Network Topologies of Shanghai Stock Index

Jianhua Zhang, Huaxi Zhou, Lu Jiang, Yougui Wang*

Department of Systems Science, School of Management, Beijing Normal University, Beijing, 100875, P. R. China

Abstract

In this paper, we analyzed time series of Shanghai stock index with complex network theory. The degree distribution of the network extracted from the original series can be well fitted with a power law, while the network from return series is governed by an exponential degree distribution. Compared with the time series of standard Brownian motion, we found that the dynamics of the original series can be identified, but the return series has the similar topology with a random one. Moreover, in the scale-free networks from original series, the small-world property is detected and the time interval distribution between connected pairs decays as an exponential function, which implies that nodes correlated with a given one appear in a Poisson process.

Keywords: Financial time series, Complex network, Time interval distribution

1. Introduction

An identification of the dynamic characters underlying a time series is of crucial importance in a wide variety of fields. Lots of methods have been proposed to characterize the dynamic mechanism, such as Lyapunov exponent, entropies and correlation dimension [1]. These traditional methods mainly focus on the overall features of the time series but can not provide more detailed determinism in a system.

During the last decade, complex network theory has stimulated explosive interests in the study of social, informational, technological and biological systems, resulting in a deeper understanding of complex systems [2, 3, 4, 5]. The study of most complex networks has been initiated by a desire to understand various real systems from the empirical data [2]. Local and global properties of complex networks are helpful to understand complex interrelations and information flows among different components, which provides an avenue between the global character and the individual activity in a complex system.

Complex networks display the spatial topological structure of a system, while the time series is the expression of the temporal dynamics. How to construct a network from a time series, i.e.
discussing dynamics from the view of topology, is a very interesting problem. Recently, a bridge between time series and complex networks has emerged [6, 7]. Zhang and Small applied it to reveal information embedded in time series firstly [6, 8, 9]. In their work, they proposed a mapping between pseudoperiodic time series and network. By dividing the series into different disjoint cycles, the nodes and their links comprising a network are formed, the former corresponds to those cycles and the later is defined as their correlations or distances in the phase space. Thus the temporal dynamics of time series are encoded into the topology of the corresponding networks, which can be quantified via the statistical properties of the network. According to these properties, the determinism of different dynamics can be identified. Later, they furthered the study by mapping the time series into a set of points in the phase space and connecting the nearest neighbors to form a network. They showed that the networks from different time series differed in local characters even with similar global statistics through exploring the local structure of these networks [10]. Following their ideas, Marwan et al. applied complex network to analyze time series by identifying the recurrence matrix with the adjacency matrix of a complex network[11]. Meanwhile, Lacasa et al. introduced another algorithm of visibility graph to convert time series to network[7, 12]. In their scheme, each element of a time series is mapped to the node of network, while the connections are based on their visibility associated with their values.

As we know, financial time series is a fundamental topic of continuing interest for its wide application in financial theory and practice. It has more volatility and uncertainty since the activities of the participants in financial market are influenced by lots of factors. There has been much work applying physics concepts and methods to the study of financial markets and financial time series[13, 14, 15, 16]. As one of the most important advances in statistical physics, complex network theory may be a powerful tool for analyzing financial time series. So far, some attempts have been made. Considering a sub-series of fixed length as the counterpart of a cycle, complex networks are applied to explore stock prices[17]. The visibility graph is also used to discuss the characters of exchange rate series[18] and the differences between fractional Brownian motions and multifractal random walks[19]. In addition, another method for network construction from time series is based on its fluctuation patterns[20, 21].

In this paper, we use complex networks to detect determinism from the financial time series. Analogue to the idea of Zhang and Small[6], the time series is divided into successive and nonoverlapping segments mapped to the nodes in networks. While the interaction among nodes are measured by their correlations. Under some constraints, the networks from financial time series can be constructed based on the correlation matrix. Comparing with the networks from the time series of Brownian motion, we discuss the difference and similarity between their statistics. Especially, we explore the real time interval between the nodes connected with each other, which provides a reference on studying long-range correlations since the networks are based on the correlation matrix.

The remainder of this paper is organized as follows: Section 2 presents the transformation rule from financial time series to complex networks; Section 3 displays the networks constructed from financial time series of China, then explores and discusses the basic statistics of these networks; in Section 4 we offer a conclusion.

2. Construction of networks

Given a financial time series \( \{X_1, X_2, \ldots, X_N\} \), e.g. the stock prices, we need to define the nodes and their connections to construct the network. Since the financial time series fluctuates over time, the series could be thought to be pseudoperiodic. As discussed in Ref. [6], the
networks are built as follows. Firstly, the time series is divided into \( m \) segments according to the local minimum (or maximum), denoted as \( \{S_1, S_2, \ldots, S_m\} \). Then every \( k \) adjacent segments are combined to construct a basic node of the network, represented by \( \{N_1, N_2, \ldots, N_p\} \), where \( p = \left[ \frac{m}{k} \right] \). For each pair of combined segments \( N_i \) and \( N_j \) (\( i, j = 1, 2, \ldots, p, i \neq j \)) with length \( l_i \) and \( l_j \) respectively, the correlation coefficient can be defined as below:

\[
\rho_{ij} = \max_{l=0,1,\ldots,l_{1}+l_{2}-1} \frac{\text{Cov}[N_i(1 : l_{1}), N_j(1 + l_{1} : l_{1} + l_{2})]}{\sqrt{\text{V}[N_i(1 : l_{1})]} \sqrt{\text{V}[N_j(1 + l_{1} : l_{1} + l_{2})]}}
\]

where \( N_i(a : b) \) represents the segment between the \( a \)th and \( b \)th elements in \( N_i \) and we suppose \( l_i < l_j \) without loss of generality. \( \text{Cov} \) stands for covariance and \( \text{V} \) stands for variance. This definition takes the difference of each node’s length into account. It calculates all the correlation coefficients moving the shorter segment \( N_i \) onto the longer one \( N_j \) and picks out the largest one as the correlation coefficient between \( N_i \) and \( N_j \). Obviously, \( \rho_{ij} = 1 \) (or \(-1\)) indicates the pair of nodes have perfect correlation (or perfect anti-correlation), while \( \rho_{ij} = 0 \) means there is no correlation between them.

Regarding the links between nodes \( N_i \) and \( N_j \), we use the correlation coefficient \( \rho_{ij} \) to describe their connection by setting a critical value \( v_c \). The transform rule is defined as

\[
W_{ij} = \begin{cases} 
1, & \text{if } |\rho_{ij}| \geq v_c; \\
0, & \text{if } |\rho_{ij}| < v_c,
\end{cases}
\]

where the condition \( W_{ij} = 1 \) means there is a link between nodes \( i \) and \( j \), \( W_{ij} = 0 \) corresponds to no connection between them. Then, the network can be represented by the adjacency matrix \( W \).

As mentioned above, the topology of the networks constructed will be mainly characterized by the combined parameter \( k \) and the filter parameter \( v_c \). The combined parameter \( k \) determines the size of the network, i.e. the magnitude of nodes in the network. Given a time series, the larger the \( k \), the smaller the size, and more elements of the series are contained in one node. Similarly, the value of \( v_c \) determines the smallest strength of the interaction among nodes in the network. If it is extremely small, the pairs with weak correlations are also connected. The physically meaningful correlations in time series will be submerged by these noises. With the increasing of \( v_c \), the connections among the points need larger and larger correlations, which will filter out the noises. Consequently, when \( v_c \) reaches a crucial value, the noises in the interaction among nodes are all filtered out, so the network captures the dynamics underlying the series. However, if the critical value becomes extremely large, some of the physically meaningful connections will be also filtered out and the number of connections will be too small to be significant.

3. Networks from the financial time series in China

The time series used here is the high-frequency data of Shanghai stock index from March 5th to March 16th in the year 2007, which is just two trading weeks. It contains 20705 points. The price series and its corresponding return series are mainly focused on. As defined above, the price series is \( \{X_1, X_2, \ldots, X_N\} \), where \( N = 20705 \). Its return series can be obtained by \( \{(\ln X_2 - \ln X_1), (\ln X_3 - \ln X_2), \ldots, (\ln X_N - \ln X_{N-1})\} \). According to the local minimum, the data can be divided into \( m \) segments \( \{S_1, S_2, \ldots, S_m\} \). The nodes are constructed by combining \( k \) adjacent segments, then the network contains \( \left[ \frac{m}{k} \right] \) nodes, which are connected based on their correlations.
Degree distribution is one of the most important characteristic properties of complex networks [2]. The degree of a node is the number of the edges connecting other nodes, its distribution describes the heterogeneous property and consequently sheds light on the evolutionary mechanism of the network structure. Figure 1 plots the degree distribution of the network extracted from the time series of Shanghai stock index under different thresholds \( v_c \) and combined parameters \( k \). As the figure shows, the structure varies with \( v_c \) and \( k \) drastically. Setting \( k = 3 \), with the increasing of \( v_c \), the degree distribution varies from a likely Gamma distribution to a power law. When \( v_c \) is small, the distribution are dominated by the noises in the interaction among nodes. With a large threshold, the noises are filtered out, the structure of the network is determined by the strong correlations, which may be the deterministic pattern in the time series. Regarding the parameter \( k \), it describes the length of the data in nodes. When the value of \( k \) changes from 3 to 5, the number of the nodes in the constructed networks decreases from 2111 to 1266. The magnitude will be too small for statistical analysis when \( k \) becomes even larger. In figure 1d – f, the dynamics of networks are displayed with increasing the value of \( k \). The power-law degree distribution is robust for different value of \( k \). The exponents are 1.65458, 1.47238 and 1.30389 corresponding to \( k = 3, 4, 5 \) respectively. The exponent is decreasing with the scale of networks, i.e. it has an inverse relation with the length of nodes. However, we can still argue that the dynamics of this time series is encoded into a scale-free network given a big enough threshold. The power-law degree distribution implies the existence of some hubs in the constructed networks. The segments corresponding to the hubs correlate with most of others and contain more information about the time series.

![Figure 1: Evolution of the network constructed from the time series of the stock index with the threshold \( v_c \) and combined parameter \( k \). Graphs c-f are in log-log coordinates and the red lines are fitting curves.](image)

Similarly, we apply this analysis to the return series of Shanghai stock index. In figure 2, we present a typical degree distribution for return series. As it shows, the distribution is approximated well by a straight line in semi-logarithmic coordinates, which indicates that the degree follows an exponential distribution and the estimate of its exponent is 0.04245. This result is not consistent with the discovery of Yang et al.[17]. They found that the degree distribution could be fitted with a Gaussian function for the threshold \( r_c \in [0.50, 0.70] \). Here, we give the
degree distribution under the condition of a larger filter value and argue that the network of return series is governed by an exponential degree distribution after filtering out influences of noises.

To check the validity of complex networks in detecting the dynamics from time series, we study the structure of networks extracted from the time series of Brownian motion. As shown in figure 3, both of the distributions are fitted by a straight line in the semi-logarithmic coordinates. This phenomenon implies that both of them obey the exponential distribution. The estimates of exponents for subgraph $a$ and $b$ are 0.08202 and 0.01271 respectively. Although one may expects a Poisson degree distribution for the networks in a first moment, there indeed exists some correlation in a random series, because the correlation is calculated from the data. Some large correlation may appear between the random data. This deviation may cause the degree distributions to be exponential instead of Poisson. Compared with the result in the case of Shanghai stock index, the features of networks from the original series are different with each other remarkably, but the distributions for return series have a similar shape, the difference is just the value of the exponents. It is not enough to distinguish them from each other. The reason for this may be the operation of converting the original series into the return series. The network of price return series has the same degree distribution with that from a random one, however, it can not be inferred that they have the same topological structure simply. The degree distribution just shows the global characteristics of networks. We note that many networks with the same global properties may have wildly different local structures [22]. The difference between the networks constructed from price return series and a random one needs to be explored in detail.

Besides the degree distribution, another two statistics on topology of networks are (I) the average clustering coefficient $C$: the ratio of the number of edges that actually exist between a vertex and its topological neighbors to the total number and (II) the average shortest distance $\langle l \rangle$: the average of the minimum number of links necessary to connect a pair of nodes in the network.
Take the network presented in figure 1 as an example, it is a scale-free network due to its power-law degree distribution. The largest connected subgraph of the networks contains 1613 nodes. We find that \( C \) of the subgraph is 0.2433 and the value of \( \langle l \rangle \) is 4.801. Compared with stochastic networks of the same size, the value of mean path length is smaller than stochastic results and the mean clustering coefficient is larger than the stochastic values, showing that the networks have small world property. Setting the time of each node as the middle time of the corresponding segment, the duration of the strong correlation for the time series could be described by the time interval among the nodes connected. As displayed in figure 4, the time interval obeys exponential distribution. This phenomenon implies that the segments correlated with a given one appear in a Poisson process.
4. Conclusions

In the paper, we applied the complex networks to analyze the financial time series. We transformed the temporal dynamics of the series to the topological structure of a corresponding network. This network-based analysis attracted special attention. With the high-frequency series of Shanghai stock index and its return, we built several networks under different values of parameters $k$ and $v_c$, which determined the length of segments converted to nodes and the connections in networks. By studying the basic statistical properties of the networks, we found some useful and interesting phenomena. The network constructed from index series exhibits scale free and small world features after filtering out the influences of noises. Whereas, the degree of the network from return series follows an exponential distribution. Subsequently, we did the same operation for the time series of Brownian motion, the degree distributions of the networks transformed from the original and return series both have the form of exponential function, which is consistent with the case of the real return. It may be caused by the operation of converting the original series into the return series. The degree statistics does not detect the local character such as volatility clustering in the real return series. It needs further study on local structure of the networks.

The nodes are connected according to the correlation between them. The topological structure of the corresponding network can provide us with more detailed information on the correlation matrix. The scale-free degree distribution of the stock index series implies the existence of some hubs in the constructed networks. These segments corresponding to hubs contain more information and play a key role in the dynamics of the time series. The time interval between the correlated segments in the scale-free networks follows a distribution of exponential function, which means that the segments correlated with a given one strongly appear in a Poisson process. The complex network provides some useful global information on the dynamics underlying the time series, however, there still has a large gap from the accurate identification of the dynamics behind the data.

Acknowledgements

This research is partly funded by the National Natural Science Foundation of China under Grants of Nos. 70771012, 70771011 and 60974084.

References

[1] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, L. S. Tsimring, The analysis of observed chaotic data in physical systems, Reviews of Modern Physics 65 (1993) 1331–1392.
[2] R. Albert, A. L. Barabási, Statistical mechanics of complex networks, Reviews of modern physics 74 (2002) 47–97.
[3] M. E. J. Newman, The structure and function of complex networks, SIAM Review (2) (2003) 167–256.
[4] S. Dorogovtsev, J. F. F. Mendes, Evolution of networks: From biological nets to the Internet and WWW, Oxford University Press, USA, 2003.
[5] S. Boccaletti, D. U. Hwang, M. Chavez, A. Amann, J. Kurths, L. M. Pecora, Synchronization in dynamical networks: Evolution along commutative graphs, Physical Review E 74 (2006) 16102.
[6] J. Zhang, M. Small, Complex network from pseudoperiodic time series: Topology versus dynamics, Physical Review Letters 96 (2006) 238701.
[7] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J. Nuño, From time series to complex networks: The visibility graph, Proceedings of the National Academy of Sciences 105 (2008) 4972.
[8] J. Zhang, X. Luo, M. Small, Detecting chaos in pseudoperiodic time series without embedding, Physical Review E 73 (2006) 16216.
[9] J. Zhang, J. Sun, X. Luo, K. Zhang, T. Nakamura, M. Small, Characterizing pseudoperiodic time series through the complex network approach, Physica D 237 (2008) 2856–2865.
[10] X. Xia, J. Zhang, M. Small, Superfamily phenomena and motifs of networks induced from time series, Proceedings of the National Academy of Sciences 105 (2008) 19601.
[11] N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, J. Kurths, Complex network approach for recurrence analysis of time series, Physics Letters A 373 (2009) 4246–4254.
[12] L. Lacasa, B. Luque, J. Luque, J. C. Nuno, The visibility graph: A new method for estimating the hurst exponent of fractional brownian motion, Europhysics Letters 86 (2009) 30001.
[13] T. Lux, M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, Nature 397 (1999) 498–499.
[14] H. E. Stanley, S. V. Buldyrev, The salesman and the tourist, Nature 413 (2001) 373.
[15] V. Plerou, P. Gopikrishnan, H. E. Stanley, Two-phase behaviour of financial markets, Nature 421 (2003) 130.
[16] X. Gabaix, P. Gopikrishnan, V. Plerou, H. E. Stanley, A theory of power-law distributions in financial market fluctuations, Nature 423 (2003) 267–270.
[17] Y. Yang, H. Yang, Complex network-based time series analysis, Physica A 387 (2008) 1381–1386.
[18] Y. Yang, J. Wang, H. Yang, J. Mang, Visibility graph approach to exchange rate series, Physica A 388 (2009) 4431–4437.
[19] X. H. Ni, Z. Jiang, W. X. Zhou, Degree distributions of the visibility graphs mapped from fractional brownian motions and multifractal random walks, Physics Letters A 373 (2009) 3822–3826.
[20] P. Li, B. Wang, An approach to hang seng index in hong kong stock market based on network topological statistics, Chinese Science Bulletin 51 (2006) 624–629.
[21] P. Li, B. H. Wang, Extracting hidden fluctuation patterns of hang seng stock index from network topologies, Physica A 378 (2007) 519–526.
[22] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (2002) 824–827.