Received April 7, 2018, accepted May 23, 2018, date of publication May 30, 2018, date of current version June 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2842069

Modeling the Complex Network of Multidimensional Information Time Series to Characterize the Volatility Pattern Evolution

SIYAO LIU, XIANGYUN GAO, WEI FANG, QINGRU SUN, SIDA FENG, XUEYONG LIU, AND SUI GUO

School of Humanities and Economic Management, China University of Geosciences, Beijing 100083, China

Key Laboratory of Carrying Capacity Assessment for Resource and Environment, Ministry of Land and Resources, Beijing 100083, China

Corresponding author: Xiangyun Gao (gxy5669777@126.com)

This work was supported in part by the Humanities and Social Sciences Planning Funds Project under the Ministry of Education of China under Grant 17YJCZH047, in part by the Beijing Natural Science Foundation under Grant 9174041, and in part by the fund from the Key Laboratory of Carrying Capacity Assessment for Resource and Environment, Ministry of Land and Resources, under Grant CCA2017.11.

ABSTRACT Time series processing and analyzing is one of the major challenges in big data research, especially in inferring the dynamical mechanism of time series. In this paper, we constructed a complex network from time series via exploring the evolutionary relationship among the volatility patterns. We introduced the symbolic method and sliding window to describe the volatility patterns of time series that contains multidimensional information. Meanwhile, we explored the evolutionary mechanism of these volatility patterns based on the topological characteristics in network. In our research, we selected six stock indices around the world as sample data. Interestingly, for the six networks, they all showed a ''petal-shaped'' structure which consists of a core and loops. Moreover, through analyzing the topological characteristics of the six networks, we discovered distinguished results of their overall characteristics and loop length distributions. Furthermore, we uncovered the media patterns which trigger the structure of network changing among core and loops. In a word, this paper described the volatility patterns and explored the topological characteristics in networks constructed from time series data, which provides a novel perspective to understand the evolutionary dynamic mechanism.

INDEX TERMS Time series analysis, complex networks, nonlinear dynamical systems, multidimensional information, volatility pattern.

I. INTRODUCTION

In big data era, time series is one of the most common data forms, and it has been extensively applied in various fields. Processing and analyzing large volumes of time series data is still a challenging endeavor in big data technology [1]. Compressing the multidimensional information together of the time series to describe the volatility patterns of it and exploring the evolutionary dynamical mechanism of the volatility patterns is always a fundamental problem which has also attracted extensive interests over several decades [2]–[7]. Going beyond some widely used techniques, such as econometrics [8], [9], neural networks [10]–[12], wavelet decomposition [13]–[16] and fractal [17], [18], intensive attentions have been focused on applying complex network to solve the problem that has been mentioned above [19]–[22]. In our research, we abstracted the volatility patterns as nodes, their evolutionary relationship as edges based on complex network theory. And then, we explored the evolutionary dynamic mechanism of the volatility patterns from the topological characteristics of the network.

In recent years, complex network has been widely developed, particularly after the notion of 'small-world' proposed by Barabási and Albert [23] and the 'scale free network' by Watts and Strogatz [24]. It provides a novel perspective for analyzing the complex system with a large amount of data and information. The method of complex network is based on the presented topological features which can characterize the connectivity of network and highly influence the dynamics of processes executed on the network [25], such as degree [26], shortest path [27], modularity [28], betweenness centrality [29],clustering coefficient [30] and other measures [31]–[33]. The method of complex network can not only help us understand the global dynamical characteristics of the system by performing data compression from a global perspective, but it can also serve as a magnifying glass to extract the characteristics of a unit or the relation between the units to help comprehend the dynamical regime from a local perspective. To make the best of its advantage in time series analysis, scholars have proposed a large number of improved methods: Zhang and Small [34] thought out a brand-new method to reconstruct the pseudo- periodic time series into a corresponding network. In the specific process of analysis, they represented each cycle as a node in the network. Lacasa *et al.* [35] proposed a method which is based on the visibility relationship among the nodes to transform the time series into the network. To express it in a more visual way they named it the Visibility Graph (VG) and widely applied it into various areas. Moreover, to improve the applicability of the method, scholars have improved it in various ways, such as Horizon Visibility Graph(HVG) [36], Limited penetrable visibility graph (LPVG) [37] and Multiscale limited penetrable horizontal visibility graph (MLPHVG) [38]. Donner *et al.* [39] borrowed the idea of Recurrence Plots to convert a time series into the corresponding complex network. Gao proposed some new models to analyze multivariable time series data, such as Multiplex Network-based Sensor Information Fusion Model, Multiplex multivariate recurrence network, Wavelet Multiresolution Complex Network, Multilayer Network and so on [40]–[47]. In addition, in nonlinear systems, described by time series, varieties of volatility patterns at different time points and the evolutionary mechanism of these volatility patterns have received much research attention. To make further contribution to this topic, some studies [19], [48]–[51]aimed to map the volatility patterns as nodes and their relations as edges in the network and analyze the topological and statistical information of the network to uncover the volatility evolution over time.

In this paper, we proposed a methodology to transform the time series into network by exploring the evolutionary relationship among the volatility patterns. We selected the Shanghai composite index, S&P500 index, DAX30 index, CAC40 index, Nikkei225 index and FTSE100 index daily price for empirical research. They respectively represent the economic development of China, the United States, Germany, France, Japan and British. As shown in the FIGURE 1, in the long term, the volatility trends of the four-dimensional information are similar (open price, close price, highest price and lowest price). However, in the short term, there exists asynchrony among their volatility trends. Compared with previous studies [48]–[52], which only considered close price in their researches, it might be more accurate to use the four-dimensional information of price as the research objects to explore the dynamical mechanism of volatility. In addition, from FIGURE 1 we can also notice that, for the same time

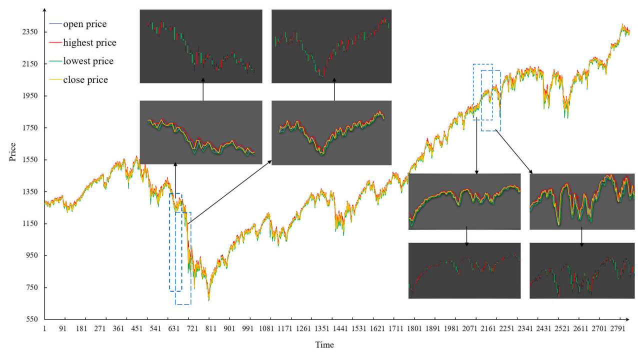

FIGURE 1. The volatility process of S&P500.

series, the information contained in different time scale is not the same. The smaller the scale is, the more detailed will be and the more variegated the volatility patterns can be described. Therefore, on the one hand, to retain much more volatility information and discover more dynamical characteristics, we concentrated on the daily price volatility. On the other hand, we calculated the quantitative relations among the four-dimensional data of the daily price (open price, close price, highest price and lowest price). After calculating the quantitative relations of the four-dimensional data, we sorted the results and symbolized them into thirteen types. In this way, the time series were converted into symbol sequence. At the same time, during the dynamic and nonlinear volatility process, the volatility patterns are various in the short term and will evolve over time, which were shown in the FIGURE 1. To describe the daily volatility patterns of the time series, we used the method of symbolization. The advantage of the symbolization is that it could not only help us filter out the redundant information and retain the core characteristics, but it could also classify the information and infer the commonality and particularity of the information. To describe the evolution process, we used the method of sliding window to dynamically divide the time series into some fragments in the same length [52] and defined the combination of symbols in each sliding window as volatility modes. After that, we selected an appropriate moving step to ensure the dynamic evolutionary process of these modes and that each fragment could contain part of the information of the former one. Due to the usage of symbolization method and sliding window, the volatility patterns possibly repeated in the evolution process and there might hide a lot of information about the relationships among the volatility patterns. Thus, we applied the method of complex network to solve this problem and analyzed the property of volatility patterns and their evolutionary relationships.

II. DATA AND METHODS

A. DATA

In this paper, we selected the daily price data of the Shanghai securities composite index (China), S&P500 index (the United States), DAX30 index (Germany), CAC40 index (France), Nikkei225 index (Japan) and FTSE100 index

(British) as sample data. The data used in this paper are from January 4, 2006, to April 20, 2017, and are downloaded from Wind.

B. CONSTRUCTION OF THE MODEL

In this paper, we proposed a model to depict the volatility patterns and their evolution mechanism in a more accurate, multidimensional and comprehensive way. We firstly compressed the multidimension information of the daily price together and used some symbols to describe it. In this way, the time series were converted into the corresponding symbol sequences. Then, we applied the sliding window theory to divide the symbol sequence into some fragments and identified the symbol combination in a sliding window as a mode. In this way, the symbol sequences were converted into the corresponding mode sequences. Finally, we considered the modes as nodes, the transformation relation among modes as edges to construct the corresponding networks.

1) VOLATILITY PATTERN IDENTIFICATION

For the stock index daily price, the volatility can be mainly reflected by four elements: the length of the real body, the color of the real body, the length of the upper shadow and the length of the lower shadow. The length of the real body shows the volatility information reflected by the open price and close price. The length of the upper shadow reflects the difference between the highest price and open price (or close price), and the lower shadow reflects the difference between the lowest price and close price (or open price). The color of the real body could reflect the rising and the dropping status of stock indices. Therefore, we described the volatility patterns from these four elements.

Step 1 (Definition of Volatility From Real Body, Color, Upper Shadow and Lower Shadow): In the analysis of real body, we learn from the previous algorithm [53] and use 'body' to describe it. For the upper shadow and lower shadow, we calculated the ratio between the real body and the extreme difference in price (the difference between the highest price and the lowest price) and used 'scale' to describe it. For the color, we selected red to represent the rising status and green to represent the dropping status. In a word, we used body, scale and color to express the volatility of daily stock index price.

$$body = \frac{|open\ price - close\ price|}{close\ price} \tag{1}$$

$$scale = \frac{|open\ price - close\ price|}{highest\ price - lowest\ price} \tag{2}$$

Step 2 (Division of Real Body and Scale): The interval division is a free parameter in our model, and different values lead to different types of volatility patterns. To avoid the arbitrariness of interval division, we divided the values of the body into four intervals: (0, 1%), [1%, 3%], (3%, 7%), [7%, 1) based on the rules proposed by Zwergel *et al.* [53] Then, we classified the scale values into three groups: (0, 30%) [30%, 70%] (70%, 1) according to the partitioning of real body.

Step 3 (Classification of Patterns):

a. When body≥7% and scale >70%, it indicates that the stock price fluctuates heavily during the day. In addition, if the close price is greater than open price, which indicates that the price rises sharply, we used "R" to illustrate such status. Otherwise, we used "G" for description.

b. When 3% < body < 7%, the scale only has value in [30%, 70%] and (70%, 1). If the value of scale is in (70%, 1), we used "r1" to represent the price rising and "g1" to represent the price dropping. If the value is in [30%, 70%], we used "r2" to indicate the price rising and "g2" to indicate the price dropping.

c. When 1% ≤ body ≤ 3%, the scale has value in (0, 30%), [30%, 70%] and (70%, 1). If it is in (70%, 1), we used "r3" to show the price rising and "g3" for the price dropping. In addition, if the scale value is in [30%, 70%], we used "r4" for the up state and "g4" for the down state. Finally, if the value is in (0, 30 %), we used "r5" to indicate that price goes up during the day and "g5" to describe that price goes down.

d. When 0 < body < 1%, it indicates there exists a small gap between the open price and close price. So, we classified this state into one group, and named it as "e".

In the process of classification, all the patterns are exclusive, and we exhibited the rule of classification in the FIGURE 2.

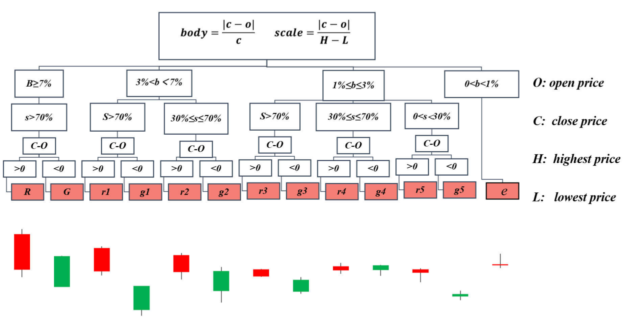

FIGURE 2. The rules of dividing volatility patterns.

2) DESCRIPTION OF THE EVOLUTION PROCESS OF THE VOLATILITY PATTERNS

For describing the evolutionary process of volatility patterns, we used sliding window to divide the time series into some segments. In addition, the method of the sliding window is superior in that the latter fragment contains $\frac{\omega - a}{\omega}(a \le \omega)$ of the information of the former one. Here, ω explains the time fragment length of each sliding window, and a represents the moving step of each sliding window. Therefore, based on this method, we can guarantee the memory of data, as well as increase the diversity of fragments to ensure the integrity of the information. Meanwhile, the evolutionary process of volatility patterns was shown when the sliding window slides.

All in all, in the specific analysis, determining the window length and the moving step length is the core content of the sliding window. The details are as follows:

Step1 (Determination of the Window Length): It is quite clear that ω is a flexible variable and its value will have direct influence on the network structure. So, it is another free parameter in the model. In our research. According to the stock market rules: trading for five days per week, except statutory holiday, we used 5 as the length of sliding window.

Step2 (Determination of the Moving Step Length): The diversity of modes and the richness of information will be influenced by the moving step length and the different value of it will also affect the network structure. In our model, the moving step length serves as a free parameter. To preserve much more volatility information of the stock market, we set the step length to be 1. By setting the sliding window length to 5 and the step length to 1, the latter fragment contains 80% of the information in the former one, which can make the fragments more memorable.

Step3 (Description of the Evolution Process): After conducting step 1 and step 2, the symbolic sequences were converted into the corresponding mode sequences.

$$\begin{cases} Y_1 = y_1 + y_2 + y_3 + \cdots + y_\omega \\ Y_2 = y_2 + y_3 + y_4 + \cdots + y_{\omega+1} \\ Y_3 = y_3 + y_4 + y_5 + \cdots + y_{\omega+2} \\ \qquad \cdots \; \cdots \\ Y_m = y_m + y_{m+1} + y_{m+2} + \cdots + y_{m+\omega-1} \quad (i = 1, 2, \ldots m) \end{cases} \tag{3}$$

From (3), we can see that ω indicates the volume of data contained in each sliding window, Y_i represents the value of the i^{th} sliding window (m in total), and y_i describes the value of the i^{th} point in the sliding window. Through the set of formulas, we can understand the evolutionary process from a mathematical standpoint: $Y_1 \rightarrow Y_2 \rightarrow \cdots Y_m$. Then, to describe the evolution process of stock index price volatility patterns in a more graphically way, we show the specific evolution process in the FIGURE 3. After calculating the value of body and scale, the symbols of the first seven days in the figure were determined as: r3, r3, r4, e, e, g4, r3, g4. According to the rules of sliding window, we figured out the modes are r3r3r4ee, r3r4eeg4, r4eeg4r3 respectively. And the specific evolutionary process of the volatility patterns is r3r3r4ee $\rightarrow$ r3r4eeg4 $\rightarrow$ r4eeg4r3.In this way, we can convert the entire sequence of symbols into an integrated mode sequence.

C. CONSTRUCTION OF THE NETWORK

After processing the data through the method mentioned above, a series of volatility modes were obtained, and these modes evolved to each other over time, as FIGURE 4 shows:

Date	Open price	Highest price	Lowest price	Close price	C-O	H-L	Body	Interval1	Scale	Interval2	Symbol	Mode
2006/1/4	1163.87	1181	1161.9	1180.96	17.09	19.1	0.01	1%≤b≤3%	0.89	s>70%	r3	
2006/1/5	1183.3	1197.83	1180.45	1197.26	13.96	17.38	0.01	1%≤b≤3%	0.80	s>70%	r3	
2006/1/6	1198.81	1215.53	1191.61	1209.42	10.61	23.92	0.01	1%≤b≤3%	0.44	30%≤s≤70%	r4	
2006/1/9	1210.31	1217.31	1205.24	1215.66	5.35	12.07	0.00	0<b<1%	0.44	——	e	
2006/1/10	1215.84	1220.75	1203.65	1220.61	4.77	17.1	0.00	0<b<1%	0.28	——	e	r3r3r4ee
2006/1/11	1219.81	1223.55	1204.04	1211.05	-8.76	19.51	0.01	1%≤b≤3%	0.45	30%≤s≤70%	g4	r3r4eeg4
2006/1/12	1208.76	1227.3	1205.28	1226.7	17.94	22.02	0.01	1%≤b≤3%	0.81	s>70%	r3	r4eeg4r3
2006/1/13	1228.15	1231.22	1214.88	1221.45	-6.7	16.34	0.01	1%≤b≤3%	0.41	30%≤s≤70%	g4	eeg4r3g4
……	……	……	……	……	……	……	……		……		……	……
2017/4/20	3165.67	3178.18	3148.18	3172.1	6.43	30	0.00	0<b<1%	0.21	——	e	g3eg4ee

Sliding Window

FIGURE 3. The formation of volatility patterns and evolution process of volatility mode.

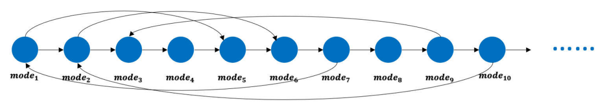

FIGURE 4. The evolution process of volatility modes.

Based on the evolutionary process shown above, we represented the process in a corresponding matrix T:

$$T = \begin{bmatrix} \omega_{1,1} & \cdots & \omega_{1,k} \\ \vdots & \ddots & \vdots \\ \omega_{k,1} & \cdots & \omega_{k,k} \end{bmatrix} \tag{4}$$

Here, $\omega_{i,j}$ represents the frequency of evolution from pattern i to pattern j.

Finally, we defined the volatility modes as nodes, the evolution relations as edges and the evolution frequency as the weight of edges to construct the corresponding volatility pattern evolution network of the stock index time series. The networks are shown in the FIGURE 5.

Based on this model, we found that the corresponding networks of these six kinds of time series all showed a 'petal-shaped' network structure, and the 'petal' consisted of a central part and some external parts. Specifically speaking, the central part is composed of nodes that have larger weighted degree in the network, and the other nodes constitute the external part of the network. In addition, all the external parts start from the central part, end with the central part, and present a 'loop' shape. The emergence of the 'petal-shaped' structure reflects the fact that, for the most part, the stock market evolves around some major volatility status, however, anomalies occasionally occur that the volatility modes evolve deviate from the core. Therefore, according to the shape of these two parts, we name the central part the 'core' and the external part the 'loop'. To understand the dynamical mechanism of the time series, we conducted a detailed research on the overall structure, core structure and loop structure of the network

III. RESULTS

In our research, we respectively established the directed weighted networks for the Shanghai Composite index,

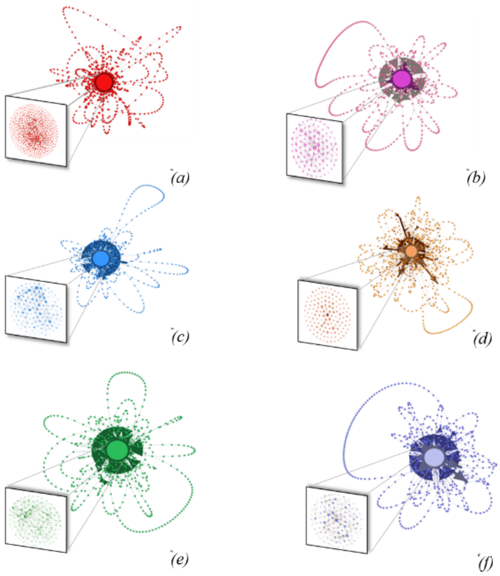

FIGURE 5. Networks of six stock indices and the figures at the bottom left illustrate the core part of the network system. In detail, (a) displays the network system of the Shanghai Composite index volatility state, (b) shows the S&P500 index, (c) illustrates the DAX30 index, (d) shows the CAC40 index, (e) shows the condition of Nikkei225 index, and (f) displays the information of the FTSE100 index.

S&P500 index, DAX30 index, CAC40 index, Nikkei225 index and FTSE100 index to explore the volatility patterns of daily price and the evolutionary mechanism of these patterns. From the network structure (FIGURE 5), we can intuitively notice that all the six stock indices show the same characteristic: 'core+loops'. Therefore, from the overall perspective, we can infer that in the real world, the stock market might be a complex system and exhibit an overall "petal- shaped" volatility dynamical mechanism.

A. ANALYSIS OF THE GLOBAL CHARACTERISTICS

In this part, we conducted a comparative analysis of the six overall networks, as shown in Table 1. From the results, we found that the Shanghai Composite index has the most abundant types of volatility modes, and the number of its nodes is twice as many as those of the other five indices. However, because of the existence of loops, for these six indices, the richness and complexity of the connections between nodes are relatively low (the values of graph densities are only 0.001 and 0.002). For the average path length (APL), it reflects the average distance between any two nodes in the network and the average efficiency of evolution among the nodes in the stock market.

From the results in TABLE 1, we found that the S&P500 has the maximum value of APL and CAC40 has the minimum value. The average clustering coefficient is an indicator for evaluating the degree of network collectivization and the closeness of nodes. The average clustering coefficients of the S&P500 and Nikkei225 are relatively high, which shows that in these two networks, nodes are more closely

TABLE 1. Topology of the whole network.

	Nodes	Edges	Graph density	Average path length	Average clustering coefficient	Average weighted degree
Shanghai composite	1775	2296	0.001	19.697	0.008	1.543
S&P500	618	818	0.002	23.024	0.042	4.597
DAX30	758	1063	0.002	14.696	0.018	3.778
CAC40	831	1130	0.002	14.093	0.019	3.471
Nikkei225	704	949	0.002	16.275	0.042	4.027
FTSE100	690	942	0.002	17.914	0.018	4.129

connected. However, for the Shanghai composite index, the network is relatively sparse. For the average weighted degree of the six networks, in the S&P500, Nikkei225 and FTSE100, the degree of conversion between nodes is much higher, but it is relatively lower for Shanghai composite index.

B. IDENTIFICATION OF KEY VOLATILITY MODES

In this part, we focused on identifying the key nodes in the network. To determine the key nodes, we firstly analyzed the statistical characteristics of the weighted degree S_i of the nodes. S_i describes the total frequency of evolution from one node to others. Therefore, it could reflect the importance and activeness of the volatility mode in the stock market.

$$S_i = \sum\nolimits_{j \in N_i} \omega_{ij} \tag{5}$$

where S_i represents the weighted degree of node i and ω_{ij} represents the frequency of evolution from node i to node j.

Then, we ranked the nodes based on the value of weighted degree and plotted the cumulative distribution curve of the weighted degree (FIGURE 6). Finally, we found the inflection points on the curves, and in this paper, we took the nodes that lie between the origin point and inflection point as the key nodes to constitute the core structure, and the other nodes to constitute the loops.

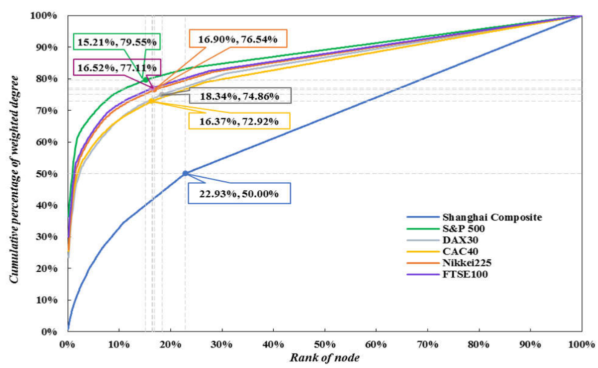

FIGURE 6. The cumulative weighted degree distribution of the Shanghai composite index, S&P 500 index, DAX30 index, CAC 40 index, Nikkei225 index and FTSE100 index.

From the weighted degree cumulative distribution of the six stock indices, we found that each network had its own inflection point. These inflection points can help us understand the different degrees of Pareto Principle. Specifically, the main volatility information of the entire network can be characterized by a small part of the volatility patterns. That is, in the Chinese stock market, which is represented by the Shanghai composite stock index, the top 405 nodes (22.82% of the total) in weighted degree can reflect nearly 50% of the volatility information of the entire network. In the American stock market which is represented by the S&P500 index, 79.44% of the volatility information of the whole market can be represented by the top 96 nodes (15.5% of the total) in weighted degree. For the German stock market, which can be represented by the DAX30 index, the top 139 nodes (accounting for 18.34% of the total) reflect 75.07% of the volatility information of the entire network. In addition, the volatility information represented by the top 136 nodes (16.37% of the total) in the weighted degree reflects 72.82% of the volatility information of the CAC400 index market, which represents the French stock market. For the Japanese stock market, by analyzing the Nikkei225 index network, we find that the 76.03% of the volatility information of the whole network can be reflected by the top 118 nodes (16.9% of the total) in the weighted degree. For the British stock market, after analyzing the FTSE100 index network, we find that the coordinate of the inflection point is (16.52%, 75.92%), which indicates that the top 114 nodes in the weighted degree (16.52% of the total) can reflect 75.92% of the volatility information of the whole network.

C. ANALYSIS OF LOOP CHARACTERISTICS

An important contribution of this article is that we found a special structure in the six networks: loops. In this part, we identified the nodes in the loops and conducted a statistical analysis of the path length of the loops. In our research, we found that the loops always started from the core and then evolved back to the core after a period. Identifying the volatility patterns in loops and measuring the duration of them could provide references to investors and policy makers. For example, they could conclude how long the volatility in loops last and when will evolve back to the core. Therefore, we explored the common laws and special phenomena that existed in the duration of the loops to provide some valuable conclusions.

We firstly identified the nodes in the loop according to the cumulative distribution of the weighted degree (mentioned in 3.1). Then, we discovered that all the ‘loops’ took key nodes as the starting points and ending points. Meanwhile, they took the non-key nodes as the intermediate process (the number of the non-key nodes is greater than one). Finally, we calculated the path length of the loops to analyze the duration of such volatility status in the stock market.

After counting the frequency of various loop lengths (FIGURE 7), we detected that for the six stock indices, most of the loop lengths are in the interval [3, 7], accounting for nearly 70% of the total. And nearly 20% of the loop lengths are in the range of [8, 15]. This statistical result reveals that when the stock market enters such state, it would have a 70% probability to end within a week and a 90% probability to end within 12 days. More importantly, we also found some special phenomena in these six markets that last for more than one month (TABLE 2).

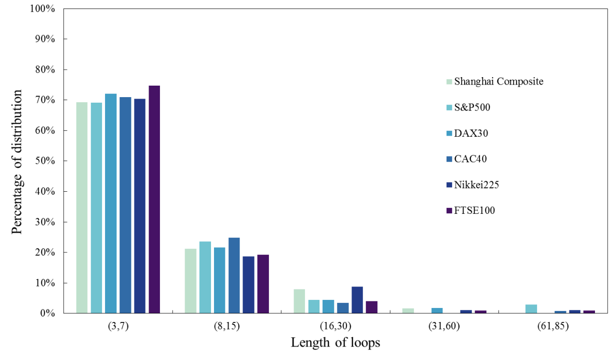

FIGURE 7. The probability distribution of loop for six stock indices.

TABLE 2. Path length of loop that last for more than one month.

Stock Index Type	Duration (day)
Shanghai composite index	37,43,51
S&P500 index	68,79
DAX30 index	34,45
CAC40 index	68
Nikkei225 index	33,67
FTSE100 index	73

For the overall network, the emergence of the loops might break the balance of the network structure, which reflects the occurrence of abnormal condition. Especially when the volatility lasts for longer than one month in the loop, the overall volatility will enter another kind of extreme state. Specifically, when evolution process happened in the core, every node has more than four choices to evolve, so it would be complex. However, when the evolution process changes from the core to the loops, we can notice that there is only one target to evolve, so the process would become simpler than that in the core. For the extreme volatility circumstance that lasts for more than one month in loops, we discovered that in the Shanghai Composite index market, the duration of the extreme circumstance is 37,43 and 51 days as shown in Table 2 and they take g4eg3g4e, r4eeg3g4, r3er3ee as the start points respectively. In the S&P500 index market, the extreme circumstances last for 68 and 79 days, and their starting nodes are eer3eg3 and r3er3ee. In the DAX30 index market, the extreme circumstances last for 34 and 45 days, their starting nodes are eer3eg1 and eeg4r4e. In the CAC40 index market, the extreme circumstance lasts for 68 days and it starts from eeg1er4. In the Nikkei225 index market, the extreme circumstances last for 33 and 67 days and they start from r3eeer3 and eer3eg3 respectively. In the FTSE100 index market, the extreme circumstance starts from eeeg3g3 and lasts for 73 days. In general, the emergence of such a structure shows different results in the six stock markets, and the results could provide some references and risk warnings for the investor and policy maker.

D. ANALYSIS OF MEDIA NODE CHARACTERISTICS

Another important contribution in our research is that we applied the statistical results of betweenness centrality to find a set of special media nodes in the core structure. More specially, these nodes are only used as the media between loops and core. These important findings could not only help us understand the evolution mechanism of volatility patterns but also provide references to investors.

As an important index for evaluating the effects of the nodes on the flow of network information, the betweenness centrality $C_B(i)$ plays a vital role in analyzing the network structure and the evolution mechanism.

$$C_B(i) = \sum\nolimits_{j<k} g_{jk}(i)/g_{jk}, \quad (j \neq k \neq i, j < k) \qquad (6)$$

where $C_B(i)$ is the betweenness centrality of node i, g_{jk} is the number of the shortest paths between two nodes j and k, and $g_{jk}(i)$ is the length of the shortest path between j and k that contains node i.

We firstly analyzed the cumulative distribution of the betweenness centrality of nodes in the whole network to determine the key media nodes (FIGURE 8). And then, we filtered the key media nodes that do not belong to the core structure to compose a new set of nodes. Finally, we compared the values of betweenness centrality of the nodes (contained in the new set) in the core network and the whole network to find the channel nodes between the core and the loop (as shown in FIGURE 9).

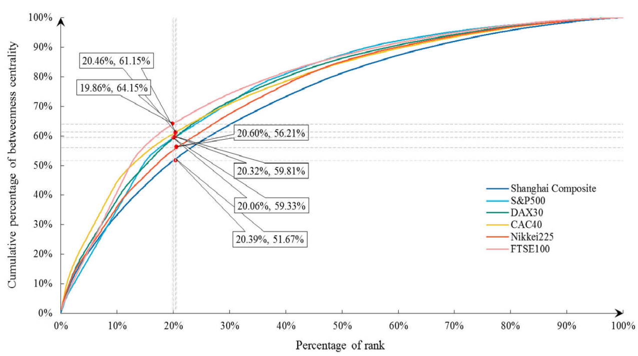

FIGURE 8. Cumulative betweenness centrality distribution for the Shanghai Composite index, S&P 500 index, DAX30 index, CAC40 index, Nikkei225 index and FTSE100 index.

Through the analysis of cumulative distribution of betweenness centrality, we found that for the six indices, there exist some nodes (approximately 20% of the total) that have a cumulative media function accounting for approximately 60% of the total (as shown in FIGRURE 8). The results shown

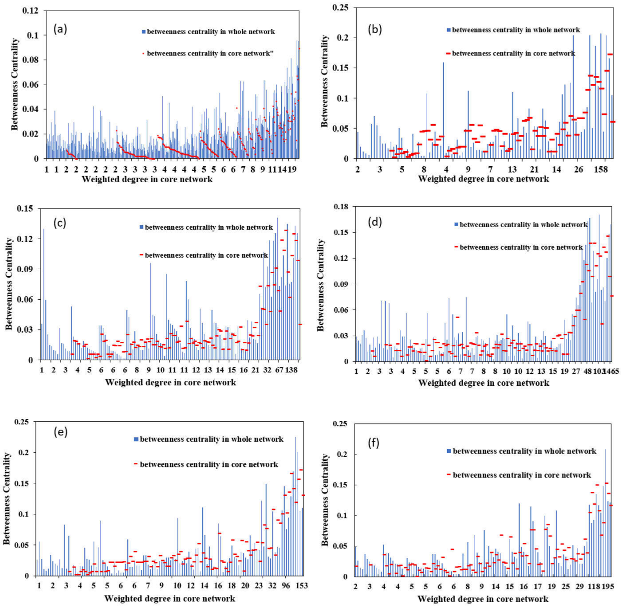

FIGURE 9. Comparison of betweenness centrality for some nodes in the core structure and the whole network: (a) explains the information of the Shanghai Composite index; (b) illustrates the results of the S&P500 index; (c) shows the comparison of the DAX30 index; (d) displays the results of the CAC40 index; (e) explains the information of the Nikkei225 index and (f) shows the results of the FTSE100 index.

that there are some representative nodes in the six networks, whose media ability could represent the ability of the whole network to a large extent. Then, by comparing the betweenness centrality of these nodes in both core and the whole network, we found an interesting phenomenon (as shown in the FIGURE 9): there are some nodes own the minimum weighted degree in the core, and their betweenness centrality values in the core and the whole network are greatly different. As shown in FIGURE 9, each blue cylinder represents the betweenness centrality of nodes in the whole network and the weighted degree of the node in the core (abscissa). Meanwhile, each read node represents the weighted degree and betweenness centrality of the node in the whole network. As we can notice above, there exist distinct gaps of the value of the betweenness centrality when the same nodes embed in different structure. For example, in (a), when the weighted degree of the nodes in the core network is 1, the betweenness centrality of them in the 'core' is 0, but in the whole network the value is relative higher. Similarly, this situation also exists in the other five networks.

Of these nodes, there are even several nodes whose betweenness centrality is 0 in the core structure, however, in the overall network these nodes are key media nodes and have a relative higher betweenness centrality. To be more precise, these nodes are the "end point" of the evolution in the core structure, as well asst the "start point" of the loops. Therefore, in the whole network, they are the important media for connecting the core and loops. Unlike other media nodes, these nodes only act as the bridges between core and loops.

From a practical point of view, the volatility patterns represented by these nodes are the important mechanisms that trigger the market volatility transformation from a frequently occurring state (nodes in the core structure) to less frequent states (represented by the nodes in the loop). As a result, the emergence of these nodes shows that the stock market volatility will have an important change in the future and that investors should pay more attention to it.

IV. CONCLUSION

In summary, we developed a novel model to transform the time series which contains multidimensional information to complex network, in the context of big data technology. In our research, we selected stock price index as research objects and synthesized multidimensional information of the price to describe the volatility patterns. Furthermore, we explored the evolutionary mechanism of these volatility patterns. Firstly, we compressed the multidimensional information on each time point into thirteen types of symbol, and then the time series were converted to the symbol sequences. Secondly, we used the method of sliding window to define the volatility modes and transformed the symbol sequence to the mode sequence. Finally, we constructed the evolutionary networks of the volatility patterns, with the modes were taken as nodes and their evolutionary relations as edges. We analyzed the topological characteristics of the network to understand the structure of the network and the dynamic mechanism of volatility evolution, including analyzing the cumulative distribution of the weighted degree, the path length of loops, the betweenness centrality and some other basic global topological indices.

In our research, we selected six stock price indices (Shanghai Composite index, S&P500 index, DAX30 index, CAC40 index, Nikkei225 index and FTSE100 index) as sample data. From the overall perspective, we firstly compared some global topological indices of the six networks, including the number of nodes, the number of edges, the graph density, the average path length, the average clustering coefficient and the average weighted degree. Then, we found an interesting phenomenon of the network structure: the network of these six indices all show an overall "petal-shaped" structure. In addition, the "petal" consists of a central part and an external part. To describe these two parts visually, we named the central part the "core" and the external part the "loops".

In the process of identifying the members of the core and loop, we analyzed the cumulative distribution of node weighted degree and found inflection points on the six curves. We then chose the nodes whose weighted degree is equal to or greater than that of the inflection points as the key nodes to form the core, and the other nodes to form the loops. In practical terms, the nodes in the core were indicative of the volatility patterns that occurred frequently in the stock market, and the nodes in the loop were indicative of the volatility patterns that occurred less frequently.

After identifying the two structures in the network, we calculated the probability of the loop path that occurred in each length interval. Meanwhile, we also clearly confirmed how long it would take evolve back to the core structure once the stock market fluctuated into the situation characterized by the loop structure. Through the analysis, it should be noted that the loop structures show varying degrees of the extreme phenomena (lasts longer than one month) in the six stock markets.

Moreover, to understand the mechanism of the evolution of the volatility patterns in a more deeper way, we extracted key media nodes in the network by analyzing the cumulative distribution of the betweenness centrality of nodes. We then intersected the key nodes with key media nodes to construct a new node set. Furthermore, we compared the value of betweenness centrality of the nodes, which are contained in the new set, between the core structure and the entire network. Finally, we concluded that these special nodes are the driving force that trigger the transformation of the network between the core and the loops.

In this paper, our primary goal is to provide an approach to describe the volatility patterns and study the evolutionary dynamics of these volatility patterns. After applying the approach to the stock market, we found a great deal of valuable information, including the transformation mechanism of the market structure and the duration of various volatility status, that might be useful to investors for decision making, risk warning and diagnostic judgments. And, the limitation of the proposed method is that it is only appropriate for the analysis of time series which owns four-dimension information.

REFERENCES

[1] M. Tahmassebpour, "A new method for time-series big data effective storage," *IEEE Access*, vol. 5, pp. 10694–10699, 2017.

[2] X. Gao, W. Fang, F. An, and Y. Wang, "Detecting method for crude oil price fluctuation mechanism under different periodic time series," *Appl. Energy*, vol. 192, pp. 201–212, Apr. 2017.

[3] H. Watanabe, Y. Sano, H. Takayasu, and M. Takayasu, "Statistical properties of fluctuations of time series representing appearances of words in nationwide blog data and their applications: An example of modeling fluctuation scalings of nonstationary time series," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 94, p. 052317, Nov. 2016.

[4] M. D. Costa and A. L. Goldberger, "Generalized multiscale entropy analysis: Application to quantifying the complex volatility of human heartbeat time series," *Entropy*, vol. 17, no. 3, pp. 1197–1203, 2015.

[5] F. Yusof and I. L. Kane, "Volatility modeling of rainfall time series," *Theor. Appl. Climatol.*, vol. 113, pp. 247–258, Jul. 2013.

[6] Y. Yan, S. Zhang, J. Tang, and X. Wang, "Understanding characteristics in multivariate traffic flow time series from complex network structure," *Phys. A, Stat. Mech. Appl.*, vol. 477, pp. 149–160, Jul. 2017.

[7] S. Scarsoglio, F. Cazzato, and L. Ridolfi, "From time-series to complex networks: Application to the cerebrovascular flow patterns in atrial fibrillation," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 27, p. 12, p. 093107, Sep. 2017.

[8] S. Celik, "The more contagion effect on emerging markets: The evidence of DCC-GARCH model," *Econ. Model.*, vol. 29, pp. 1946–1959, Sep. 2012.

[9] J.-M. Kim, H. Jung, and L. Qin, "Linear time-varying regression with a DCC-GARCH model for volatility," *Appl. Econ.*, vol. 48, no. 17, pp. 1573–1582, Aug. 2016.

[10] K. Muralitharan, R. Sakthivel, and R. Vishnuvarthan, "Neural network based optimization approach for energy demand prediction in smart grid," *Neurocomputing*, vol. 273, pp. 199–208, Jan. 2018.

[11] J. C. M. Oliveira, K. V. Pontes, I. Sartori, and M. Embiruçu, "Fault detection and diagnosis in dynamic systems using weightless neural networks," *Expert Syst. Appl.*, vol. 84, pp. 200–219, Oct. 2017.

[12] D. Pradeepkumar and V. Ravi, "Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network," *Appl. Soft Comput.*, vol. 58, pp. 35–52, Sep. 2017.

[13] Y. Bai, Y. Li, X. Wang, J. Xie, and C. Li, "Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions," *Atmos. Pollut. Res.*, vol. 7, pp. 557–566, May 2016.

[14] S. Huang, H. An, X. Gao, and X. Hao, "Unveiling heterogeneities of relations between the entire oil–stock interaction and its components across time scales," *Energy Econ.*, vol. 59, pp. 70–80, Sep. 2016.

[15] S. Huang, H. An, X. Gao, S. Wen, and X. Hao, "The multiscale impact of exchange rates on the oil-stock nexus: Evidence from China and Russia," *Appl. Energy*, vol. 194, pp. 667–678, May 2017.

[16] A. Meng, J. Ge, H. Yin, and S. Chen, "Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm," *Energy Convers. Manage.*, vol. 114, pp. 75–88, Apr. 2016.

[17] S. Bianchi and M. Frezza, "Fractal stock markets: International evidence of dynamical (in)efficiency," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 27, no. 7, p. 071102, Jul. 2017.

[18] B. Wu and T. Duan, "The fractal feature and price trend in the gold future market at the Shanghai futures exchange (SFE)," *Phys. A, Stat. Mech. Appl.*, vol. 474, pp. 99–106, May 2017.

[19] X. Y. Gao, H. Z. An, and W. Fang, "Research on fluctuation of bivariate correlation of time series based on complex networks theory," *Acta Phys. Sinica*, vol. 61, pp. 1321–1323, May 2012.

[20] Z.-K. Gao, Q. Cai, Y.-X. Yang, N. Dong, and S.-S. Zhang, "Visibility graph from adaptive optimal kernel time-frequency representation for classification of epileptiform EEG," *Int. J. Neural Syst.*, vol. 27, p. 1750005, Jun. 2017.

[21] M. Wang *et al.*, "Systemic risk and spatiotemporal dynamics of the consumer market of China," *Phys. A, Stat. Mech. Appl.*, vol. 473, pp. 188–204, May 2017.

[22] R. Du *et al.*, "A complex network perspective on interrelations and evolution features crossMark of international oil trade, 2002–2013," *Appl. Energy*, vol. 196, pp. 142–151, Jun. 2017.

[23] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–552, Oct. 1999.

[24] D J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.

[25] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, "Characterization of complex networks: A survey of measurements," *Adv. Phys.*, vol. 56, no. 1, pp. 167–242, 2007.

[26] S. N. Dorogovtsev, J. F. Mendes, and J. G. Oliveira, "Degree-dependent intervertex separation in complex networks," *Phys. Rev. E, Stat. Nonlinear Soft Matter Phys.*, vol. 73, p. 056122, May 2006.

[27] S. Dorogovtsev and J. F. Mendes, "The shortest path to complex networks," *Physics*, vol. 71, pp. 47–53, Apr. 2004.

[28] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E, Stat. Nonlinear Soft Matter Phys.*, vol. 69, p. 026113, Feb. 2004.

[29] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, Mar. 1977.

[30] M. E. Newman, "Scientific collaboration networks. I. Network construction and fundamental results," *Phys. Rev. E, Stat. Nonlinear Soft Matter Phys.*, vol. 64, p. 016131, Jun. 2001.

[31] S. S. Shenorr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of *Escherichia coli*," *Nature Genet.*, vol. 31, pp. 64–68, Apr. 2002.

[32] L. Vincent, "Graphs and mathematical morphology," *Signal Process.*, vol. 16, pp. 365–388, Apr. 1989.

[33] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, vol. 433, pp. 392–395, Jan. 2005.

[34] J. Zhang and M. Small, "Complex network from pseudoperiodic time series: Topology versus dynamics," *Phys. Rev. Lett.*, vol. 96, no. 23, p. 238701, Jun. 2006.

[35] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, "From time series to complex networks: the visibility graph," *Proc. Nat. Acad. Sci. USA*, vol. 105, pp. 4972–4975, Apr. 2008.

[36] M. G. Ravetti, L. C. Carpi, B. A. Gonçalves, A. C. Frery, and O. A. Rosso, "Distinguishing noise from chaos: Objective versus subjective criteria using horizontal visibility graph," *PLoS ONE*, vol. 9, p. e108004, Sep. 2014.

[37] T.-T. Zhou, N.-D. Jin, and Z.-K. Gao, "Limited penetrable visibility graph for establishing complex network from time series," *Acta Phys. Sin.*, vol. 61, no. 3, pp. 355–367, 2012.

[38] Z.-K. Gao, Q. Cai, Y.-X. Yang, W.-D. Dang, and S.-S. Zhang, "Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series," *Sci. Rep.*, vol. 6, Oct. 2016, Art. no. 35622.

[39] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, "Recurrence networks—A novel paradigm for nonlinear time series analysis," *New J. Phys.*, vol. 12, pp. 129–132, Mar. 2010.

[40] Z. Gao, W. Dang, C. Mu, Y. Yang, S. Li, and C. Grebogi, "A novel multiplex network-based sensor information fusion model and its application to industrial multiphase flow system," *IEEE Trans. Ind. Inform.*, to be published.

[41] Z. Gao, Y. Yang, L. Zhai, N. Jin, and G. Chen, "A four-sector conductance method for measuring and characterizing low-velocity oil–water two-phase flows," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 7, pp. 1690–1697, Jul. 2016.

[42] Z.-K. Gao, W.-D. Dang, Y.-X. Yang, and Q. Cai, "Multiplex multivariate recurrence network from multi-channel signals for revealing oil-water spatial flow behavior," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 27, p. 035809, Mar. 2017.

[43] Z.-K. Gao, P.-C. Fang, M.-S. Ding, and N.-D. Jin, "Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow," *Experim. Thermal Fluid Sci.*, vol. 60, pp. 157–164, Jan. 2015.

[44] Z.-K. Gao, S. Li, W.-D. Dang, Y.-X. Yang, Y. Do, and C. Grebogi, "Wavelet multiresolution complex network for analyzing multivariate nonlinear time series," *Int. J. Bifurcation Chaos*, vol. 27, p. 1750123, Jul. 2017.

[45] Z.-K. Gao, Y.-X. Yang, P.-C. Fang, N.-D. Jin, C.-Y. Xia, and L.-D. Hu, "Multi-frequency complex network from time series for uncovering oil-water flow structure," *Sci. Rep.*, vol. 5, Feb. 2015, Art. no. 8222.

[46] Z.-K. Gao, Y.-X. Yang, P.-C. Fang, Y. Zou, C.-Y. Xia, and M. Du, "Multiscale complex network for analyzing experimental multivariate time series," *Europhys. Lett.*, vol. 109, p. 30005, Feb. 2015.

[47] Z.-K. Gao, S.-S. Zhang, W.-D. Dang, S. Li, and Q. Cai, "Multilayer network from multivariate time series for characterizing nonlinear flow behavior," *Int. J. Bifurcation Chaos*, vol. 27, p. 1750059, Apr. 2017.

[48] X. Gao, H. An, W. Fang, X. Huang, H. Li, and W. Zhong, "Characteristics of the transmission of autoregressive sub-patterns in financial time series," *Sci. Rep.*, vol. 4, p. 6290, Sep. 2014.

[49] M. Wang and L. Tian, "From time series to complex networks: The phase space coarse graining," *Phys. A, Stat. Mech. Appl.*, vol. 461, pp. 456–468, Nov. 2016.

[50] M. Wang, L. Tian, and R. Du, "Research on the interaction patterns among the global crude oil import dependency countries: A complex network approach," *Appl. Energy*, vol. 180, pp. 779–791, Oct. 2016.

[51] M. Wang, Y. Chen, L. Tian, S. Jiang, Z. Tian, and R. Du, "Fluctuation behavior analysis of international crude oil and gasoline price based on complex network perspective," *Appl. Energy*, vol. 175, pp. 109–127, Aug. 2016.

[52] H. An, X. Gao, W. Fang, X. Huang, and Y. Ding, "The role of fluctuating modes of autocorrelation in crude oil prices," *Phys. A, Stat. Mech. Appl.*, vol. 393, pp. 382–390, Jan. 2014.

[53] B. Zwergel, C. Klein, and H. Fock, "Performance of candlestick analysis on intraday futures data," *J. Derivatives*, vol. 13, pp. 28–40, Oct. 2017.

SIYAO LIU received the B.S. degree in software engineering from the Taiyuan University of Technology, Taiyuan, China, in 2016. She is currently pursuing the Ph.D. degree in management science and engineering with the China University of Geosciences, Beijing.

Her research interests include complexity science in economics and management, time series analysis and application, data mining, big data analysis, and nonlinear system analysis.

XIANGYUN GAO received the Ph.D. degree in management science and engineering from the China University of Geosciences, Beijing, China, in 2015.

From 2016 to 2017, he was a Visiting Scholar with Pennsylvania State University, USA. He is currently an Assistant Professor with the Humanities and Economic Management, China University of Geosciences. He is the author of two books and 28 articles in *Scientific Reports*, *Physical Review E*, *Physica A: Statistical Mechanics and its Applications*, *Applied Energy*, and so on. His research interests include complexity science in economics and management, econometrics, time series analysis, nonlinear system analysis, data mining, and big data analysis.

WEI FANG received the Ph.D. degree in management science and engineering from the Beijing Institute of Technology in 2007.

Since 2007, he has been a Lecturer with the Department of Humanities and Economic Management, China University of Geosciences, Beijing. He is the author of three books and 12 articles. His research interests include complex network analysis and application, data mining, and big data analysis.

QINGRU SUN received the B.S. degree in management from Hebei University in 2015. She is currently pursuing the Ph.D. degree with the Department of Humanities and Economic Management, China University of Geosciences, Beijing.

She has published an article as the first author using complex network theory. Her research interests include energy economics, time series analysis, and economic complex network.

SIDA FENG received the B.S. degree in management from Xi'an International Studies University in 2013 and the master's degree from Sichuan University in 2016. She is currently pursuing the Ph.D. degree with the Department of Humanities and Economic Management, China University of Geosciences, Beijing.

She has published two papers as the first author using complex network theory. Her research interests include economic complex network analysis and big data analysis.

XUEYONG LIU received the M.S. degree in communication engineering from the China University of Geosciences, Beijing, where he is currently pursuing the Ph.D. degree with the Department of Humanities and Economic Management.

His research interests include pattern recognition and nonlinear time series analysis. He is mainly focused on energy financial research, especially the financial volatility effects among the energy stock market, time series analysis, and computer science.

SUI GUO received the B.S. degree in economics from Liaoning University in 2010. He is currently pursuing the Ph.D. degree with the Department of Humanities and Economic Management, China University of Geosciences, Beijing.

He majors in time series analysis and complex network application. He has published an article as the first author in a Q1 journal and has participated some energy and economic projects from related research institutes and public sectors.