Physica A 505 (2018) 903–918

Contents lists available at ScienceDirect

Physica A

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

Investigating the features of pairs trading strategy: A network perspective on the Chinese stock market

Danyan Wen [a], Chaoqun Ma [a,*], Gang-Jin Wang [a,*], Senzhang Wang [b]

[a] Business School, Hunan University, Changsha 410082, China
[b] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

HIGHLIGHTS

- A research framework is proposed based on cointegration theory and complex network.
- The static and dynamic features of pairs trading strategy are examined.
- A few number of linkages between potential assets survived from one time to the next.
- Stocks within the same industry are inclined to form cointegration relationships.
- The re-occurrence circles of edges indicate that the trading pairs are unsteady.

ARTICLE INFO

Article history:
Received 8 February 2018
Received in revised form 27 March 2018
Available online 13 April 2018

Keywords:
Pairs trading
Complex network
Community
Margin-trading and short-selling
Cointegration
Dynamics

ABSTRACT

We systematically investigate the features of pairs trading strategy from a network perspective. Based on the cointegration theory, we construct the full graphs (FGs) and minimum spanning trees (MSTs) in terms of the cointegration matrix by using daily stock prices from the Chinese stock market around the launch of the margin-trading and short-selling (MTSS) programme. The static and dynamic features of the networks are analyzed with respect to the formation and evolution of edges. Our results show (i) that the assets allowed MTSS or within the same industry are more inclined to form trading pairs, (ii) that the cointegration relationships among trading pairs are as fragile as only a few number of them survived from one period to the next, (iii) that the movement of cointegration relationships is independent to stock market trends, confirming that pairs trading is a market neutral strategy, and (iv) that over 45% trading pairs are excluded in the re-occurrence circles of edges within 12 months, indicating that a considerable number of trading pairs are unsteady and investors should carefully pick pairs for trading in such short-term speculations.

1. Introduction

On 31 March 2010, China launched a long-awaited pilot scheme to lift the ban on margin-trading and short-selling (MTSS),[1] allowing 90 stocks to be purchased on margin or/and sold short. The list of underlying stocks has changed over time, and expanded to 971 constituent stocks (excluding exchange-traded-funds, ETFs) by the end of 2016. It introduced

* Correspondence to: Business School of Hunan University, 11 Lushan South Road, Yuelu District, Changsha 410082, China.
E-mail addresses: cqma1998@hnu.edu.cn (C. Ma), wanggangjin@hnu.edu.cn (G.-J. Wang).

1 Margin-trading and short-selling (MTSS) programme is widely referenced as "Rong Zi Rong Quan" in Chinese, which means that investors can buy stocks on margin and sell them shortly.

https://doi.org/10.1016/j.physa.2018.04.021
0378-4371/

a completely new trading mechanism and was widely regarded as a milestone for the development of the Chinese stock market. Under this background, a large number of rarely-used strategies in the Chinese stock market are employed to enhance the profitability of investments, where one of the typically used strategies is pairs trading.

As a classical quantitative investment strategy, pairs trading strategy or statistical arbitrage was first developed and put into practice by Nunzio Tartaglia, while working for Morgan Stanley in the 1980s [1]. Thereafter, it has been widely employed by institutional investors and professional traders in financial markets. The general rule of pairs trading strategy is, that one should seek two stocks whose prices have long-run equilibrium first, and then trace them all through a period. When their prices are deviating from the equilibrium abnormally, namely the short-term deviation, the investor should take long and short positions simultaneously. As the spread narrows back to the equilibrium, a profit results. Thus, the prices of the two stocks are expected to converge to a mean in the future [2].

Pairs trading can be generally employed in two steps, the selection of asset pairs and the seeking of trading opportunities. Presently, there are four methodologies for the selection of trading pairs, including the minimum sum of distance square [1], the stochastic spread [3], the relative coefficient, and the cointegration theory [4]. Among these four methodologies, the cointegration theory cannot only describe the price trends, but also accurately measure the extent of the price difference that deviates from the long-run equilibrium [5]. Due to the above two advantages, the cointegration theory has been widely applied in pairs trading for stock selection [2,4,6]. The employment of cointegration theory in pairs trading dates back to 1980s. Granger in 1981 [7] and Granger & Weiss in 1983 [8] have precisely defined the concept of cointegration on asset prices. Subsequently, a strict prove of the integration theory and the operation steps of pairs trading are provided by Granger and Engle [9,10]. Since then, a growing number of literature provides empirical studies based on the cointegration theory. Lin et al. [5] introduce the cointegrated methodology into the Australia stock market and achieve the excess return consecutively. By adopting this methodology, Hong and Raul [11] find that pairs trading in the U.S. stock market delivers significant profits. Moreover, some researchers employ this methodology to investigate the properties of pairs trading in other stock markets [12,13].

With respect of the cointegration relationships, previous studies usually focus on the theoretical analysis or the profitability of several specific trading pairs while ignoring the macro-analysis of all the potential asset pairs. Since the stocks in a financial system are usually correlated and heavily synchronized [14,15], it is necessary to systemically investigate a large number of stocks for the better selection of portfolios. On the one hand, for the investors, it is possible that a reasonable portfolio constructed by part of the asset pairs may be unreasonable from the perspective of the whole system and the whole time period. On the other hand, with the rapid development of algorithmic trading systems, various portfolios using the pairs trading strategies are applied by computer programs, requiring that the synchronization of asset pairs and portfolios containing pairs trading should be examined carefully.

To achieve such objectives, we try to examine the cointegration relationships systematically among stocks using the assets listed in the MTSS programme of the Chinese stock market. We propose the use of the complex network theory to deal with this problem for two reasons. On the one hand, the cointegration relationship can be viewed as a special kind of synchronization movement from the network perspective [16]. On the other hand, the complex network theory provides us with many effective tools to measure the properties of systems [17]. Therefore, we propose a framework to form the pairs trading networks based on the cointegration theory and complex networks for analysis. Concretely, we first construct the cointegration matrix using correlation efficient as weights mainly based on the cointegration approach. Based upon the weighted cointegration matrix, we build the full graphs (FGs) and the minimum spanning trees (MSTs) of stock collections. Then, we analyze the cointegration networks (including the FGs and MSTs) by focusing on two problems, (i) how the cointegration relationships form and (ii) how they evolve. To sum up, our contributions can be summarized as follows.

(i) We propose a research framework based on daily stock prices by using the cointegration networks to investigate the static and dynamic features of pairs trading from a systematic perspective.
(ii) Our results show that the stocks in the MTSS programme or the stocks within the same industry are more inclined to form cointegration relationships, but only a few of them can survive from one period to the next.
(iii) The movement of the cointegration networks indicates that pairs trading is market neutral and the movement of edges suggests that there are a considerable number of unsteady trading pairs that may enhance the risk of pairs trading.

The rest of this paper is organized as follows. Section 2 introduces the methodology used in this paper. Section 3 provides the basic statistics of the constructed cointegration networks. Section 4 shows the analysis on the static and dynamic features of the cointegration networks. Finally, Section 5 concludes this paper.

2. Methodology

In this section, we first introduce the research framework of this study. Then, we provide the methodologies on how to construct the trading pairs and how to construct the cointegration networks, e.g., the full graphs (FGs) and minimum spanning trees (MSTs). Based on this, we finally propose some measures for analyzing the static and dynamic features of the cointegration networks.

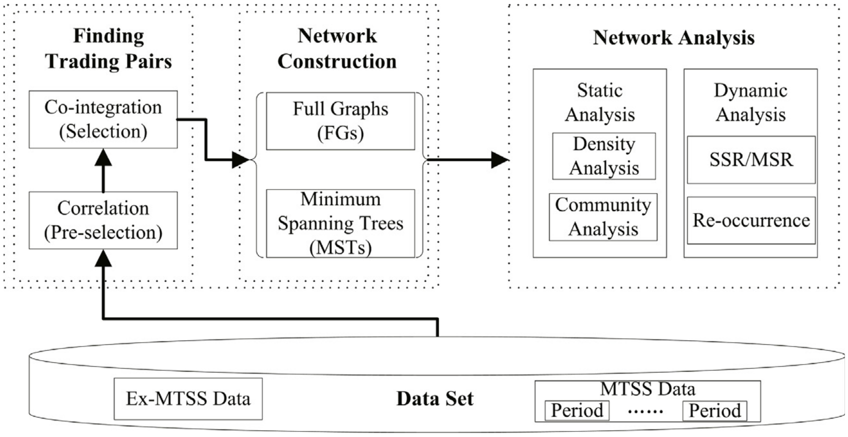


Fig. 1. Research framework.

2.1. The research framework

In order to better understand the static and dynamic features of pairs trading strategy from the perspective of networks, we propose a synthesis framework including the construction and analysis processes of the cointegration networks.

We present our research framework in Fig. 1 inspirited by Liu et al. [17]. There are four main parts in the framework: data set, finding trading pairs, network construction, and the analysis on the constructed networks. The data set part is the basis of our research; the second and third parts can be regarded as the process of constructing the cointegration networks; and the last part is the analysis on the constructed networks. The data flowing process of our research can also be seen from the figure. We first apply the correlation and cointegration methods successively to select the trading pairs based on the daily stock prices. Then, we construct the FGs and MSTs based on these trading pairs. Finally, we analyze the cointegration networks statically and dynamically to investigate the interpretation of edges' formation and evolution. The four parts to construct the framework are further described in detail as follows.

(i) Data set. We use the daily stock prices before and after the announcement of MTSS programme to investigate the pairs trading strategy.
(ii) Finding trading pairs. We construct the trading pairs by using the cointegration theory. Moreover, we employ the correlation statistic to pre-select the potential asset pairs.
(iii) Network construction. We construct the FGs and MSTs based on the cointegration matrix for analysis.
(iv) Network analysis. At last, we focus on investigating the features of trading pairs by mining the static and dynamic properties of the FGs and MSTs. Specifically, we conduct the density analysis and community analysis to mine the static properties. Moreover, we also use the single-step surviving rate (SSR), multi-step surviving rate (MSR) and the re-occurrence of edges to investigate the dynamic properties of the networks.

2.2. Pairs trading strategy

The basic process of pairs trading is that investors should identify two potential stocks that move together first and then take long and short positions simultaneously when their prices diverge abnormally. Here we attempt to construct the pairs trading portfolios mainly based on the cointegration approach. Specifically, we use the correlation coefficients to pre-select the potential asset pairs by setting a threshold. Then, we employ the cointegration theory to check whether the pre-selected pairs are co-integrated.

Pre-selection Using Correlation. The perception of pairs trading strategy is to see to what extent the two potential stocks are correlated. Correlation is a statistical term measuring the relationship between two stocks that have similar trends. Following [2,18,19], here we use the correlation statistic to pre-select the potential stocks for refining the trading pairs, because the correlation measure can identify the highly-correlated stocks.

Considering two stocks X and Y, let P^X and P^Y be the prices of the two stocks, respectively. We use Pearson Correlation Coefficient ρ to measure the correlation of stocks X and Y as follows.

$$\rho_{XY} = \frac{\sum_{t=1}^{T}(P_t^X - \overline{P}^X)(P_t^Y - \overline{P}^Y)}{\sqrt{\sum_{t=1}^{T}(P_t^X - \overline{P}^X)^2 \sum_{t=1}^{T}(P_t^Y - \overline{P}^Y)^2}}, \tag{1}$$

where T is the observation over the investigated period, and $\overline{P}^X$ and $\overline{P}^Y$ denote the mean prices of stocks X and Y, respectively, which are given by

$$\overline{P}^X = \frac{1}{T}\sum_{t=1}^{T} P_t^X, \text{ and } \overline{P}^Y = \frac{1}{T}\sum_{t=1}^{T} P_t^Y. \tag{2}$$

ρ_{XY} describes the strength of a relationship between two stocks X and Y, ranging from -1 to 1. Specifically, the larger positive value ρ_{XY} is, the larger positive correlation the correlation between stocks X and Y, indicating that the stock pairs are highly matched. Following Miao [2], we select the stock pairs by setting $\rho_{XY} = 0.90$ as a threshold.

With the correlation coefficients as the filter, we can get the correlation matrix $\mathbf{Cr}_t$ of the selected stocks at time t.

Selection Using Cointegration. The definition of cointegration is given by Engle and Granger [9,10], which states that a specific linear combination of two non-stationary time series could be stationary. In other words, the cointegration relationship attempts to capture the statistical feature of the paired stocks whose prices can move together in a lockstep way. The cointegration theory provides a powerful tool for investigating common trends of multivariate time series and a sound methodology for modeling both of the long-term equilibrium and short-term deviations in a system.

The basic idea of cointegration is to identify the pairwise assets whose prices could be linearly combined to produce a stationary time series. If the prices of stocks X and Y, namely $\{P_t^X, P_t^Y\}$, are assumed individually non-stationary, there should be a parameter β to guarantee the stationary of the following equation.

$$P_t^X - \beta P_t^Y = c + \varepsilon_t, \tag{3}$$

where c is a mean of the cointegration model, and ε_t is stationary and mean-reverting process. If a process is stationary, it will be denoted as I(0). Furthermore, if a non-stationary time series becomes stationary after first differencing, it will be denoted as I(1). Thus, two integrated series P_t^X and P_t^Y are cointegrated if P_t^X and $P_t^Y \sim I(1)$, but there exists a parameter β such that $z_t = P_t^X - \beta P_t^Y \sim I(0)$.

Here we use the Augmented Dickey Fuller (ADF) test [20], which is regarded as the most stationary test, to verify the regression cointegration. The basic idea of ADF centers on the regression residual ε_t to determine whether it has a unit root.

In the ADF test, testing for the presence of the unit root in ε_t is defined as follows,

$$\Delta Z_t = \alpha + \gamma t + \beta Z_{t-1} + \sum_{i=2}^{p} \delta_i \Delta Z_{t-i} + \mu_t, \tag{4}$$

where α is a constant, γ is the coefficient on a time trend, p is the lag order of the autoregressive process, and μ_t is an error term and serially uncorrelated. If parameters α and γ are both equal to zero, Eq. (4) can model a random walk. If α does not equal to zero but γ does, Eq. (4) can model a random walk with a drift.

Then the unit root test for the residual ε_t is carried out with the null hypothesis $H_0 : \beta = 0$. A statistical form of the ADF test is given by

$$ADF\ test = \frac{\hat{\beta}}{SE(\hat{\beta})}, \tag{5}$$

where $\hat{\beta}$ is the coefficient estimated by OLS. The standard errors $SE(\hat{\beta})$ can be obtained by

$$SE(\hat{\beta}) = \sqrt{\frac{\sum_{t=1}^{T}(\Delta Z_t - \widehat{\Delta Z_t})^2}{(T-2)\sum_{t=1}^{T}(Z_t^2 - T\overline{Z}_t^2)}}. \tag{6}$$

If the ADF test statistic is less than the critical value, the null hypothesis $H_0 : \beta = 0$ will be rejected, indicating that there is no unit root and the residual ε_t is stationary. Thus, the two stock prices $\{P_t^X, P_t^Y\}$ are cointegrated.

With the methodology introduced above, we can filter these stock pairs that do not have cointegration relationships based on the correlation matrix $\mathbf{Cr}_t$. Finally, we obtain the weighted cointegration matrix $\mathbf{Ci}_t$ of the adopted stocks at time t by setting those entries without cointegration relationships as 0. Next, we will construct the FGs and MSTs based on $\mathbf{Ci}_t$ following a prevalence methodology [21].

2.3. Network construction

To capture the information of the weighted cointegration matrix $\mathbf{Ci}_t$, we build the FGs and MSTs based on the weighted cointegration matrix following [22–24]. Since a tiny variation of constructing methodologies (e.g., the size of the threshold) may cause great differences on the characters of constructed network, we provide the construction process of the two cointegration networks.

Full Graphs (FGs). The full graphs of the cointegration matrix can be viewed as the topology of the cointegration relationships of the adopted assets. We get the full graphs of the weighted cointegration matrix ($\mathbf{Ci}_t$) by using a strict restriction on the trading pair relationships. As mentioned previously, the edges between pairwise assets should pass the ADF tests of cointegration. We further get a matrix with non-symmetric entries and force it to a symmetric matrix by saving the symmetric elements only. By this way, we finally form the full graph matrix $\mathbf{G}_t$. Therefore, the full graph matrix $\mathbf{G}_t$ can be obtained as follows,

$$\mathbf{G}_t = \frac{1}{2}(\mathbf{Ci}_t + \mathbf{Ci}_t{}') - \mathrm{abs}(\frac{1}{2}(\mathbf{Ci}_t - \mathbf{Ci}_t{}')), \tag{7}$$

where abs(·) denotes the absolute value of all the entries in matrix, and $\mathbf{Ci}_t{}'$ refers to the transpose of cointegration matrix $\mathbf{Ci}_t$.

Minimum Spanning Trees (MSTs). Spanning trees are particular types of graphs that connect all the vertices in a graph without any cycle. Considering that the MSTs can retain lots of salient features of the full graphs, we follow Onnela et al. [22] and Wang et al. [25], and construct the MSTs based on the FGs.

Full graph matrix $\mathbf{G}_t$ is applied to determine the MST network connecting N nodes of $\mathbf{G}_t$ with $N-1$ edges such that the total edge weight is as small as possible [26]. To some extent, the MSTs are usually constructed by $N-1$ important correlations from the connected edges of the FGs [27]. In the spirit of Keskin et al. [28], we further employ Kruskal's [29] algorithm to construct the MSTs ($\mathbf{M}_t$). Specifically, the construction procedure of MST is presented as follows.

Step 1 Construct a graph $\mathbf{M}_t$ using all the nodes in $\mathbf{G}_t$, where each node in $\mathbf{M}_t$ is a separate tree.
Step 2 Create set S_t containing all the edges in graph $\mathbf{G}_t$.
Step 3 Remove an edge with the minimum weight from S_t. If the removed edge connects two different trees in $\mathbf{M}_t$, add it to the forest $\mathbf{M}_t$.
Step 4 Repeat step 3 until S_t is empty or $\mathbf{M}_t$ is not yet spanning. Finally, we get the minimum spanning tree $\mathbf{M}_t$ at time t.

For a full technical introduction on the MST approach, see Refs. [26,30–32].

2.4. Network analysis

We intend to analyze the cointegration networks with respect to the static and dynamic properties, focusing on the formation and evolution of the edges. Specifically, we firstly investigate the formation of edges from the perspective of network density and community. Subsequently, we investigate the evolution of edges by calculating their single-step surviving rate (SSR) and multi-step surviving rate (MSR). Finally, we study the re-occurrence of edges for discovering the economic meaning of the cointegration networks.

2.4.1. Analysis on edge formation

Density. In order to investigate the formation of trading pairs, we investigate the networks systemically from the perspective of network density. The network density measures how strong the vertices of a graph are connected. It is defined as the fraction of actual connections over the total possible connections. For an undirected cointegration network, e.g., full graph $\mathbf{G}_t(E_t, V_t)$, the density d at time t is defined as:

$$d_t = \frac{2|E_t|}{2|V_t|(|V_t|-1)} = \frac{|E_t|}{|V_t|(|V_t|-1)}, \tag{8}$$

where $|E_t|$ denotes the number of edges and $|V_t|$ denotes the number of vertices of the networks. Since the MTSS programme remarkably changes the trading rules of the Chinese stock market, we use the stocks both on and off the MTSS list for comparison. Particularly, we compare density d of the networks by using the stocks *on/off* the MTSS programme and those *before/after* the announcement of the programme to investigate whether the programme enhances the cointegration relationships of stocks. Obviously, the larger the density is, the closer the relationships are.

Community. The community structure is widely found in the real world complex networks (e.g., Refs. [33,34]), as well as the financial networks (e.g., Refs. [35,36]). Thus, the vertices in the network can be usually clustered into different groups that are compactly linked internally and sparsely connected externally. In this paper, we investigate the community structure of the full graphs (FGs) and the minimum spanning trees (MSTs) from the following two perspectives. First, as a lot of portfolios are constructed within industries [1,4], it is necessary to make an analysis on the community structure through the perspective of industries. Second, since the state-owned assets have been concerned by more and more researchers, we investigate the communities through the perspective of asset ownerships, e.g., state-owned or non-state-owned.

We use the Louvain method of Blondel et al. [37] to investigate the community structure of the networks. To capture the edge connectivity among different communities, we use the cut ratio to measure the connection degree among different communities,

$$cut-ratio = \frac{\sum_{X\in C_1, Y\in C_2} w_{XY}}{\sum_{X\in C_1, Z\in C_1} w_{XZ}}, \tag{9}$$

where C_1 and C_2 denote two different communities, and w_{XY} and w_{XZ} represent the weight of edges E_{XY} and E_{XZ}, respectively.

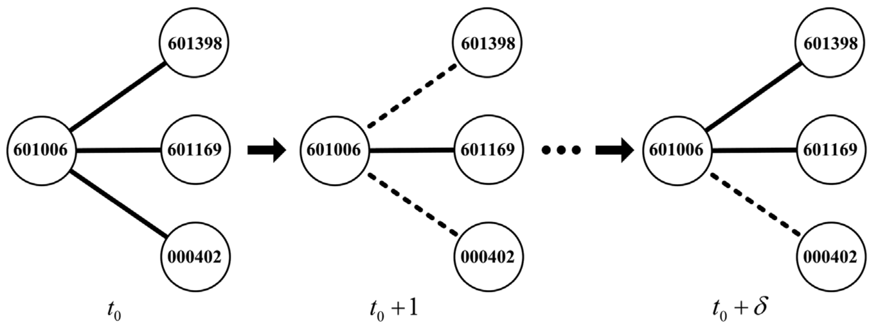


Fig. 2. Toy example of network evolution. The ego-network is center with stock 601006. The time horizon of from t_0 to $t_0+\delta$ when the relationship between 601006 and 601398 re-appears.

2.4.2. Analysis on edge evolution

To reduce the risk of pairs trading, investors are interested in the trading pairs with steady cointegration relationships. Thus, one natural question is how robust or the consistent the cointegration relationships are. In this paper, we investigate this dynamic feature of the cointegration networks to answer this question. We first analyze the rate of edge surviving by using the single-step surviving rate (SSR) and multiple-step surviving rate (MSR), which is associated with the robustness of the edges during the evolution of networks. Then, we propose the use of re-occurrence circles to investigate the evolution of edges, which is associated with the consistency of the trading pairs.

Edge Surviving Rate. In this subsection, we introduce two concepts developed by Onnela et al. [22], i.e., the single-step survival rate (SSR) and the multi-step survival ratio (MSR) to investigate the robustness of links over time and the long-term evolution of the dynamic networks, respectively.

The SSR is defined as the ratio between the number of common edges found in two consecutive networks and the total number of edges, i.e.,

$$SSR(t) = \frac{1}{M}|E_t \cap E_{t+1}|, \tag{10}$$

where M is the total number of edges of the network. For the MSTs, $M_{\mathrm{MST}} = N - 1$ and for the FGs, $M_{\mathrm{FG}} = \max|E_t|(t = 1, \ldots, k)$,[2] E_t and E_{t+1} are the set of edges of the network at times t and $t+1$, respectively. $\cap$ is the intersection operator, k is the number of network evolution series, and $|\cdot|$ denotes the number of observations in the investigated set. Similarly, MSR is defined as the fraction of edges found in multiple consecutive networks in common, i.e.,

$$MSR(t_0, \delta) = \frac{1}{M}|E_{t_0} \cap E_{t_0+1} \cap \cdots \cap E_{t_0+\delta-1} \cap E_{t_0+\delta}|, \tag{11}$$

where t_0 is the initial time and δ is the time step length.

Edge Re-occurrence Circles. Pairs trading is a well-known short-term speculation strategy. If the trading pairs are not steady, risk-adjusted returns from such strategy will not be positive. On the contrary, if the trading pairs are steady, their prices will converge to the long-run equilibrium following to the motivation of pairs trading. Therefore, we investigate the stability of trading pairs by the unique phenomenon of the "re-occurrence" of edges in the cointegration networks.

The "re-occurrence" of cointegration edges, i.e., the Appearance–Disappearance–Appearance ("A–D–A") circles, is associated with the stability of cointegration relationships. Fig. 2 provides a toy example of the edge evolution. If a trading pair is steady, the corresponding edges will be expected to keep "Appearance" all through a period, like the relationship between stocks 601006 and 601169 surviving from t_0 to $t_0+\delta$. However, in many cases, the cointegration relationships are not as strong as surviving for long time. Some of them will disappear for a certain period, like the relationship between stocks 601006 and 601398 disappeared for δ time units. Others will disappear for even more time, like the relationship between stocks 601006 and 000402. Since a huge majority of the cointegration relationships cannot survive for all the time, it is very interesting to investigate their re-occurrence phenomena. In terms of the cointegration networks, we focus on investigating the "A–D–A" of edges (e.g., the edge between stocks 601006 and 601398 shown in Fig. 2). Here, we use the following indices to study the stability of cointegration edges, (i) how long the disappeared time of an "A–D–A" circle is on average. The shorter the disappeared time is, the steadier the cointegration relationship is. (ii) what the probability of an "A–D–A" circle disappeared for δ time units is.

Specifically, we measure the disappeared time "A–D–A" circles of edge E by counting the number of time units when it disappeared. We use the time span between its successive appearances, e.g., in Fig. 2, the disappeared time of relationship

2 To fairly compare the evolution of edges, we use the largest number of all FGs for M.

Table 1
Data set of the MTSS programme.

Year	# MTSS assets	Start date	End date	#Exchange days
2010	90	31 March	31 December	185
2011	285	4 January	31 December	244
2012	287	4 January	31 December	243
2013	713	4 January	31 December	238
2014	913	4 January	31 December	245
2015	913	4 January	31 December	244
2016	971	4 January	30 December	242
Used	80	31 March 2010	30 December 2016	1643

Table 2
Descriptive statistics of the correlation coefficients $\{\rho_{ij}; i < j\}$.

	Mean	Min.	Max.	Std.	Skewness	Kurtosis
$\{\rho_{ij}; i < j\}$	0.4553	0.1559	0.8999	0.1226	0.5997	3.5207

between stocks 601006 and 601398 is δ time units. We use P_δ^r to represent the probability of an "A–D–A" circle of E that disappeared for δ time units. For simplicity, P_δ^r is obtained by using the ratio between the number of "A–D-A" circles with δ time units and the number of all the "A–D–A" circles in the investigated period. It is calculated as follows,

$$P_\delta^r = \frac{|E_\delta^{\mathrm{ADA}}|}{\sum_{\delta=2,3,\ldots}^{n} |E_\delta^{\mathrm{ADA}}|}, \tag{12}$$

where $|E_\delta^{\mathrm{ADA}}|$ represents the number of "A–D–A" circles that disappeared for δ time units, and n represents the largest time units of that an edge disappeared during the investigated period.

3. Statistics on cointegration networks

We will provide the basic statistics of our data set and the cointegration networks to make an overview on the data of our experiments. These statistics including the distribution of correlation coefficients and the basic statistics of cointegration networks.

3.1. Data set

We obtain daily closing prices of the stocks that allowed MTSS in the Chinese stock market from Wind Database (http://www.wind.com). The time period spans over 31 March 2010 (the start time of the MTSS programme) to 30 December 2016. Table 1 illustrates the basic information of the stocks listed in the MTSS programme. In order to investigate the evolution of the networks, we use the assets listed in the MTSS programme all through the time. Finally, 1,643 daily closing prices of 80[3] assets are adopted in our data set.

Moreover, the adopted stocks contain 22 assets from financial industry, 43 assets from industrials; and 53 of all the stocks are from state-owned corporations. We obtain the industry and ownership information of the assets from Wind database.

3.2. Statistics of correlation coefficients

Here we examine the probability density function (PDF) of all the correlations among the adopted assets. The graphical results are shown in Fig. 3. In Table 2, we present the descriptive statistics of correlation coefficients. Both Fig. 3 and Table 2 show that: (i) the PDF of the correlation coefficients is asymmetric with a positive value at its center, namely the mean value is 0.455, suggesting that the correlations of the stocks are neither independent nor negatively correlated, but tightly correlated, and (ii) right skewness and leptokurtosis exist in the correlation coefficients $\{\rho_{ij}; i < j\}$, indicating that the PDF of $\{\rho_{ij}; i < j\}$ is asymmetric, fat-tailed, and non-Gaussian.

3.3. Statistics of cointegration networks

We provide an overview of the FG and MST with an analysis on its basic properties. Fig. 4 illustrates the FG and MST constructed by the methodology proposed in Sections 2.2 and 2.3 using the daily stock prices spanning over 31 March 2010 to 30 December 2016. Each node in the figure represents an asset and each edge between a node pair denotes the cointegration relationship of them. The FG in Fig. 4(a) looks quite dense, where the average degree is 10.62 and the average path length is 3.52. Based on the above analysis, we know that the FG is a tightly connected small world network. The MST is shown

3 10 of the 90 stocks in 2010 were excluded from the MTSS programme in the following years.

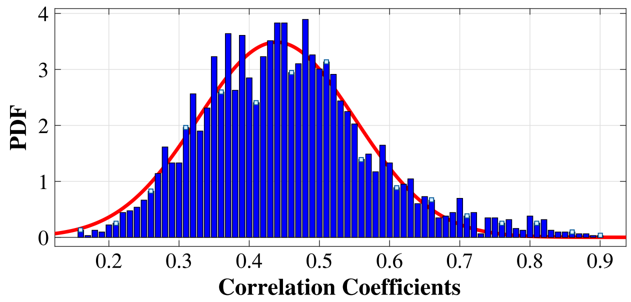
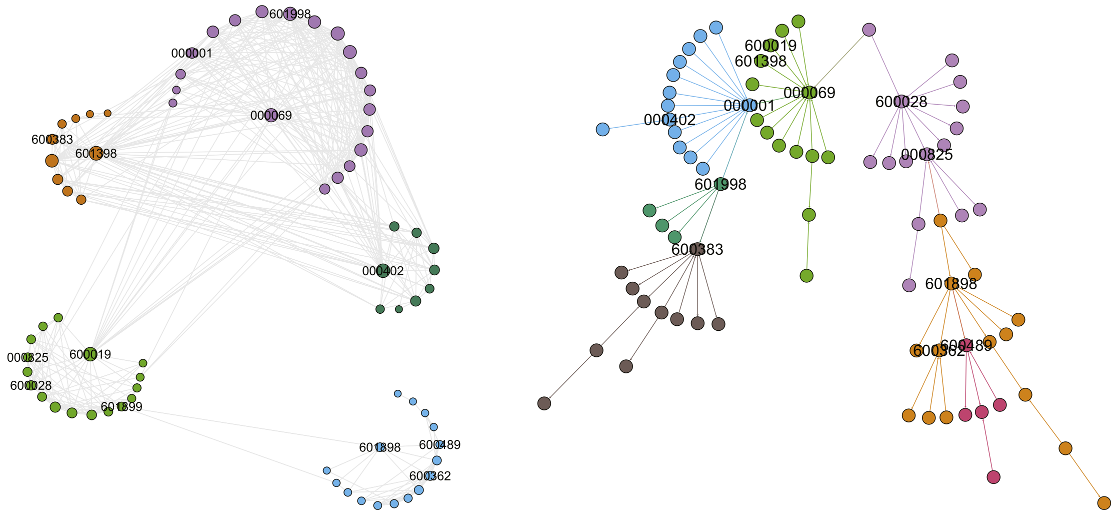


Fig. 3. Probability distribution function (PDF) of correlation coefficients.



(a) Full graph. (b) Minimum spanning tree.

Fig. 4. (a) Snapshot of the full graph of the cointegration network. The size of each node is related to its degree. (b) Snapshot of the minimum spanning tree of the cointegration network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in Fig. 4(b), and its average degree is 0.99 and the average path length is 5.34. The nodes are highly cointegrated and the average path length is within 6, even when we use the least edges (cointegration relations) to connect the nodes (assets), confirming again that the adopted assets are highly cointegrated. To sum up, by a general analysis of the two graphs, we know that the prices of the MTSS stocks are highly synchronized.

Some of the nodes, namely hub nodes, are cointegrated with many other nodes, and have significant impact on the network. In Fig. 4, we illustrate some of the hub nodes that act critically important both in the FG and the MST. For example, node 600028 (degree: 28 in the FG, 11 in the MST, China National Petroleum Corporation, CNPC) and node 601398 (degree: 27 in the FG and 2 in the MST, Industrial and Commercial Bank of China, ICBC). These node degrees are about one third of the node number of the FG and one sixth of the MST. Moreover, they also act as structural holes of the MST, suggesting that such assets play a leading role in the networks because many assets are cointegrated with them. In fact, ICBC is a giant company which is widely regarded as the indicative stocks of market.

To sum up, by an overview on the cointegrated networks, we know that they are highly correlated, small world network and with a considerable number of high degree nodes, suggesting that the adopted assets of the MTSS programme are highly synchronic with a few of them that play leading roles in the network.

4. Analysis on the cointegration networks

In this section, we first analyze the formation of edges by investigating the density and community of the cointegration networks. Then, we will discuss the economical meaning of the cointegration edges' evolution.

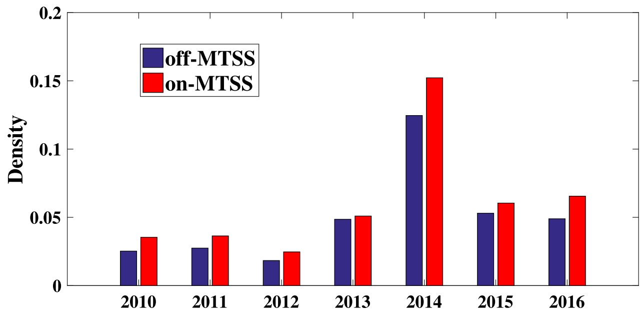
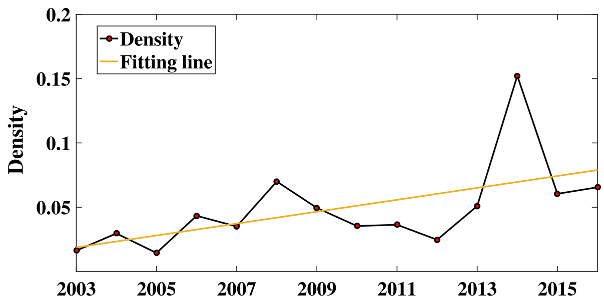
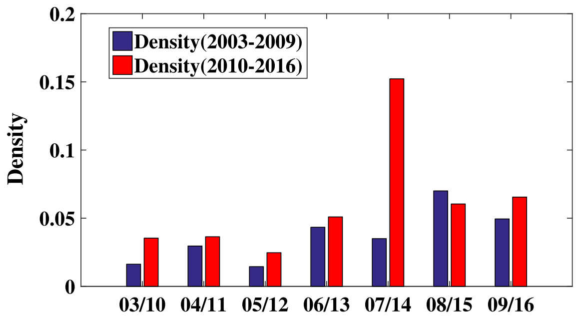


Fig. 5. Density of the on/off MTSS networks.



(a) Density of the networks.



(b) Density of the networks for comparison.

Fig. 6. (Color online) (a) Density series of the adopted stocks. (b) Density series of the adopted stocks for comparison.

4.1. Density

The network density can be viewed as a general property of the cointegration networks. The denser the networks are, the more synchronized the stocks' prices are. Therefore, we conduct the density analysis on the FGs to examine whether the MTSS programme boasts the synchronization of stock prices.

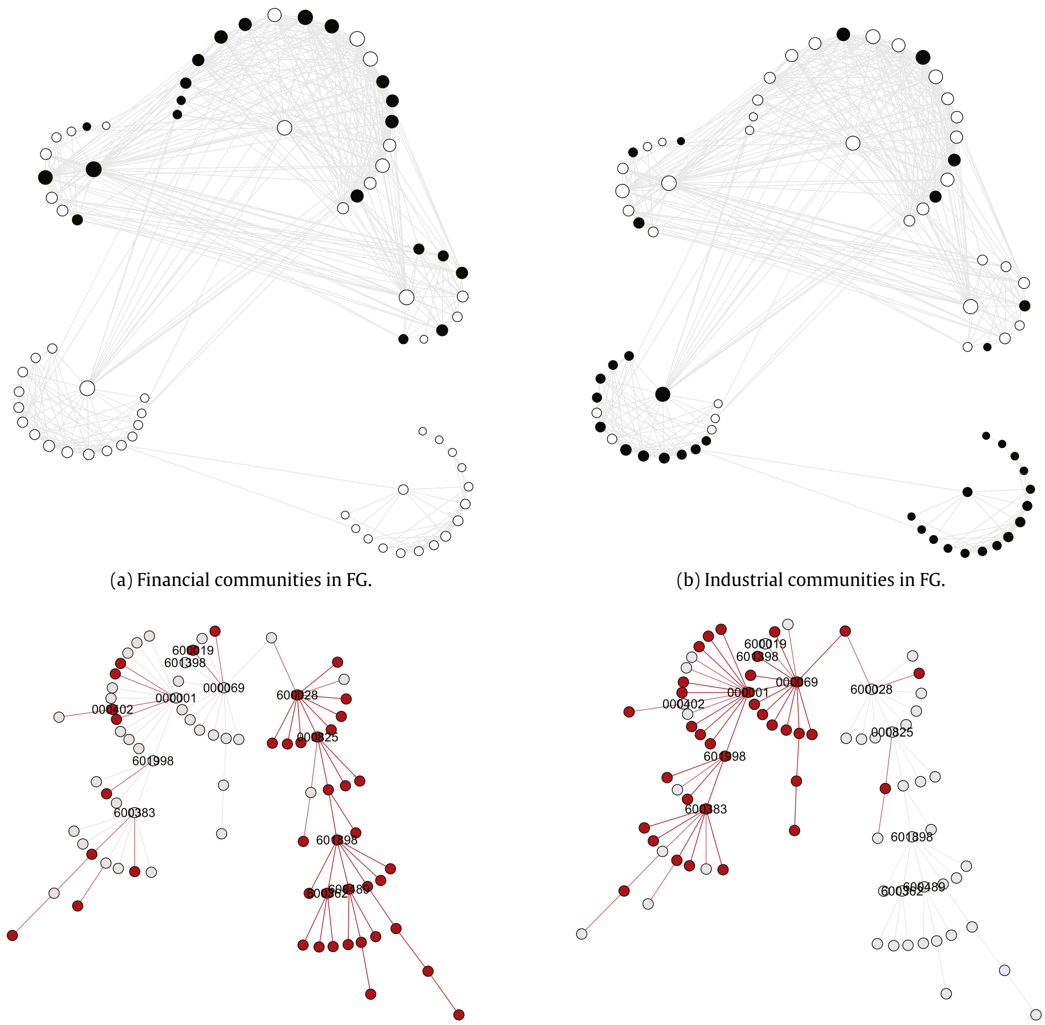
We use the assets that are on/off the MTSS programme to see whether the programme boasts the synchronization of stock prices. Specifically, we calculate the density of both the assets listed and the equal number of assets that unlisted in the programme following Eq. (8). The unlisted stocks are randomly chosen from all the assets that unlisted in the MTSS programme. Fig. 5 shows the density of the networks constructed by the assets on and off the MTSS programme from 2010 to 2016. The results show that the networks constructed by the assets on the MTSS programme are more denser than those off the programme. Furthermore, we also investigate the evolution of density of the listed stocks.[4] Fig. 6 shows the density evolution of full graphs. From Fig. 6(a) we find that the density of the cointegration networks is growing up with fluctuation. To explicitly show the density evolution of FGs, we compare the pairwise densities by separating the time by 2010, the launching of MTSS. The results are shown in Fig. 6(b). From the figure one can see that in each paired years, the network density after 2010 is larger than that before. Such a result indicates that the MTSS programme does boast the forming of the cointegration relationships among the stocks listed in the programme. Such a result also provides evidence that the MTSS programme has shaped the market structure and promoted the level of co-integration of stocks' prices.

To sum up, the assets listed in the MTSS programme show denser connections than those not included in the programme and before the programme. We conclude that the MTSS programme boasts the synchronization of stock prices.

4.2. Community

The community structures of the cointegration networks are shown in Fig. 4. The FG (see Fig. 4(a)) can be roughly separated into 5 clusters with different sizes, where they have 20 (purple colored), 17 (green colored), 16 (blue colored), 10 (orange colored), and 9 (dark green colored), respectively. In addition, one can also see that the upper three clusters are more tightly connected comparing with the bottom two clusters. The MST (see Fig. 4(b)) has at least 5 grades and can be roughly divided into two cascades, the left part and the right part.

4 Some of the MTSS assets were unlisted before 2010, here we only use 42 stocks that have prices all through the end.



(a) Financial communities in FG.

(b) Industrial communities in FG.

(c) Financial communities in MST.

(d) Industrial communities in MST.

Fig. 7. (Color online) (a) Financial communities in the FG (●: financial sector). (b) Industrial communities in the FG (●: industrial sector). (c) Financial communities in the MST (●: financial sector). (d) Industrial communities in the MST (●: industrial sector).

Since pairs trading is often conducted within industries [1,4], it is interesting to investigate the networks with respect to industries. In this paper, we investigate the two largest groups of the adopted assets, the financial group and the industrial group of the FG and the MST. The results are shown in Fig. 7. From Fig. 7(a) and (b), we find that most of the financial assets are in the upper three clusters and most of the industrial assets are in the bottom two clusters of the full graph. Similarly, from Fig. 7(c) and (d), one can see that the financial corporations are roughly distributed in the left cascade and the industrial corporations in the right cascade of the MST. The results indicate that both of the FG and the MST keep the industrial information of the adopted assets. Since the cointegration networks are a macro-analysis on the correlation of the adopted assets, the two graphs imply that the intrinsic correlation information of the price information of these assets. The clearly separation of these two communities in both of the two networks indicates that the cointegration relationship of the potential assets are highly related to their industries.

We measure connectivity of the financial and industrial groups in both of the FG and MST by using the cut ratio. According to Eq. (9), the cut ratios of both the financial and industrial communities are 0.1034 and 0.3667 in the minimum spanning tree, respectively. In addition, the cut ratios of the two communities are less than 1 in the full graph. The results of cut ratio indicate that both of the financial and industrial communities are clearly separated from the cointegration networks. Therefore, the sectoral information can be used to form portfolios.

Most of the nodes are included in their own industrial communities. But there are some of them separated from their own industrial community members, e.g., node 600016 in the financial communities of FG, node 600739 and 601111 in the industrial communities of MST. As the MST omits most cointegration relationships, it is not strange to find outliers of the

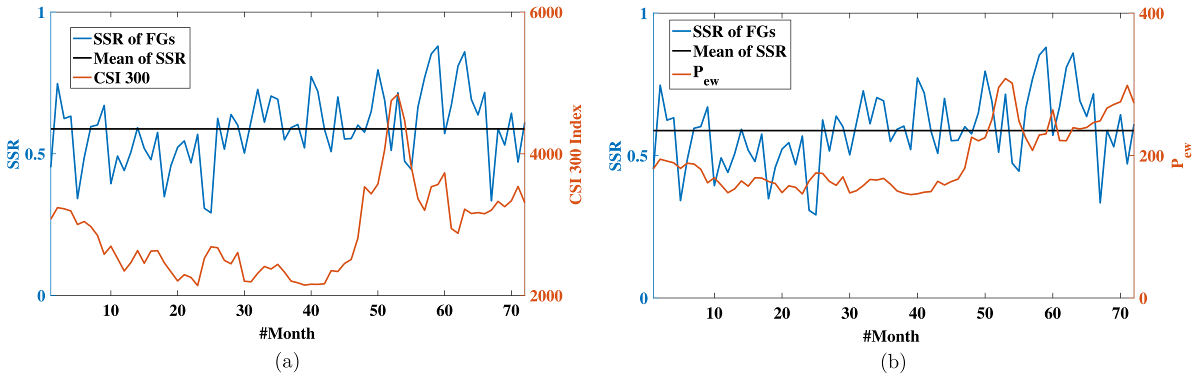


Fig. 8. (Color online) (a) Monthly evolution of SSR of the full graphs (FGs) *vs.* the CSI 300 index. (b) Monthly evolution of SSR of the full graphs (FGs) *vs.* the equal-weighted stock price of the adopted assets.

communities. Thus here we mainly discuss the outliers in the FG. Taking node 600016 (China Minsheng Bank, CMBC) as an example, it is regarded as the finance corporation in the grouping data. However, its price is highly correlated with only two nodes (600048, Poly Real Estate and Yunnan Baiyao Group 000538). A natural question is that why the CMBC is not included in the financial communities of the full graph. A reasonable explanation is, unlike other state-owned financial institutions, that Minsheng Bank is widely considered as the first privately operated bank of China. Its prices are often associated with the private corporations in the markets. Thus, it is not strange that the movement of its stock prices shows different features.

Summarizely, both of the two networks constructed by the price series show clear community structure that are highly related to their industries, suggesting that the cointegration networks preserve the stock industry information, which is in line with the findings by Gatev et al. [1], i.e., the outliers of the communities can be explained by the background of the assets.

4.3. Investigation of edge evolution

To explore the dynamics of cointegrated networks, we try to investigate the robustness and consistency of the networks through several measures, i.e., SSR, MSR, and the re-occurrence of edges.

Specifically, we build the monthly price-based cointegration networks to capture the stock market evolution. We set the width of the rolling window as one year,[5] and the step size as one month. As the data set ranges from 31 March 2010 to 30 December 2016, we finally obtain 73 time-varying networks.

4.3.1. Single-step surviving rate

Full Graph. Fig. 8 presents the curve of monthly evolved SSRs of the FG. The average SSR is 0.588, indicating that about 40% of the edges change over the period (one month). Such a figure shows us a new perspective of pairs trading that the cointegration relationship of stock prices evolves quite quickly.

Considering the SSR of FGs is based on the movement of stock prices, we propose a natural question that whether the SSR is related to the movement of the stock market trend, for example, the CSI 300 index[6] and the mean of the adopted stock prices. In this paper, we investigate the relationships between the SSR of full graphs and the later two indices. Fig. 8 illustrates the CSI 300 index and the equal-weighted price (P_{ew}) of the stock collection. We find that the vast majority of the CSI 300 and P_{ew} are largely overlapped, indicating that they are highly related to each other. In contrast, it seems that the trends of SSR curve are not related to the monthly evolved value of the two indices. To further explore the relationships among SSR, CSI 300 and P_{ew}, we compute the correlation coefficients among the three series. Table 3 reports the results of their correlation coefficients and p-values. As the coefficients are extraordinarily low and the p-values over 0.05, we can infer that the SSR is not related to both P_{ew} and CSI 300. Such an inference indicates that the pairs trading fluctuation curve is independent to the stock market movement. In other words, the risk of pairs trading strategy is independent to the market trend or pairs trading is a market neutral strategy which is in line with the motivation of pairs trading.

Minimum Spanning Tree. The time-varying SSRs of the MST networks are presented in Fig. 9. There are about 41.7% SSR points of the MST networks are under 0.3, denoting that more than half of the consecutive MSTs at times t and $t+1$ are not identical, as only a few of them are survived in each step. Moreover, the average SSR value of the MST is 0.322, suggesting

[5] We use a nature year as a time horizon, e.g., from the first exchange day of January to the last exchange day of December, or from the first exchange day of March to the last exchange day of February of the next year.

[6] We adopt the China Securities Index 300 (CSI 300), a widely accepted index of the Chinese stock market, as the stock market index, which incorporates 300 major stocks listed in the Shanghai and Shenzhen Stock Exchanges. The listed stocks were revised in June, 2016..

Table 3
Coefficients and p-values among SSR (full graphs), CSI 300 and P_{ew}.

Index A	Index B	Coefficient	p-value
SSR	P_{ew}	0.1775	0.1357
SSR	CSI 300	0.1557	0.1691
P_{ew}	CSI 300	0.8841	0.0000

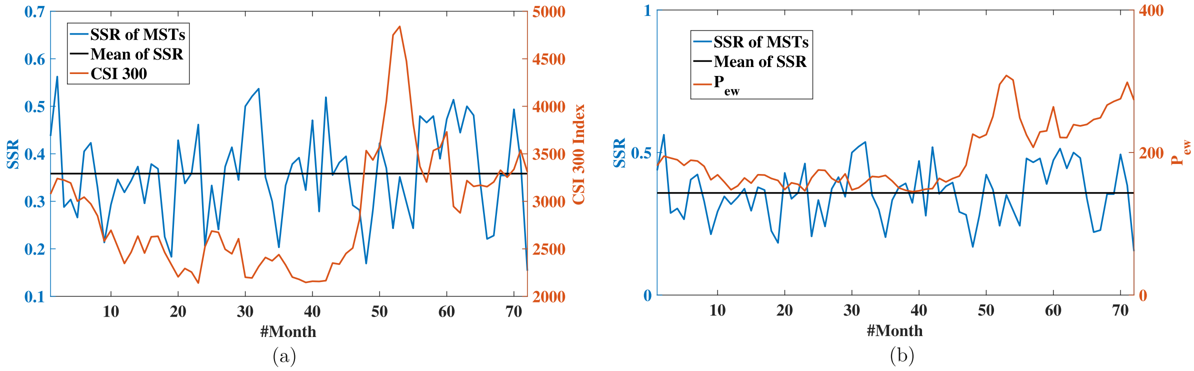


Fig. 9. (Color online) (a) Monthly evolutional SSR of the MSTs *vs.* the CSI 300 index. (b) Monthly evolutional SSR of the MSTs *vs.* the equal-weighted stock price of the adopted assets.

Table 4
Coefficients and p-values among SSR(MST), CSI 300 and P_{ew}.

Index A	Index B	Coefficient	p-value
SSR	P_{ew}	−0.0510	0.6705
SSR	CSI 300	−0.0806	0.5012
P_{ew}	CSI 300	0.8841	0.0000

that a huge majority (near 70%) of the links between the monthly evolved networks could not survive from one time to the next. As the MST omits most of the connections in every transition step, it is not strange that the fluctuation of MST is much larger than that of the FGs.

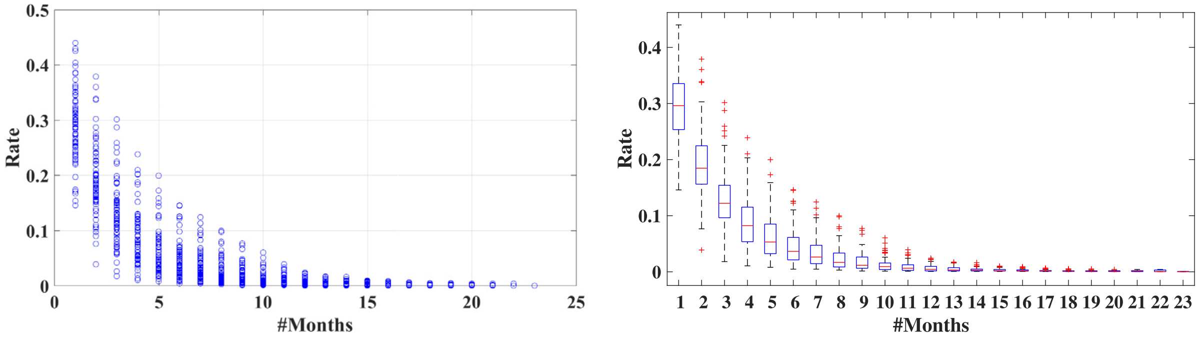
Similarly, to observe the correlations between the SSR of MSTs and stock market movement, we further compare the SSR of MSTs with the CSI 300 index and the equal-weighted stock price of the adopted stocks. Fig. 9 illustrates the CSI 300 index and the equal-weighted price (P_{ew}) of the stock collection. From the figure one can find that similar to that in the FGs, the trends of SSR curve in the MSTs are not related to both the CSI 300 and the equal-weighted price P_{ew}.

We compute the correlation of the SSR curve of MSTs with both of the two index series. Table 4 illustrates the results of their correlation coefficients and p-values. From the table one can see that the coefficients of SSR are negative, and both of the absolute values are quite low and the p-values over 0.05, indicating that the correlation is insignificant. Therefore, we can infer that the SSR of MST networks is not related to either CSI 300 or P_{ew}, confirming again that the selected pairs trading movement is independent to the stock market trend.

4.3.2. Multiple-step surviving rate

In this subsection, we investigate the multiple-step surviving rate (MSR) of the graph edges, which is associated with the consistency and stability of pairs trading. The MSR is calculated by Eq. (11) and drawn in the following figures. Fig. 10(a) shows the distribution of MSRs by month, where each point in the figure represents a MSR data element calculated from the evolutional FGs. We can find that for most of the networks, the majority of edges can only survive for one month, and the surviving time of a graph in two months is about 0.37. With the growing of month, the survive rate descends drastically till the 7th month, and then descends slowly. From Fig. 10(a) one can also see that for each edge number, most of the points are centered around their medium, indicating that most networks evolve similarly.

Fig. 10(b) shows the evolution of MSR by the box graph. From the figure one can observe that the mean value of the MSR of edges descends rapidly first, and then descends slowly. Moreover, the variation sizes (see the box) of edges surviving rate are also descending drastically with time, most of the edges in the networks will not survive in many circumstances. By further investigating the figure, one can see that only a few edges in most of the graphs can survive over 12 months. In other words, most graphs will be completely reconstructed in one year. In each time step, some outliers (red points out of the box range) represent the nodes that do not follow the major surviving rate distribution.



(a) MSR of the FGs.

(b) Box-graph of FGs' MSR.

Fig. 10. (Color online) Multiple-step surviving rate (MSR) of the full graphs. Rate distribution (a) and Box graph (b).

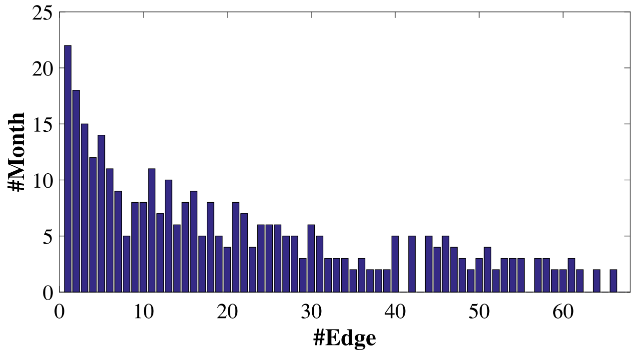


Fig. 11. Lasting time of the connected components.

4.4. Further investigation of edge evolution

In this subsection, we investigate the lasting time of edges and subgraphs. Moreover, we investigate the stability of the cointegration networks from the perspective of "re-occurrence" of edges.

4.4.1. Lasting time of edges

The duration of an edge can be treated as the lasting time of the cointegration relationship between the paired assets. Therefore, the MSR of cointegration networks can be viewed from another perspective. Specifically, we investigate the lasting time of edges and subgraphs to analyze the stability and consistency of the pairs trading strategy.

Note that the lasting time of subgraphs indicates the static relationships of some sub-parts of the graph. Fig. 11 shows the lasting time of cointegrated subgraphs where the x-axis is the number of component edges and the y-axis is the number of months that the components' lasting. From the figure one can see that, first, the longest time of the edges is 22 months, about two years, indicating that the cointegrated relationship is very steady. The steady cointegration relationship suggests less risk than those unsteady edges. Practically, if we form the portfolios using these paired assets, the portfolios are expected to have steady cointegration relationships for long time, and the risk of pairs trading will be reduced.

The lasting time of the connected subgraphs over 30 edges is within five months. Although the size over 30 grows larger than previous components, the lasting time of these components can also be compared to their previous components. Taking a deep investigation of these components, we know that most of the large components that can last for long time survived the period of 2014–2016. There was a stock disaster in the Chinese stock market in that period, and many stock prices rose up drastically before the disaster and drove down drastically after it. Therefore, many of the stocks are cointegrated as they are moving together steadily.

4.4.2. Re-occurrence of edges

As mentioned in Section 2.4.2, we investigate the probability of edges that are involved in the "A–D–A" circles and the time between the two "A". The results are shown in Fig. 12 and Table 5. Fig. 12 shows the disappeared time distribution of the edges, where the x-axis represents the disappeared time that an edge lasting and y-axis represents the logarithm of the probability of the disappeared time. To vividly describe the disappeared time distribution, we draw the fitting line

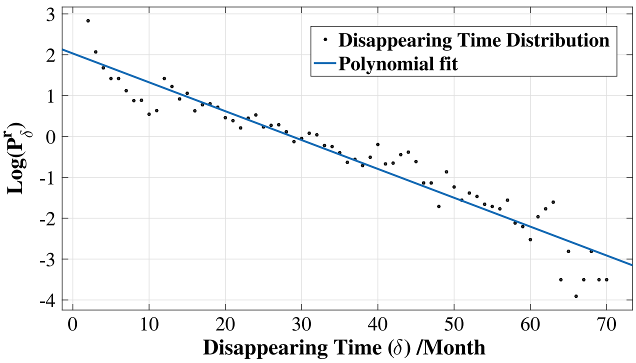


Fig. 12. Disappeared time distributions.

Table 5
Statistics on network evolution.

Month	Count	Percent	Cumulated percent
2	1693	16.97%	16.97%
3	790	7.92%	24.89%
4	535	5.36%	30.25%
5	412	4.13%	34.38%
6	412	4.13%	38.51%
7	306	3.07%	41.58%
8	240	2.41%	43.99%
9	242	2.43%	46.42%
10	172	1.72%	48.14%
11	188	1.88%	50.02%
12	412	4.13%	54.15%

in the figure which follows, $y = -0.071 \log P_\delta^r + 2.031$. The disappeared time distribution suggests that the lasting time of a majority of edges between two appearances are about two or three months. With the time going on, the probability of disappeared for δ months (P_δ^r) gets smaller and smaller, but the descending rate is getting slower. Table 5 shows the distribution of the disappeared time, which indicates that most of "A–D–A" circles last less than 12 month. Furthermore, by analyzing the statistics we know that only 54.15% pairwise assets are included in the "A–D–A" circles within one year. The results indicate that it is uneasy to find trading opportunities for the other 45% of the candidate trading pairs due to their unsteady price relationships, which may be a high risk for many investors.

By analyzing the evolution of edges with a focus on the stability of edges, we know that there are some steady edges that can last for many months. However, over 45% trading pairs are excluded in the re-occurrence circles of edges within 12 months. This means that a considerable number of trading pairs are unsteady, thus the investors should carefully pick pairs for trading in such short-term speculations.

4.5. Discussion

Regarding the potential application on pairs trading and portfolio optimization [38], it is very interesting to discuss the experimental results of this paper. Like much literature in this field to employ networks to detect trading risk and opportunity [39–41], we use the cointegration networks for macroscopical analysis over the market with respect of the pairs trading strategy. Note that we pay no attention to the construction of portfolios. However, our main findings are inspirational for the application of pairs trading strategy. For example, our finding shows that the co-integration relationships are not consistent but change quickly. The pairs trading strategy is widely regarded as market-independent, but we find it can be influenced by the state of market from the view of network. The co-integration relationships formed in a good state may not be consistent when the market changes to bad. As the pairs trading strategy is a well-known short-term speculation strategy, the unsteady of co-integration relationships will heavily influence the performance of the strategy. Our future work will focus on the influence of the role of market states on the specific profitability of the pairs trading strategy.

5. Conclusions

The Chinese stock market lifts the ban on margin-trading and short-selling (MTSS) providing the opportunity for pairs trading strategy. In this study, we have systemically analyzed the static and dynamic features of the potential asset pairs for the pairs trading strategy from a network perspective. Our constructed cointegration networks (FGs and MSTs) based on the cointegration theory and complex networks indicate:

(i) The cointegration networks including the FGs and MSTs are small-world, and highly correlated networks.
(ii) The stocks in the MTSS programme or the stocks within the same industry are more inclined to form cointegration relationships.
(iii) From the perspective of evolution by month, only a few number of linkages between assets survived from one time to the next.
(iv) The movement of the cointegration networks confirms that pairs trading is a market neutral strategy.
(v) The movement of edges reflects that there are a considerable number of unsteady trading pairs that may enhance the risks of pair trading.

In summary, our study provides a new view for the pairs trading strategy which is different from previous literatures in this area. The results indicate that it is necessary to analyze the features of the cointegration networks for the potential asset pairs systematically. The new findings may help investors to manage their portfolios and obtain robust rewards.

Acknowledgments

The work of DW and CM was supported by NSFC, China (Grant nos. 71431008, 71521061 and 71790593). The work of GJW was supported by NSFC, China (Grant no. 71501066) and the Hunan Provincial Natural Science Foundation of China (Grant no. 2017JJ3024).

References

[1] W.N.G. Evan Gatev, K.G. Rouwenhorst, Pairs trading: performance of a relative-value arbitrage rule, Rev. Financ. Stud. 19 (2006) 797–827.
[2] G.J. Miao, High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach, Int. J. Econ. Finance 6 (2014) 96–110.
[3] R.J. Elliott, J.V.D. Hoek, W.P. Malcolm, Pairs trading, Quant. Finance 5 (2005) 271–276.
[4] G. Vidyamurthy, Pairs Trading: Quantitative Methods and Analysis, John Wiley & Sons, 2004.
[5] Y.X. Lin, M. Mccrae, C. Gulati, Loss protection in pairs trading through minimum profit bounds: A cointegration approach, J. Appl. Math. Decis. Sci. 2006 (2006) 1–14.
[6] C. Alexander, A. Dimitriu, The cointegration alpha: Enhanced index tracking and long-short equity market neutral strategies, Icma Centre Discussion Papers in Finance, 2002.
[7] C.W.J. Granger, Some properties of time series data and their use in econometric model specification, J. Econometrics 16 (1981) 121–130.
[8] C.W.J. Granger, A.A. Weiss, Time series analysis of error-correction models, in: Studies in Econometrics Time, 1983, pp. 255–278.
[9] R.F. Engle, B.S. Yoo, Forecasting and testing in co-integrated systems, J. Econometrics 35 (1987) 143–159.
[10] R.F. Engle, C.W.J. Granger, Co-integration and error correction: representation, estimation, and testing, Econometrica 55 (1987) 251–276.
[11] G. Hong, R. Susmel, Pairs trading in the Asian ADR market, Working paper, 2003.
[12] C. Yang, Y. Chen, L. Niu, Q. Li, Cointegration analysis and influence ranka network approach to global stock markets, Physica A 400 (2014) 168–185.
[13] C. Tu, Cointegration-based financial networks study in chinese stock market, Physica A 402 (2014) 245–254.
[14] G.J. Wang, Z.Q. Jiang, M. Lin, C. Xie, H.E. Stanley, Interconnectedness and systemic risk of China's financial institutions, Emerg. Mark. Rev. (2017). http://dx.doi.org/10.1016/j.ememar.2017.12.001.
[15] S.W.Y.Y.B.F. Xi Zhang, Yunjia Zhang, Y.S. Philip, Improving stock market prediction via heterogeneous information fusion, Knowl.-Based Syst. 143 (2018) 236–247.
[16] X.B. Lu, B.Z. Qin, Synchronization in complex networks, Phys. Rep. 469 (2011) 93–153.
[17] X. Liu, H. An, H. Li, Z. Chen, S. Feng, S. Wen, Features of spillover networks in international financial markets: Evidence from the G20 countries, Physica A 479 (2017) 265–278.
[18] N. Huck, Pairs trading and outranking: The multi-step-ahead forecasting case, European J. Oper. Res. 207 (3) (2010) 1702–1716.
[19] H. Rad, R. Low, Y. Kwong, R. Faff, The profitability of pairs trading strategies: distance, cointegration and copula methods, Quant. Finance 16 (2016) 1541–1558.
[20] D.A. Dickey, W.A. Fuller, Distribution of the estimates for autoregressive time series with a unit root, J. Amer. Statist. Assoc. 74 (1979) 427–431.
[21] G. Bonanno, G. Caldarelli, F. Lillo, S. Miccich, N. Vandewalle, R.N. Mantegna, Networks of equities in financial markets, Eur. Phys. J. B 38 (2004) 363–371.
[22] J.P. Onnela, A. Chakraborti, K. Kaski, J. Kertsz, A. Kanto, Dynamics of market correlations: taxonomy and portfolio analysis, Phys. Rev. E 68 (2003) 056110.
[23] G.J. Wang, C. Xie, Tail dependence structure of the foreign exchange market: A network view, Expert Syst. Appl. 46 (2015) 164–179.
[24] G.J. Wang, C. Xie, S. Chen, Multiscale correlation networks analysis of the US stock market: a wavelet analysis, J. Econ. Interact. Coord. 12 (2017) 561–594.
[25] G.J. Wang, C. Xie, H.E. Stanley, Correlation structure and evolution of world stock markets: Evidence from pearson and partial correlation-based networks, Comput. Econ. 51 (2018) 607–635.
[26] R.N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1999) 193–197.
[27] C.G. Gilmore, B.M. Lucey, M.W. Boscia, Comovements in government bond markets: A minimum spanning tree analysis, Physica A 389 (2010) 4875–4886.
[28] M. Keskin, B. Deviren, Y. Kocakaplan, Topology of the correlation networks among major currencies using hierarchical structure methods, Physica A 390 (2011) 719–730.
[29] J.B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, Proc. Amer. Math. Soc. 7 (1956) 48–50.
[30] T. Aste, W. Shaw, T.D. Matteo, Correlation structure and dynamics in volatile markets, New J. Phys. 12 (2010) 085009.
[31] G.J. Wang, C. Xie, Correlation structure and dynamics of international real estate securities markets: A network perspective, Physica A 424 (2015) 176–193.
[32] G.J. Wang, C. Xie, K. He, H.E. Stanley, Extreme risk spillover network: Application to financial institutions, Quant. Finance 17 (2017) 1417–1433.
[33] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99 (2002) 7821–7826.
[34] M.E. Newman, Modularity and community structure in networks, Proc. Natl. Acad. Sci. USA 103 (2006) 8577–8582.
[35] G. Buccheri, S. Marmi, R.N. Mantegna, Evolution of correlation structure of industrial indices of U.S. equity markets, Phys. Rev. E 88 (2013) 493–494.

[36] D.M. Song, M. Tumminello, W.X. Zhou, R.N. Mantegna, Evolution of worldwide stock markets, correlation structure, and correlation-based graphs, Phys. Rev. E 84 (2011) 026108.
[37] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (2008) 155–168.
[38] L. Zhao, W. Li, A. Fenu, B. Podobnik, Y. Wang, H.E. Stanley, The q-dependent detrended cross-correlation analysis of stock market, J. Stat. Mech. Theory Exp. 2018 (2018) 23402.
[39] F. Pozzi, T. Di Matteo, T. Aste, Spread of risk across financial markets: better to invest in the peripheries, Sci. Rep. 3 (1) (2013) 1665.
[40] G. Peralta, A. Zareei, A network approach to portfolio selection, J. Empir. Finance 38 (2016) 157–180.
[41] L. Zhao, G.J. Wang, M. Wang, W. Bao, W. Li, H.E. Stanley, Stock market as temporal network, 2017, Arxiv Prepr.