Physica A 502 (2018) 545–553

Contents lists available at ScienceDirect

Physica A

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

Emergence and temporal structure of Lead–Lag correlations in collective stock dynamics

Lisi Xia [a], Daming You [a], Xin Jiang [b,c,*], Wei Chen [d,e]

[a] Business School, Central South University, 410083, Changsha, China
[b] LMIB & School of Mathematics and Systems Science, Beihang University, 100191, Beijing, China
[c] Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, 100191, Beijing, China
[d] Harvard Medical School & Brigham Women Hospital, Channing Division of Network Medicine, 181 Longwood Ave, Boston, MA 02115, USA
[e] Center for Complex Network Research, Department of Physics, Northeastern University, Boston, MA 02115, USA

HIGHLIGHTS

- We observe strongly positive lagged correlations among stocks in Chinese stock market.
- The correlation is not constant throughout the period but emerges in certain periods.
- Dynamic lead–lag structures are uncovered in the form of temporal network structures.
- We show significant market events can be distinguished in the Jaccard matrix diagram.

ARTICLE INFO

Article history:
Received 18 August 2017
Received in revised form 4 February 2018
Available online 4 March 2018

Keywords:
Lead–lag relationships
Jaccard matrix
Temporal networks
Chinese stock market

ABSTRACT

Understanding the correlations among stock returns is crucial for reducing the risk of investment in stock markets. As an important stylized correlation, lead–lag effect plays a major role in analyzing market volatility and deriving trading strategies. Here, we explore historical lead–lag relationships among stocks in the Chinese stock market. Strongly positive lagged correlations can be empirically observed. We demonstrate this lead–lag phenomenon is not constant but temporally emerges during certain periods. By introducing moving time window method, we transform the lead–lag dynamics into a series of asymmetric lagged correlation matrices. Dynamic lead–lag structures are uncovered in the form of temporal network structures. We find that the size of lead–lag group experienced a rapid drop during the year 2012, which signaled a re-balance of the stock market. On the daily timescale, we find the lead–lag structure exhibits several persistent patterns, which can be characterized by the Jaccard matrix. We show significant market events can be distinguished in the Jaccard matrix diagram. Taken together, we study an integration of all the temporal networks and identify several leading stock sectors, which are in accordance with the common Chinese economic fundamentals.

1. Introduction

Financial markets are a primary complex system composed of massive components [1,2]. The dynamic interactions among various components typically induce many interesting structure patterns or evolving organizations [3,4]. Extracting

* Corresponding author.
E-mail address: jiangxin@buaa.edu.cn (X. Jiang).

https://doi.org/10.1016/j.physa.2018.02.112

a deep insight into the behavior and structure of these dynamic interactions is not only of great scientific significance but also helpful for economic portfolio risk estimation [5]. Actually, since the milestone work of Markowitz [6], understanding the correlations among asset prices has been considered as an efficient way for pursuing optimal portfolios and reducing systemic risks [7].

As a typical integrated financial system under large amounts of regulation, the stock market exhibits complex correlation behaviors characterized by continuous fluctuations and transient of prices and volumes. Various measures derived from linear and nonlinear characteristics of price return dynamics have been established as useful indicators for risk modeling and forecasting [8]. Methods from network theory are also widely applied, such as recent works considering heterogeneous coupling [9] and evolution of cooperation [10,11]. Generally, the correlation between pairs of elements is often investigated, accomplished with the analysis of the corresponding correlation matrix. For the analysis of correlation, related topics include the estimation and dynamics of symmetric and asymmetric correlation [12,13], detrending the correlation [14], lead–lag effects [15,16] and so on. Much of the significant research [17,18] in financial lead–lag relationships in the last two decades has validated that statistically significant lagged correlation relationships exist in financial markets. Studies of intra-day seasonalities of stock dynamics suggest that correlation between stocks increases throughout the day [19]. In [20], the authors successfully analyze lead–lag relationships at arbitrarily high frequencies without additional imputation bias.

For the correlation matrix, random matrix theory (RMT) is widely used to characterize the spectral properties of a sample correlation matrix [21,22]. Clustering algorithms [23] and the principal component analysis (PCA) [24] can also be applied to seek for the emergence of multi-scale structures from the correlation matrix. Other popular approaches involve the minimum spanning tree (MST) [25], hierarchical structure analysis and financial networks [26,27]. The dynamical evolution of cross-correlations is also investigated by introducing moving time-window method [28].

Since lots of works have been done to investigate the synchronous correlations, an understanding of asynchronous, or lagged correlation is also required to reveal lead–lag relationships in stock markets. In this work, we consider dynamical evolutions of lagged correlations among 218 stocks in the Chinese Stock Market. In particular, we consider lagged correlations with different values of time lag in a moving time window. In each time window, an asymmetric lagged correlation matrix is introduced to measure the lead–lag relation strengths. By defining strong lagged-correlation as weighted edges, we uncover dynamic lead–lag structures expressed by temporal networks [29]. Our analysis reveals how the collective lead–lag phenomenon emerges and evolves during the period, how robust the lagged-correlation structures are, and quantifies the distribution of leading stocks assigning to stock sectors. Our results present evidence for the existence and dynamics of strong lead–lag relationships in the Chinese stock market.

2. Lagged correlation of stock returns

In this paper, we analyze the historical daily closing prices of stocks from the Shanghai and Shenzhen Stock 300 Index (hereafter CSI 300), which is used to replicate the performance of 300 stocks traded in the Shanghai and Shenzhen stock exchanges. The data set is obtained from the Wind database.[1] In order to check the dynamic behavior of lead–lag correlations during a sufficient long period, we select trading data from January 4, 2010 to December 28, 2016 which contains 1698 trading days. To make sure that all the testing stocks have enough data during this period, we filter the original data and choose 218 stocks of complete records listed on the CSI 300 index.

Before proceeding, we first transform the daily closing price series to additive return series as

$$G_i(t) \equiv log(P_i(t + \Delta t)) - log(P_i(t)). \tag{1}$$

Here $P_i(t)$ is the price of stock i at time t, and Δt is the time resolution scale. Since we mainly discuss daily return series, Δt is set to be 1 day. We consider a lagged correlation analysis to investigate the emergence of lead–lag structures between all possible stock pairs. The discrete lagged-correlation between stock i and stock j is defined as

$$C_{ij}(d) = \frac{\sum_{t=1}^{N-d}[(G_i(t) - \langle G_i\rangle)(G_j(t-d) - \langle G_j\rangle)]}{\sqrt{\sum_{t=1}^{N-d}(G_i(t) - \langle G_i\rangle)^2}\sqrt{\sum_{t=1}^{N-d}(G_j(t-d) - \langle G_j\rangle)^2}}. \tag{2}$$

Here d is the time lag, which takes integral values $0, \pm 1, \pm 2, \pm 3, \ldots$. If $d = 0$, which means there is no time lag, Eq. (2) degenerates into the standard synchronous Pearson's cross-correlation coefficient [16]

$$\rho_{ij} = \frac{\sum_{t=1}^{N}[(G_i(t) - \langle G_i\rangle)(G_j(t) - \langle G_j\rangle)]}{\sqrt{\sum_{t=1}^{N}(G_i(t) - \langle G_i\rangle)^2}\sqrt{\sum_{t=1}^{N}(G_j(t) - \langle G_j\rangle)^2}}. \tag{3}$$

Since we are interested in the dynamics of the lagged-correlation defined in Eq. (2), we introduce a moving window rolling through the whole time line. Over each time window, the corresponding $C_{ij}(d)$ is calculated with given d and stock pair (i, j). Here length of the moving window T is fixed to be 300 trading days and moving step $\Delta T = 1$ trading day.

[1] www.wind.com.cn/en/edb.html.

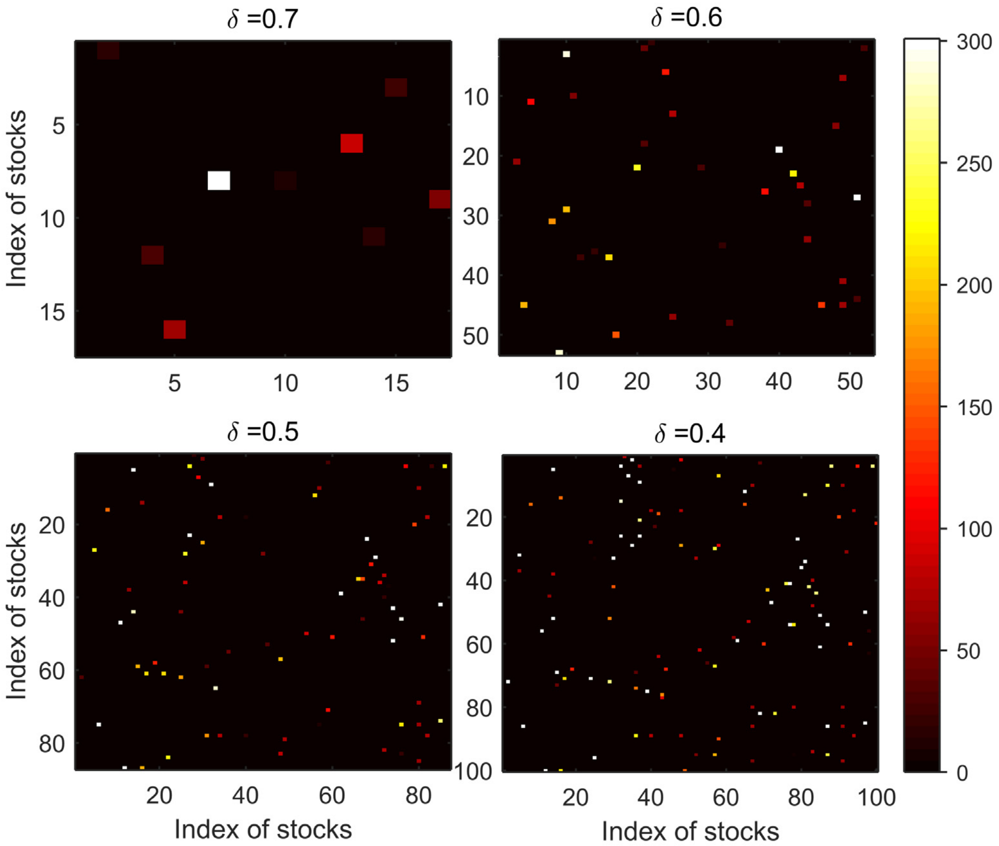

Fig. 1. Emergence of lead–lag correlations between all pairs of stocks. Illustrations of matrix $S_{i,j}$ for four values of δ. Here time lag $d = 1$. Black points indicate there is no lead–lag correlation between the considered pair of stocks while colored points identify the emerging frequency of lead–lag correlations. To make a compact illustration, in the picture we only present stock indices which ever have lead–lag relations with others. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We show high positive lagged-correlations between numbers of stock pairs emerge as time evolves. Generally, correlation coefficient values may range from $+1$ to -1, where $+1$ indicates a perfect positive relationship and -1 indicates a perfect negative relationship. 0 means no relationship exists. In this work, we introduce the widely used threshold model to characterize the positive lead–lag relationship between stocks. That is, for any pair of stocks, a lead–lag relationship is created if the corresponding correlation coefficient $C_{ij}(d)$ becomes larger than a given threshold δ. Concretely, by setting a fixed time lag d and threshold δ, we introduce a signal function $s^t_{i,j}$ to identify whether there exists a lead–lag relationship between stock i and stock j during time window t. Here

$$s^t_{i,j} = \begin{cases} 1 & C_{ij}(d) > \delta \\ 0 & else. \end{cases} \tag{4}$$

We varies the threshold δ and make a statistic of lead–lag relations over all moving windows with given d. In this sense, the summation of matrix $s^t_{i,j}$ over index t can show how many times a lead–lag relationship emerges during the whole testing time period. We denote

$$S_{i,j} = \sum_t s^t_{i,j}. \tag{5}$$

In Fig. 1, for $d = 1$, we set four values of the threshold $\delta = 0.7, 0.6, 0.5, 0.4$ and check $S_{i,j}$ for all time windows. It can be clearly observed that, as δ decreases, more and more lead–lag relations emerge. For instance, if δ is set to be 0.5, 87 stocks, a ratio about 40%, have lead–lag relationships with others during the whole period. In Fig. 2, we give an example of a lagged correlation between a selected pair of stocks. We show three corresponding lagged-correlation coefficients $C_{ij}(d)$ between 2 selected stocks as the time window moves. We observe that a high positive 1 day lagged-correlation suddenly emerges and lasts for a period of more than 200 days, as the blue line shows. For $d = 0, -1$, the correlation keeps weak during the whole period.

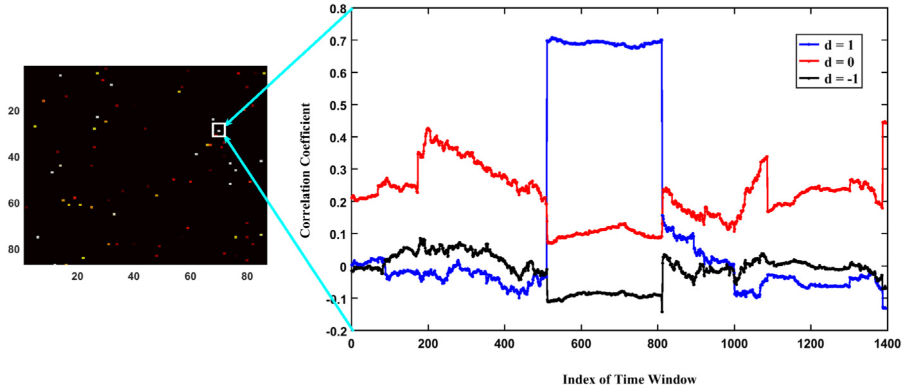

Fig. 2. An example of a long lasting lagged correlation. Left panel: Left bottom picture in Fig. 1. A sample stock pair (i, j) which has long lasting lagged correlation is identified by a square. Right panel: Lagged-correlation coefficients $C_{ij}(d)$ between stock i: Yutong Bus Co., Ltd and stock j: Sunshine City Group Co., Ltd. Red dots: cross-correlation with no time lags. Blue dots: Lagged-correlation $C_{ij}(1)$. Black dots: Lagged-correlation $C_{ij}(-1)$. A high positive 1 day lagged-correlation is observed at the middle of the time line, where stock j behaves as a leading factor of stock i. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

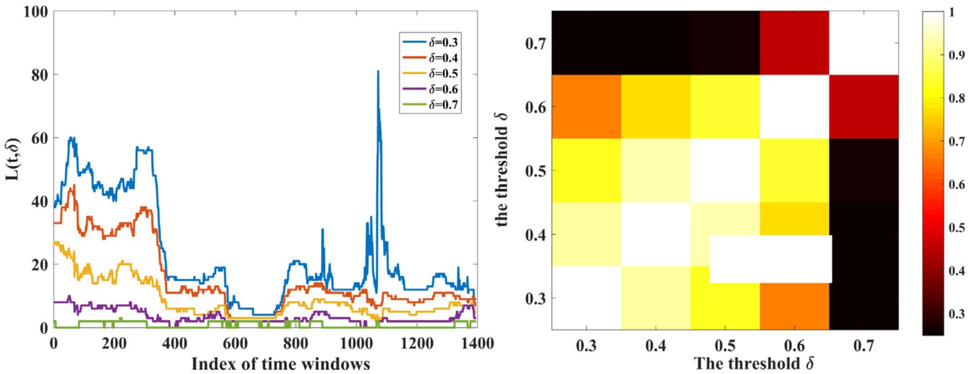

Fig. 3. Patterns and similarities of time series $L(t, \delta)$. Left panel: $L(t, \delta)$ changes in time with various δ values. Right panel: The similarity (correlation coefficients) between $L(t, \delta)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In order to give an intuitive description of the relationship between the emergence of lagged-correlation and the threshold δ, we check the total number of lagged correlation coefficients larger than δ between all pairs of stocks, which is denoted by $L(t, \delta) = |\{C_{ij}(d)|C_{ij}(d) > \delta\}| = \sum_{(i,j)} s^t_{i,j}$, as a function of δ. In Fig. 3(left panel), for $\delta = 0.3, 0.4, 0.5, 0.6, 0.7$, we show how $L(t, \delta)$ changes with threshold δ over all time windows. It is shown that for smaller δ, there is relatively larger value of $L(t, \delta)$ during the whole period. As the threshold increases, $L(t, \delta)$ descends accordingly (refer to the colored lines in Fig. 3(left panel)). In order to see whether this result is robust in time as δ varies, we consider the similarity between different $L(t, \delta)$. In general, If $L(t, \delta)$, which actually are time series, share a common feature or pattern, they should be quite similar. Here we use correlation coefficient between $L(t, \delta)$ to measure the similarity. In Fig. 3(right panel), we manifest a similarity matrix between $L(t, 0.3)$, $L(t, 0.4)$, $L(t, 0.5)$, $L(t, 0.6)$, $L(t, 0.7)$. It is clear that $L(t, 0.7)$ has a quite different pattern which differs a lot from the others, while $L(t, \delta)$ with values from 0.3 to 0.6 are similar. Thus, we argue that the results related to $L(t, 0.4)$ or $L(t, 0.5)$ should be robust in time according to the high similarities shown in the matrix.

As the following discussed method can be similarly applied to any cases with various thresholds, without loss of generality, we choose a fixed value for δ in the following discussion. Actually, if the coefficient's value lies between 0.5 and 1, we consider it to be a high positive correlation. In order to provide an overall analysis of the lagged-correlations, we construct lagged-correlation matrix M with a given lag d in time window t. The elements $M^t_{i,j}(d)$ is defined as

$$M^t_{i,j}(d) = \begin{cases} C_{ij}(d) & i \neq j \\ 0 & i = j. \end{cases} \tag{6}$$

Here the superscript t in $M^t_{i,j}(d)$ denotes the index of time window. According to Eq. (2), we have $C_{ij}(d) = C_{ji}(-d)$, which indicates that $M^t_{i,j}(d) = M^t_{j,i}(-d)$. The matrix M thus constructed is asymmetric. To simplify, the diagonal elements are set to be zero as we only consider lagged cases.

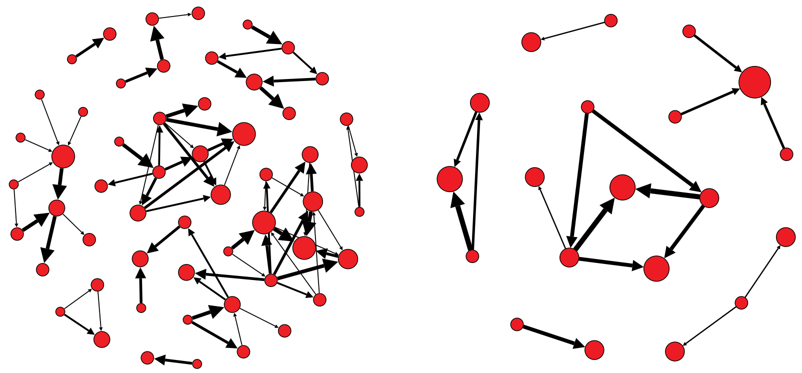

Fig. 4. Different lag-correlation structure within a time window. As edges point to lead stocks, node with highest in-degree identify the most leading stock. In both panels node size indicates in-degree and edge width indicates time lag d (thin–thick as weight d varies from 1 to 4). Left panel: Temporal lag-correlation network generated within the 30th time window (from 2010.02.12 to 2011.05.16). 55 nodes and 69 edges, with 9 weak connected components. Right panel: Network generated within the 1000th time window (from 2014.02.24 to 2015.05.14).20 nodes and 17 edges, with 6 weak connected components.

Table 1
Statistics of lagged-correlation coefficients.

Period	T	d	$\langle C_{ij} \rangle$	σ_C	Skewness	Kurtosis
2010.1.7–2016.12.23	1692	0	0.304	0.107	0.502	3.754
2010.1.7–2016.12.23	1692	1	0.015	0.036	0.071	3.940
2010.1.7–2016.12.23	1692	2	−0.012	0.029	0.178	5.471
2010.1.7–2016.12.23	1692	3	$-5.501 * 10^{-5}$	0.027	0.505	7.457
2010.1.7–2016.12.23	1692	4	0.017	0.029	0.618	7.064
2010.1.7–2011.4.7	300	0	0.297	0.149	0.405	3.054
2010.1.7–2011.4.7	300	1	$5.334 * 10^{-4}$	0.065	0.371	6.636
2010.1.7–2011.4.7	300	2	−0.007	0.057	0.354	6.319
2010.1.7–2011.4.7	300	3	0.027	0.056	0.561	8.072
2010.1.7–2011.4.7	300	4	−0.006	0.054	0.628	9.923
2012.11.21–2014.2.24	300	0	0.224	0.134	0.728	4.242
2012.11.21–2014.2.24	300	1	−0.001	0.067	0.098	3.925
2012.11.21–2014.2.24	300	2	0.001	0.060	0.119	4.799
2012.11.21–2014.2.24	300	3	−0.018	0.059	0.247	5.108
2012.11.21–2014.2.24	300	4	0.021	0.064	0.188	3.834

Before we discuss the structure of lagged-correlations, we first report several statistical quantities of the estimated lagged-correlation $C_{ij}(d)$ in Table 1. Concretely, we show the mean $\langle C_{ij}(d) \rangle$ and standard deviation σ_C of entries $C_{ij}(d)$ in the lagged-correlation matrix M. Skewness and kurtosis of $C_{ij}(d)$ with various d values are also presented. We check these statistical quantities within different periods. The length of the moving window T is set to be 1692 and 300. We note that for both long period ($T = 1692$) and short period ($T = 300$), the skewness of $C_{ij}(d)$ always take positive values, which indicates the tail of the lagged-correlation distribution on the right is longer than the left side. In fact, for $d > 0$, $C_{ij}(d)$ with larger positive values are located at the longer tails.

In the following analysis of M, we investigate the structure of M with various d values. As we state in the previous section, if $M^t_{i,j}(d) > 0.5$, we consider the leading stock j is strongly cross-correlated with the lagging stock i at d days in time window t. However, after calculating $M^t_{i,j}(d)$ over all the time windows, we find that for a given d, the number of strong lag-correlation pairs in a time window is relatively small. In fact, most coefficient values distribute with a mean value around zero, as shown in Table 1. A small fraction of bright points in Fig. 2(left panel) indicates the structure of the strong lag-correlation might be of microscale or mesoscale compared with the whole correlation structure. In Fig. 2(right panel), we compares the dynamic pattern of the average synchronous correlation $\langle M^t_{i,j}(0) \rangle$ with the lagged one $\langle M^t_{i,j}(1) \rangle$. We observe the averaged synchronous correlation fluctuates a lot during the economic period while the average lagged correlation stays smoothly around 0, which confirms that the emergence of strong lag-correlation may also not be detected by statistical methodology within a relatively large time scale.

From Fig. 2 one can see that the strong lagged correlation between two stocks is active only for restricted periods of time. The existence of strong lag-correlation is a temporal phenomenon. In order to understand how frequently the strong lagged correlation structure emerges and how it evolves with time, one needs a principled, consistent way of characterizing the temporal interactions between stocks and the time of the interaction. A common approach to achieve this goal is to represent the lagged correlation structure as a temporal directed network [29–31]. In this network, lag-correlated stocks are represented by network nodes, while the lead–lag relationships can be represented as directed weighted links, where the direction is from the lag node to the lead one and the weight denotes the time lag d. In this way, we can generate a

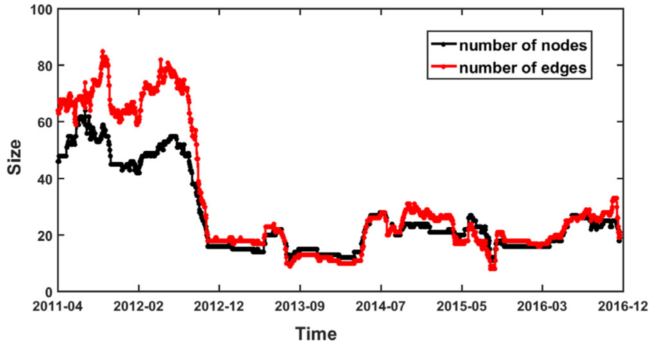

Fig. 5. Evolution of sizes for the lead–lag group. Red lines: number of nodes in the lag-correlation temporal networks. Black line: number of edges in the lag-correlation temporal networks. Results are evaluated in the period 2011–2016. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

series of temporal networks from the lagged-correlation matrix M obtained in Eq. (6). To emphasize and visualize the strong correlation, we set a threshold $\delta = 0.5$ for the strength of the lagged correlation. If $M^t_{i,j}(d) > \delta$, there is a directed link points from node i to node j with weight d. In the following discussion, we study temporal networks in which edges have weight $d = 1, 2, 3, 4$.

In Fig. 4, we present a visual comparison of two temporal network structures corresponding to the lagged-correlation relationship in different time windows. This picture intuitively illustrates that lag-correlation structure differs a lot in these two networks. Especially, in the left panel, the stocks form several complex subgraphs indicating multi lead–lag relationship might exist. While in the right panel, only some simple subgraphs, such as triangles, two-node chains, etc., form the lead–lag pattern. It is interesting that both of the networks are not connected, which means we might not find a stock that can lead the whole stocks in the network.

A basic pattern of the lagged correlation structure is the size of the corresponding temporal network including the number of nodes and edges. Fig. 5 illustrates how the size of lead–lag group in China's stock market varies during year 2010 to 2016. The size of lead–lag stock group experienced a rapid drop during the year 2012. Before the middle of year 2012, the number of edges is relatively larger than the node number, which indicates the lead–lag correlation network is more heterogeneous, including some high degree nodes. While after 2013, the size of edges becomes the same as the node numbers, indicating most lead–lag structures become simple symmetric patterns, which corresponds to the patterns shown in Fig. 5. Actually, during the year 2012, in order to tackle the imbalances in the economy development, the Chinese government has unveiled economic measures aimed at promoting a more balanced economic mode instead of the previous rapid growth one. China cut its economy growth rate down to 7.5%, which is the lowest rate ever since 2004. We argue that the lag-correlation size reflects this evolution. Indeed, the shrink of size signaled a move towards a rebalancing of stock market.

3. Stability of lag-correlation structures

Another interesting problem is whether there is any persistent patterns in these temporal networks, which may explain to what extent the lead–lag relationships persist during the sampling time windows. We find that some validated links and subgraphs are indeed largely persistent throughout the period. In order to give a quantitative analysis of the persistent patterns, we provide a detailed analysis on the constructed temporal network ensembles. First, we introduce the following Jaccard Index [32] to quantify the proportion of shared links between two networks:

$$J(m, n) = \frac{|L_m \cap L_n|}{|L_m \cup L_n|}. \tag{7}$$

Here L_m is the edge set of network m. Each edge we considered is distinguished by direction and weight. In this sense, a large value of the Jaccard Index indicates that a similar lead–lag relationship pattern persists in two networks. In Fig. 6, upper panel, we provide a matrix of Jaccard index $J(m, n)$ between each pair of directed edge sets corresponding to all the lead–lag correlation networks generated in each time window. We find that the persistence of lead–lag relationship may across periods of different lengths. The Jaccard indices are generally high in the neighborhood time windows, however, decay to zero rapidly as time interval $|m - n|$ becomes larger, indicating the lead–lag relationship are preferred to exist in a certain single period (identified by the dark red area along the diagonal of the matrix) and never appear again in the following period. Further, the Jaccard indices are quite heterogeneous throughout the whole time period. Actually, the size of the dark red square areas varies corresponding to different economic periods. As these dark red squares are formed by $J(m, n)$ of large values, we denote them as "cores" of the Jaccard indices. It is shown that the sizes of cores change as the stock market fundamental varies. To investigate this, we plot the corresponding CSI 300 Index, as an indicator of the stock market fundamental, in the lower panel, Fig. 6. Red arrows divide the time into several typical economic periods, which correspond

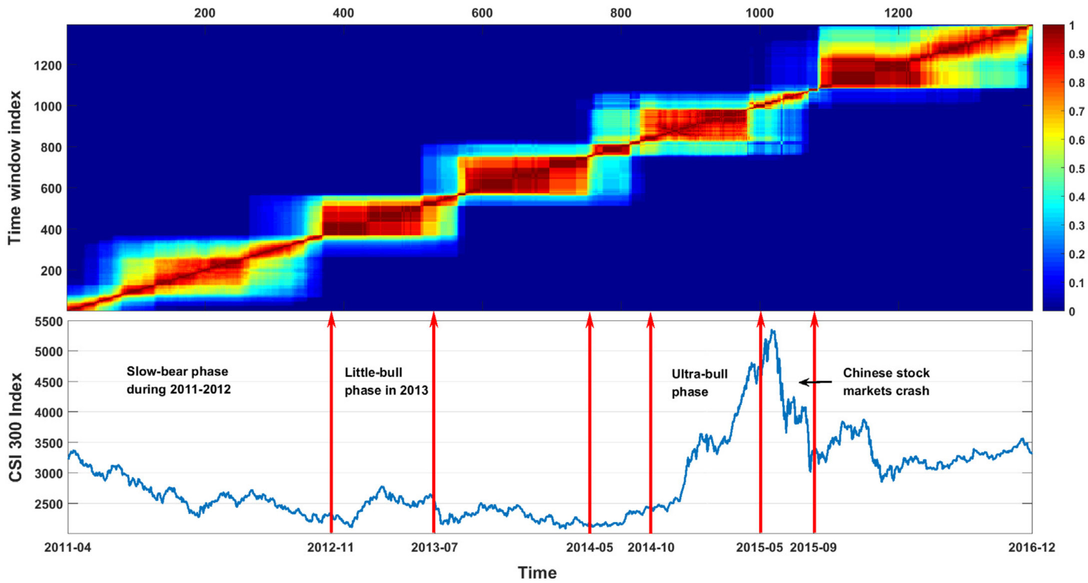

Fig. 6. Dynamic patterns in the Jaccard index matrix. Upper panel: A matrix of Jaccard indices between all sets of edges corresponding to temporal networks for all time windows. Lower panel: Corresponding CSI 300 index in the same time period. The Jaccard indices matrix is visualized as a rectangle instead of a square in order to make a comparison with the CSI 300 index. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to several featured stock market phases. We note each bull or bear market corresponds to a distinct core in Jaccard matrix. For example, the ultra-bull market and giant crash in Chinese stock market 2015 can be clearly identified in the Jaccard matrix.

We also examine the distributions of edge persistence for all networks generated in Fig. 6. Here the edge persistence refers to the fraction of temporal networks in which a given edge appears. By quantifying the persistence of these edges, we observe that a large proportion (about 18%) of edges seem to be persistent for about 300 days, which is approximately the length of one economic year. In Fig. 7, most of the individual edges are persistent less than 150 days, a number of edges last for 150 to 300 days. We argue that stable lead–lag correlation can be revealed from the edge persistence analysis.

Focusing for simplicity in an integration of the whole time interval, we study an integrated network [33] which is obtained by intergrading all the temporal networks, i.e., a union of all the node sets and edge sets generated over all time windows. We demonstrate the integrated network is a typical scale-free network with a power-law degree distribution. A visualization plot of this network is shown in Fig. 8(left panel). Red nodes identify the hubs while blue ones behave as leaves. In fact, we also give distributions of degrees, in-degrees and out-degrees in Fig. 8(right panel). The largest in-degree is 34 while the largest out-degree is 38, which indicates that during the period the largest number of a stock to lead others or to be leaded by others is less than 40. Thus, we argue that this number might be considered as an influence radius for the lead–lag relationship in Chinese stock market.

We also investigate, during the entire economic period, what are the common features of leading stocks, which are represented by nodes with nonzero in-degrees in the integrated network. Generally, it is often believed that the most liquid stock sectors tend to be leaders, since it should much less time for the liquid sectors to incorporate market information into prices. In Table 2, from the point of view of market, we present which stock sectors the leading stocks belong to and the number of stocks in each sector. Table 2 illustrates the discriminatory leading power of each stock sector by measuring the proportion of leading stocks. The strongest leading sector is the Financial, which is in agreement with the fact that financial market develops rapidly in China these years. Actually, the financial sector, which includes banks and securities companies, is almost a necessary supporting component for the stock market. The second largest sector is Real estate, which might be a principal element for the GDP's growth in China for recent years. In fact, the real estate sector in China was growing so rapidly that it attracts massive investigators. The following main leading sectors are Information technology, Electrical equipment and Medicine. Since the most dominant sector of China's economy remains its manufacturing and industries, we argue these emerging industries might accumulate more and more liquidity in the stock market and become significant leaders. We also notice that the leading stocks contains relatively smaller amount of traditional industry stocks including metals & mining and energy, indicating a loss of liquidity and attraction in these areas.

Note that in this paper, the stocks we studied come from the Shanghai and Shenzhen Stock 300 Index, which consists of the largest and representative companies from various sectors. Previous studies have shown that these stocks have strong relationship with Chinese economy [34–36]. Therefore, although to some extent Chinese stock market is correlated with the policy of central government, in fact, these stocks from the CSI 300 as a whole show stability since the economy of China

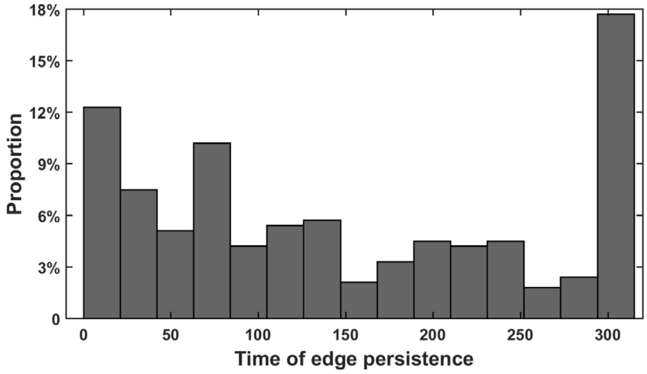

Fig. 7. Distribution of edge persistence. Distribution of the edge persistence lengths shows a similarity at the level of individual edges. The time for the edge persistence ranges from 1 to 301 days. All the edges generated in Fig. 6 are analyzed. We find the validated edges are generally persistent within 150 days or more than 300 days.

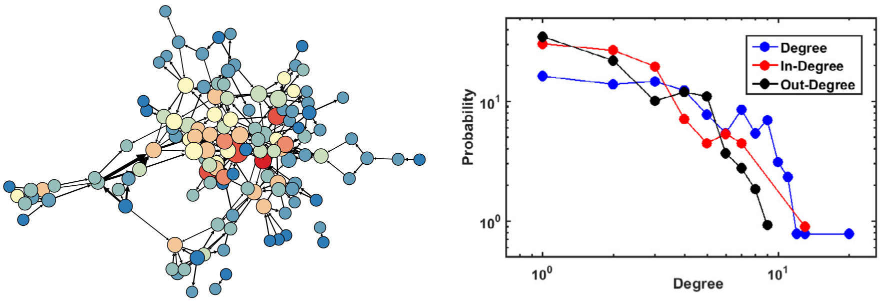

Fig. 8. Left panel: An integrated network with 128 nodes and 288 directed edges. Nodes with small degrees are colored by blue while nodes with large degrees are colored by red. Mediums are identified by yellow. Right panel: Distribution for degree (blue), in-degree (red) and out-degree (black) of nodes in the network. Data is plotted in log–log axis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
List of sectors for leading stocks.

Sector	Number	Proportion
Financials	17	15.3%
Real estate	16	14.4%
Information technology	12	10.8%
Electrical equipment	12	10.8%
Medicine	12	10.8%
Food & staples retailing	6	5.4%
Machinery	5	4.5%
Electric utilities	5	4.5%
Telecommunication	4	3.6%
Automobiles	4	3.6%
Metals & mining	4	3.6%
Media	4	3.6%
Transportation	3	2.7%
Aerospace & defense	3	2.7%
Chemicals	2	1.8%
Energy	2	1.8%

has developed steady in the last decades, except for the period of global economic crisis. From the perspective of network, it is also quite interesting to carry out a detailed exploration of the relationship between Chinese stock market and the government policy, and this may also be our further research direction.

4. Conclusion

In a summary, we investigate the emergency and structure of lead–lag relationships on the Chinese stock market. A lagged cross-correlation function is used to estimate the lead–lag strength between different pair of stocks. We observe high positive lagged-correlation between pairs of stocks emerges as time evolves. Lead–lag relationships between various stocks show different behaviors. To gain a more extensive understanding of the lagged-correlation structure, we introduce an asymmetric lagged-correlation matrix to construct temporal correlation networks. We find that the size of lead–lag group in China's stock market varies a lot during the 2010 to 2016 period. The largest lead–lag group appeared between 2011 and 2012, which is a period for a rebalancing of Chinese stock market. By analyzing Jaccard index of the temporal network ensembles, we find lead–lag relationships display a non-constant profile during different economic periods. Bull or bear market can be identified according to the variation of strong lead–lag correlations. The study of distributions of lead–lag persistence shows that stable lead–lag correlations mainly last either less than 150 days or more than 300 days, indicating diversities in the persistence diagram. Finally, we analysis the leading stock sectors in the past evolution of Chinese stock market, principal leading stocks are mainly from financial, real estate, IT, electrical equipment and medicine sectors, which corresponds to the development of China's economy.

In the future, we plan to use the lead–lag structure of stock market to develop some investment tragedies. This is inspired from the idea of correlation arbitrage in forex actions.

Acknowledgments

Lisi Xia and Daming You would like to acknowledge support from NSFC, China Grants No. 71573283. This work is partially supported by NSFC, China Grants No. 11201017, No. 11290141 and No. 11571028. Besides, Xin Jiang also would like to acknowledge support from the National Key Research and Development Program of China under Grants No. 2017YFB0701702.

References

[1] R.N. Mantegna, H.E. Stanley, Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, 1999.
[2] S. Sinha, A. Chatterjee, A. Chakraborti, B.K. Chakrabarti, Econophysics: An Introduction, John Wiley & Sons, 2010.
[3] H.E. Stanley, L.A.N. Amaral, S.V. Buldyrev, P. Gopikrishnan, V. Plerou, M.A. Salinger, Proc. Natl. Acad. Sci. USA 99 (suppl 1) (2002) 2561–2565.
[4] S. Zhao, Q. Xie, Q. Lu, X. Jiang, W. Chen, Europhys. Lett. 112 (2) (2015) 28002.
[5] T.M. Barnhill Jr., W.F. Maxwell, J. Bank. Financ. 26 (2) (2002) 347–374.
[6] H. Markowitz, J. Financ. 7 (1) (1952) 77–91.
[7] G. Marti, F. Nielsen, M. Bińkowski, P. Donnat, 2017, arXiv preprint arXiv:1703.00485.
[8] B. Podobnik, D. Horvatic, A.M. Petersen, H.E. Stanley, Proc. Natl. Acad. Sci. USA 106 (52) (2009) 22079–22084.
[9] C.Y. Xia, X.K. Meng, Z. Wang, PLoS One 10 (6) (2015) e0129542.
[10] M. Chen, L. Wang, S. Sun, et al., Phys. Lett. A 380 (1–2) (2016) 40–47.
[11] M. Chen, L. Wang, J. Wang, et al., Appl. Math. Comput. 251 (2015) 192–202.
[12] A. Ang, J. Chen, J. Financ. Econ. 63 (3) (2002) 443–494.
[13] S.S. Borysov, A.V. Balatsky, PLoS One 9 (8) (2014) e105874.
[14] R. Rak, S. Drozdz, J. Kwapień, P. Oświęcimka, Europhys. Lett. 112 (4) (2015) 48001.
[15] N. Huth, F. Abergel, J. Empir. Finance 26 (2014) 41–58.
[16] C. Curme, M. Tumminello, R.N. Mantegna, H.E. Stanley, D.Y. Kenett, Office of Financial Reseach Working Paper, 2015.
[17] C. Curme, M. Tumminello, R.N. Mantegna, et al., Quant. Finance 15 (8) (2015) 1375–1386.
[18] D.Y. Kenett, X. Huang, I. Vodenska, et al., Quant. Finance 15 (4) (2015) 569–578.
[19] R. Allez, J.P. Bouchaud, New J. Phys. 13 (2) (2011) 025010.
[20] F. De Jong, T. Nijman, J. Financ. 4 (2–3) (1997) 259–277.
[21] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, T. Guhr, H.E. Stanley, Phys. Rev. E 65 (6) (2002) 066126.
[22] A. Sensoy, S. Yuksel, M. Erturk, Physica A 392 (20) (2013) 5027–5045.
[23] M. MacMahon, D. Garlaschelli, Phys. Rev. X 5 (2013) 021006.
[24] Y. Shapira, D.Y. Kenett, E. Ben-Jacob, Eur. Phys. J. B 72 (4) (2009) 657–669.
[25] G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, Phys. Rev. E 68 (4) (2003) 046130.
[26] K.T. Chi, J. Liu, F.C. Lau, J. Empir. Finance 17 (4) (2010) 659–667.
[27] W.J. Xie, M.X. Li, H.C. Xu, W. Chen, W.X. Zhou, H.E. Stanley, Europhys. Lett. 116 (2) (2016) 28006.
[28] F. Ren, W.X. Zhou, PloS One 9 (5) (2014) e97711.
[29] P. Holme, Eur. Phys. J. B 88 (9) (2015) 234.
[30] P. Holme, J. Saramäki, Phys. Rep. 519 (3) (2012) 97–125.
[31] R. Pfitzner, I. Scholtes, A. Garas, C.J. Tessone, F. Schweitzer, Phys. Rev. Lett. 110 (19) (2013) 198701.
[32] R. Real, J.M. Vargas, Syst. Biol. 45 (3) (1996) 380–385.
[33] M. Starnini, R. Pastor-Satorras, Phys. Rev. E 87 (6) (2013) 062807.
[34] Y. Qiao, C. Ying, Econ. Res. J. 4 (2001) 32–40.
[35] Q. Sun, W.H. Tong, J. Financ. Econ. 70 (2) (2003) 183–222.
[36] S. Yao, D. Luo, World Econ. 32 (5) (2009) 667–691.