*, 46(5):1839, dec 1991.
- [15] W. Q. Huang, X. T. Zhuang, and S. Yao. A network analysis of the Chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 388(14):2956–2964, 2009.
- [16] J. B. Kruskal. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*, 7(1):48, feb 1956.

- [17] M. Majapa and S. J. Gossel. Topology of the South African stock market network across the 2008 financial crisis. *Physica A*, 445:35–47, 2016.
- [18] R. Mantegna. Hierarchical structure in financial markets. *European Physical Journal B*, 11(1):193–197, 1999.
- [19] M. E. J. Newman. Hierarchical structure in financial markets. *SIAM Review*, 45(1):167–256, 2003.
- [20] A. Nobi, S. E. Maeng, G. G. Ha, and J. W. Lee. Effects of global financial crisis on network structure in a local stock market. *Physica A: Statistical Mechanics and its Applications*, 407(xxxx):135–143, 2014.
- [21] C. Piccardi. On parameter estimation of chaotic systems via symbolic time-series analysis. *Chaos*, 16(4), 2006.
- [22] G. J. Székely and M. L. Rizzo. Brownian distance covariance. *Annals of Applied Statistics*, 3(4):1236–1265, 2009.
- [23] 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*, 389(16):3240–3249, 2010.
- [24] T. You and A. Fiedor Paweł and Hołda. Network Analysis of the Shanghai Stock Exchange Based on Partial Mutual Information. *Journal of Risk and Financial Management*, 8(4):266–284, 2015.
- [25] R. Zhuang, B. Hu, and Z. Ye. Minimal spanning tree for Shanghai-Shenzhen 300 stock Index. *2008 IEEE Congress on Evolutionary Computation, CEC 2008*, pages 1417–1424, 2008.

Appendix 1

Figure 22. MST of Pre-Turbulence based on MI. Colors represent sectors (see table 3).

Figure 23. MST of Turbulence based on MI. Colors represent sectors (see table 3).

Figure 24. MST of Post-Turbulence based on MI. Colors represent sectors (see table 3).

Figure 25. MST of Pre-Turbulence based on Correlation Distance. Colors represent sectors (see table 3).

Figure 26. MST of Turbulence based on Correlation Distance. Colors represent sectors (see table 3).

Figure 27. MST of Post-Turbulence based on Correlation Distance. Colors represent sectors (see table 3).

Appendix 2

Table 3. Companies, sectors and the corresponding colors

Sector	Number of companies	Color
Consumer Discretionary	15	Mint
Consumer Staples	11	Lavender
Energy	7	Red
Financials	14	Green
Health Care	14	Yellow
Industrials	13	Orange
IT	17	Brown
Materials	1	Grey
Utilities	4	Blue

Symbol	Company Name	Sector
AAPL	Apple Inc.	Information Technology
ABT	Abbott Laboratories	Health Care
ACN	Accenture plc	Information Technology
AGN	Allergan plc	Health Care
AIG	American International Group Inc.	Financials
ALL	Allstate Corp.	Financials
AMGN	Amgen Inc.	Health Care
AMZN	Amazon.com	Consumer Discretionary
AXP	American Express Inc.	Financials
BA	Boeing Co.	Industrials
BAC	Bank of America Corp	Financials
BIIB	Biogen	Health Care
BK	The Bank of New York Mellon	Financials
BLK	BlackRock Inc	Financials
BMJ	Bristol-Myers Squibb	Health Care
C	Citigroup Inc	Financials
CAT	Caterpillar Inc	Industrials
CELG	Celgene Corp	Health Care
CHTR	Charter Communications	Consumer Discretionary
CL	Colgate-Palmolive	Consumer Staples
CMCSA	Comcast Corporation	Consumer Discretionary
COF	Capital One Financial Corp.	Consumer Finance
COP	ConocoPhillips	Energy
COST	Costco	Consumer Staples
CSCO	Cisco Systems	Information Technology
CVS	CVS Health	Consumer Staples
CVX	Chevron Corporation	Energy
DHR	Danaher Corporation	Health Care
DIS	The Walt Disney Company	Consumer Discretionary
DUK	Duke Energy	Utilities
DWDP	DowDuPont	Materials
EMR	Emerson Electric Co.	Industrials
EXC	Exelon	Utilities
F	Ford Motor	Consumer Discretionary
FDX	FedEx	Industrials
FOX	21st Century Fox	Consumer Discretionary
FOXA	21st Century Fox	Consumer Discretionary
GD	General Dynamics	Industrials
GE	General Electric	Industrials
GILD	Gilead Sciences	Health Care
GM	General Motors	Consumer Discretionary

Symbol	Company Name	Sector
GOOG	Alphabet Inc	Information Technology
GOOGL	Alphabet Inc	Information Technology
GS	Goldman Sachs	Financials
HAL	Halliburton	Energy
HD	Home Depot	Consumer Discretionary
HON	Honeywell	Industrials
IBM	International Business Machines	Information Technology
INTC	Intel Corporation	Information Technology
JNJ	Johnson & Johnson Inc	Health Care
JPM	JP Morgan Chase & Co	Financials
KMI	Kinder Morgan	Energy
KO	The Coca-Cola Company	Consumer Staples
LLY	Eli Lilly and Company	Health Care
LMT	Lockheed Martin	Industrials
LOW	Lowe's	Consumer Discretionary
MA	MasterCard Inc	Information Technology
MCD	McDonald's Corp	Consumer Discretionary
MDLZ	Mondel?z International	Consumer Staples
MDT	Medtronic Inc.	Health Care
MET	Metlife Inc.	Financials
MMM	3M Company	Industrials
MO	Altria Group	Consumer Staples
MRK	Merck & Co.	Health Care
MS	Morgan Stanley	Financials
MSFT	Microsoft	Information Technology
NEE	NextEra Energy	Utilities
NFLX	Netflix	Information Technology
NKE	Nike	Consumer Discretionary
NVDA	Nvidia Corporation	Information Technology
ORCL	Oracle Corporation	Information Technology
OXY	Occidental Petroleum Corp.	Energy
PEP	PepsiCo	Consumer Staples
PFE	Pfizer Inc	Health Care
PG	Procter & Gamble Co	Consumer Staples
PM	Phillip Morris International	Consumer Staples
QCOM	Qualcomm Inc.	Information Technology
RTN	Raytheon Company	Industrials
SBUX	Starbucks Corporation	Consumer Discretionary
SLB	Schlumberger	Energy
SO	Southern Company	Utilities
SPG	Simon Property Group, Inc.	Real Estate
T	AT&T Inc	Telecommunication Services
TGT	Target Corporation	Consumer Discretionary
TXN	Texas Instruments	Information Technology
UNH	UnitedHealth Group	Health Care
UNP	Union Pacific Corporation	Industrials
UPS	United Parcel Service	Industrials
USB	U.S. Bancorp	Financials
UTX	United Technologies	Industrials
V	Visa Inc.	Information Technology
VZ	Verizon Communications	Telecommunication Services
WBA	Walgreens Boots Alliance	Consumer Staples
WFC	Wells Fargo	Financials
WMT	Walmart	Consumer Staples