Novel Method of Identifying Time Series Based on Network Graphs

YING LI,[1] HONGDUO CAÖ,[1] AND YONG TAN[2]

[1]*Department of Management Science, Business School, Sun Yat-Sen University, Guangzhou 510275, China; and* [2]*Michael G. Foster of Business, University of Washington, Seattle, Washington 98195-3200*

Received January 4, 2011; revised June 7, 2011; accepted June 9, 2011

In this article, we propose a novel method for transforming a time series into a complex network graph. The proposed algorithm is based on the spatial distribution of a time series. The characteristics of geometric parameters of a network represent the dynamic characteristics of a time series. Our algorithm transforms, respectively, a constant series into a fully connected graph, periodic time series into a regular graph, linear divergent time series into a tree, and chaotic time series into an approximately power law distribution network graph. We find that when the dimension of reconstructed phase space increases, the corresponding graph for a random time series quickly turns into a completely unconnected graph, while that for a chaotic time series maintains a certain level of connectivity. The characteristics of the generated network, including the total edges, the degree distribution, and the clustering coefficient, reflect the characteristics of the time series, including diverging speed, level of certainty, and level of randomness. This observation allows a chaotic time series to be easily identified from a random time series. The method may be useful for analysis of complex nonlinear systems such as chaos and random systems, by perceiving the differences in the outcomes of the systems—the time series—in the identification of the systemic levels of certainty or randomness. © *2011 Wiley Periodicals, Inc. Complexity 17: 13–34, 2011*

Key Words: time series; complex network; chaos; random; diverging speed

Correspondence to: Ying Li, Department of Management Science, Business School, Sun Yat-Sen University, Guangzhou 510275, China (e-mail: mnsliy@mail.sysu.edu.cn) or Hongduo Cao, Department of Management Science, Business School, Sun Yat-Sen University, Guangzhou 510275, China (e-mail: caohd@mail.sysu.edu.cn)

1. INTRODUCTION

Models based on complex networks have been applied to describe many real-world systems. The Internet, protein networks, firm networks, finance networks, and scientific collaboration networks are all examples of models based on complex networks that have been applied to describe real-world systems [1–3]. As well, the interdisciplinary nature of complex networks has brought about important developments in a wide range of fields, including mathematics, physics, computer and in-

DOI 10.1002/cplx.20384
Published online 30 August 2011 in Wiley Online Library (wileyonlinelibrary.com)

formation sciences, biology, and the social sciences [4]. The importance and the range of networks have prompted innovative scholarly research. However, the study of networks mainly focuses on, but is not limited to, the structure of complex networks and dynamics. Previous literature, including the representative work of Barabási [1] and Boccaletti [5], provides a comprehensive overview of the characteristics and systemic applications of complex networks—and the graphical method of analyzing the networks. In this way, a complex network can effectively outline the structure of the systems—the nodes representing the subjects of a system (which can be person, computer, or gene), and the edges modeling the relationship among different subjects. Complex networks have also been used to depict the characteristics of some abstract systems, such as sequences of symbols, and time series. Sinatra [6] introduced a method to transform a generic corpus of strings (sequences of symbols) into a weighted directed network in which nodes are motifs, while the directed links are based on statistically significant co-occurrences of two motifs. For time series, as well, complex networks can be a modeling tool. Time series, as the output of a system, can reflect the inherent characteristics of a system. Although a complete mathematical analytic model is difficult to directly establish for a complex nonlinear dynamic system, analysis of time series is usually a research approach for understanding the original system. If a time series is mapped to a network, the structural characteristics of the generated network graph can reflect the kinetics mechanism of the time series.

This article provides a novel method to describe, express, and distinguish a time series by complex network graph. This method differs from those previously developed, in that we focus on the spatial distribution properties of time series and map the spatial characteristics to network graphs. In addition, by multireconstructing phase space, the corresponding change pattern of generated network graphs can be used as a criterion to distinguish various time series (especially chaotic time series and random time series).

2. RELATED REFERENCES

Fourier analysis, as the theoretical basis for period analysis, gives the mathematical method for period analysis of time series. The classical methods for time series analysis are primarily based on statistical theory. Among theories, linear time series analysis is relatively mature, including autoregressive (AR; [7]), moving average (MA; [8]), autoregressive moving average (ARMA; [9]), and autoregressive integrated moving-average (ARIMA) models [10]. Nonlinear models include the Markov switching regime (MSR) model [11], the Threshold autoregression (TAR) model [12] and the Smooth transition autoregression (STAR) model [13], and the TV-STAR model [14]. Time series analysis based on nonlinear dynamics was started in the early 1980s. In essence, this method focusing on nonlinear time series, and based on dynamics, takes advantage of geometric properties and kinetic properties of phase space orbit to identify the nature of the time series (chaotic or random). The method based on nonlinear dynamics generally works by calculating dynamics invariants such as attractor dimension D (G-P method) [15], measure entropy K [16], and Lyapunov exponents. Schreiber [17] reports on the method development and application to measure time series on the basis of the theory of deterministic chaos.

Many time series analysis applications based on dynamics theory make use of phase space reconstruction, so that the discrete time series is an equal interval sampling of the outputs of a system:

$$x(t_1), x(t_2), \ldots, x(t_n).$$

The above one-dimension time series $x(t_i)$ hides the essential characteristics of the complex structure of dynamic systems. To obtain the geometric structure of phase space of the dynamic system from the time series, Packard et al. [18] use the time delay to reconstruct phase space, by embedding the one-dimensional time series into the m-dimensional space:

$$X(t) = [x(t), x(t-\tau), x(t-2\tau), \cdots, x(t-(m-1)\tau)],$$

Here, $X(t)$ denotes the dynamic state of the system at time t. The parameter τ is the delay time, and m denotes the embedding dimension of the space.

As an example, for a discrete time series x_1, x_2, …, x_n, the point in embedded phase space with $\tau = 1$ and $m = 3$ is as follows:

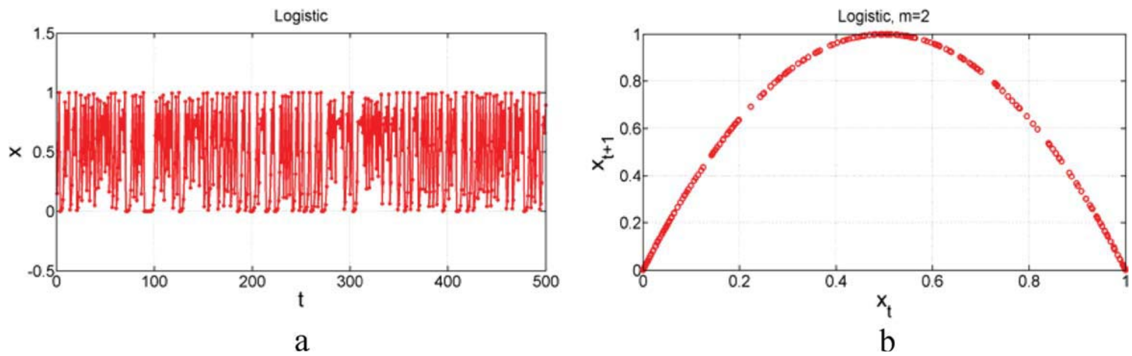
$$\begin{matrix} x_3 & x_4 & & x_n \\ x_2 & x_3 & \ldots & x_{n-1} \\ x_1 & x_2 & & x_{n-2} \end{matrix}.$$

Accordingly a mapping $\parallel$ from original dynamic system to embedded space R^m is established. Takens [19] proved an embedding theorem for a time series, demonstrating that $\parallel$ is a smooth one-to-one mapping while $m > 2D + 1$, and the geometric properties of attractors in embedded space is equivalent to that in an original dynamic system, where D is the fractal dimension of the attractor. The track with some regular patterns (attractors or strange attractors) will be restored in embedded space. In other words, in the reconstruction phase space, the phase portrait of reconstruction space R^m maintains diffeomorphism with that of the underlying system.

For example, a one-dimensional time series such as a logistic chaotic series is chaotic, but shows a parabolic pat-

DOI 10.1002/cplx

FIGURE 1



Reconstructing logistic time series in two-dimensional embedded phase space. (a) Logistic time series and (b) logistic mapping in embedded space with $m = 2$.

tern in two-dimensional embedded space, as shown in Figure 1.

Figure 1 demonstrates the method of phase space reconstruction that can display a regular pattern of the chaotic series in a two-dimensional embedded phase space. Therefore, the method of phase space reconstruction is an efficient method to observe the characteristics of complex time series. Hence, this article is based on the method of phase space reconstruction.

The time series can reflect dynamic characteristics of a system, but generally the dynamic characteristics of the time series are very complex. As an example, it is hard to distinguish chaotic time series from random time series, because the chaotic time series has a definitive internal mechanism (generated by a deterministic mapping), but with a random appearance. To address this issue, new methods and tools have recently begun to develop. Mapping time series to a network is one of the methods for researching chaotic and random time series.

Zhang and Small [20] have constructed complex networks for pseudoperiodic time series by associating each series cycle with a node and defining links between nodes by temporal correlation measures. Lacasa et al. [21] have proposed connecting one node with others that can be seen from the top of graph, according to temporal order. This method transforms periodic time series into regular graphs, random time series into random graphs, and fractal time series into scale-free networks. One application of the method is to estimate the Hurst exponent of fractional Brownian motion [22]. Further, Luque et al. [23] suggest another horizontal visibility graph method to solve the theoretical analysis of their method, and Hirata et al. [24] have developed an approach for reproducing distance matrices and original time series from recurrence plots. This procedure converts a time series in a weighted graph using RPs (Recurrence Plots), and then reproduces the original time series from the graph. Xu et al. [25] have embedded the time series in an appropriate state space with a given time delay and embedding dimension, and then have selected a fixed number of nearest neighbors to each point to connect the nodes. The main difference between our method and those detailed here is the connecting definition. Our connecting method is based on the distance between different points in the phase space. The link definition in the above references includes such approaches as correlation coefficient [20], visibility graphs [21], and the nearest neighbors method [25]. Primary among these methods is correlation coefficient. We provide a detailed comparison of the methods that are based on space distance, and that are based on correlation coefficient in principle, effect, and application scope [26]. Different connecting methods will result in different graphs. For example, a linear series will generate a fully connected graph when based on correlation coefficient, but generates a tree[1] when based on distance. The comparison shows the method based on distance has obvious advantages in distinguishing chaotic series from random series, compared with that based on correlation coefficient [25].

3. METHOD TO CONSTRUCT THE NETWORK GRAPH

3.1. Definition of Method

For the development of our concepts, we study equidistant discrete time series where the time interval is normalized

[1] *In mathematics (more specifically, in graph theory), a tree is an undirected graph in which any two vertices are connected by exactly one simple path. In other words, any connected graph without cycles is a tree [27].*

DOI 10.1002/cplx

to 1, explicitly, $x = \{x_1, x_2, \ldots, x_t, \ldots, x_n; t = 1, 2, \ldots, n\}$. To construct a corresponding network graph, we first define the distance between data points, and then construct the graph according to an algorithm based on the distance. This algorithm reflects the spatial distribution property of a time series. That is, we directly relate the spatial clustering property of a time series to that of a corresponding network graph. A data point with more neighbors will be more probably to be connected.

The algorithm of constructing a network graph includes three parts: the definition of the node of a network graph, the definition of the distance between any two data points, and the rule of connecting. The detailed descriptions appear below.

3.1.1. Definition of Node

A node is defined as a point in an m-dimensional reconstructed phase space. For a time series, $x_1, x_2, \ldots, x_t, \ldots, x_n$; $t = 1, 2, \ldots, n$, in a reconstructed phase space

$$X(t) = [x(t), x(t-\tau), x(t-2\tau), \cdots, x(t-(m-1)\tau)],$$

where m denotes the dimension of the embedding space, and τ denotes the time delay. If $\tau = 1$, then the total number of nodes k is $n - m + 1$. (In this article, a node in network corresponds to a vector in the reconstructed phase space.)

3.1.2. Definition of Distance

The Euclidean distance between two nodes i and j, $X_i = (x_{i_1}, x_{i_2}, x_{i_3}, \ldots, x_{i_m})$ and $X_j = (x_{j_1}, x_{j_2}, x_{j_3}, \ldots, x_{j_m})$, is defined as

$$d_{ij} = \sqrt{|x_{i_1} - x_{j_1}|^2 + |x_{i_2} - x_{j_2}|^2 + |x_{i_3} - x_{j_3}|^2 + \cdots |x_{i_m} - x_{j_m}|^2}.$$

For the special case when $m = 1$, the distance is simply $d_{ij} = |x_i - x_j|$.

3.1.3. Rule of Connection

We define the connecting rule as follows. Let d_{max} denotes the maximum distance. $\Delta = d_{max}/(k - 1)$ is called the judgment distance, where k is the total number of nodes. Two nodes, i and j, will be connected only if $d_{ij} \leq \Delta$.

3.2. Discussion of the Method

This method is based on the reconstructed phase space method of constructing a network. The connection between the different nodes depends on the judgment distance Δ. Therefore, Δ is a key point of this method.

In the process of constructing a network, all distances between any two nodes in the phase space are first calculated, and then are compared with the judgment distance Δ. The distance d_{ij} is a microlevel measure of distance between any two nodes, but the maximum distance of d_{max} is a relative macrolevel value because it is the maximum of the entire distance set. While in the formula of Δ, Δ is defined by d_{max}. Therefore, Δ is also a parameter reflecting the state of the system at macrolevel. The essence of the comparison between d_{ij} and Δ is comparing the microvalue of distance space from a time series to the macrovalue of distance space. The relation between micro- and macrolevels can reflect the entire feature of a time series.

The judgment distance Δ can reflect the division to the distance space, and then can also allow our algorithm to identify the divergent speed of time series.

First, we will analyze the relation between Δ and the size n of series. In the formula $\Delta = d_{max}/(k - 1)$, where $k = n - m + 1$, the series size of n appears in the denominator, and d_{max} appears in the numerator. However, d_{max} is also related to n. When n increases, the data samples increases. The space of the set of all distance $\{d_{ij} | i = 1, 2, \ldots, k; j = 1, 2, \ldots, k; i \neq j\}$ will increase. There will then be an opportunity for d_{max} to increase. The change direction of Δ with n increasing depends on whether it is d_{max} or n that changes faster.

The growth rate of d_{max} with n increasing, reflects the diverging speed of time series. If the time series oscillates in a range or converges, as shown in Figure 2, the value of d_{max} cannot be larger than $[\mathrm{Lim}(x_{max}) – \mathrm{Lim}(x_{min})]$. Hence, even if d_{max} changes, the increasing range is limited. That is, d_{max} grows at a slower pace than n. But for divergent sequences, d_{max} will increase with n increasing. The faster the time series diverges, the faster the d_{max} increases with n increasing. We analyze the meaning Δ for a time series with $m = 1$.

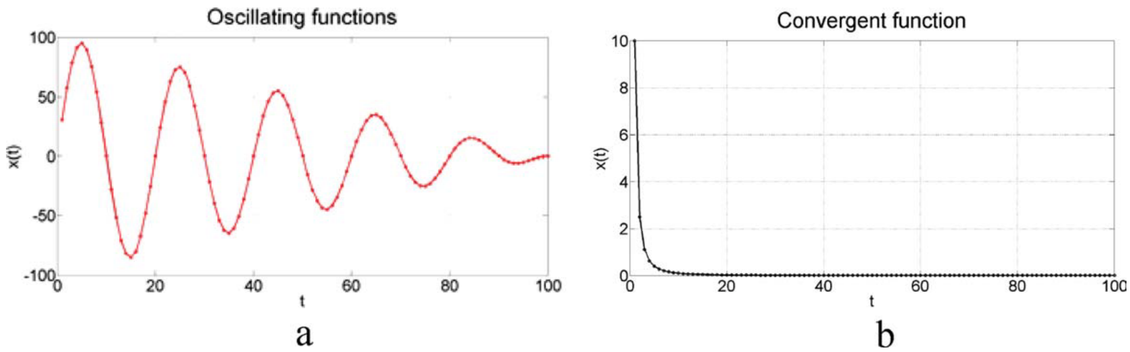
For our study, we selected a random series and a linear series for comparison. First, we generated a random sequence, ranking it from small to large. Then, we construct a linear sequence that starts from the minimum of above random series, and ends in the maximum. Figure 3 shows the original random sequence denoted by A, and the linear sequence B for which the starting point is at the minimum value, and the end point is at the maximum value of the series A. Figure 3(b) shows the ranked A, represented by A′, and sequence B.

In Figure 3(a), we first generated a random series, then generated a linear series, for which the starting point is the minimum point of the random series, and the end point is the maximum point of the random one. In Figure 3(b), the random series is ranked to compare with the linear series.

Next (as shown in Figure 4), we calculate the distance between adjacent points, and rank it from small to great. Figure 4(a) plots the distances between adjacent points from the sequences A′ and B. The distances between adjacent

DOI 10.1002/cplx

FIGURE 2



(a) Oscillating series and (b) covergent series.

points of a linear sequence B, denoted by d_0, all are equal. The distance between adjacent points from the sequence A′ is denoted by d'_i. Figure 4(b) plots the ranked d'_i.

As shown in Figure 4, compared with d_0, some d'_i are less than d_0, and others are greater than d_0. The least distance between one point and all other points exists among the distances between the two adjacent points. That is, after ranking, the possible minimum distance from a point x_i to any other point is $d_{i-1,i}$ or $d_{i,i+1}$. The relation can be expressed as $d_{i-1,i}$ or $d_{i,i+1} = \min(\{d_{ji}|i,j = 1, 2, \ldots, n; i \neq j\})$. The reason is as follows:

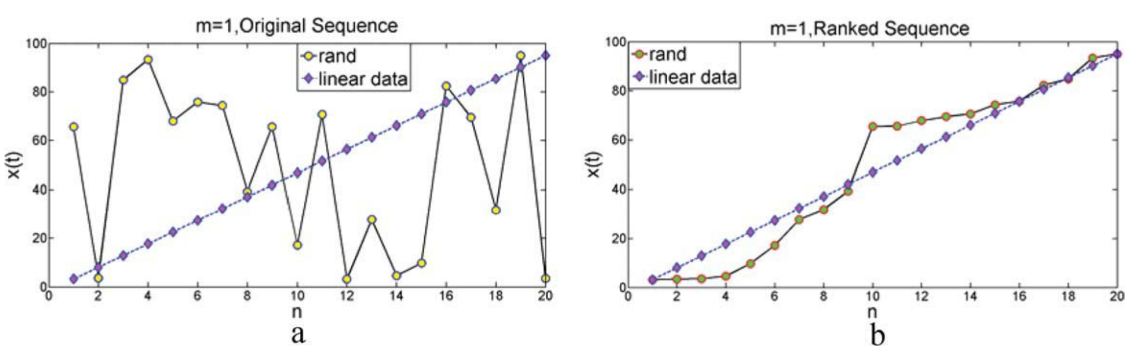
$$\text{If } i < j, d_{j,i} = d_{i,i+1} + d_{i+1,i+2} + \cdots d_{j-1,j}, \quad i = 1, 2, \ldots, m; i \neq j.$$
$$\text{If } i > j, d_{j,i} = d_{j,j+1} + d_{j+1,j+2} + \cdots d_{i-1,i}, \quad i = 1, 2, \ldots, m; i \neq j.$$

That is, $d_{i,j} \in D$, D is the space spanned by $\{d_{i-1,i}|i = 2, 3, \ldots, n\}$. Hence, $d_{i-1,i}$ or $d_{i,i+1} = \min(\{d_{ji}|i,j = 1, 2, \ldots, n; i \neq j\})$.

For the above linear series and random series, the $d_{\max}$ are equal, while for the linear time series, $d_0 = d_{\max}/(n - 1)$, where the parameter n is the size of series. Therefore, the d_0 of the linear time series is just the Δ of random series A and linear series B. In other words, $d_0 = \Delta_B = \Delta_{A'} = \Delta_A$. According to our algorithm, the two nodes of the generated network, as well as the data point for $m = 1$, will connect to each other when the distance between this point and other points is less than d_0, but will not connect to each other when the distance is greater than d_0.

Therefore, for any sequence, when ranked from small to large, the two adjacent nodes will connect if the adjacent distance $d_{i-1,i}$ is less than d_0 equaling to a linear partition of space between the maximum and the minimum of the sequence. This state means that the divergent speed of the sequence is slower than the linear divergent speed. Otherwise, the two adjacent nodes will not connect, when $d_{i-1,i}$ is greater than d_0, which means that the divergent speed of the sequence is faster than the linear divergent

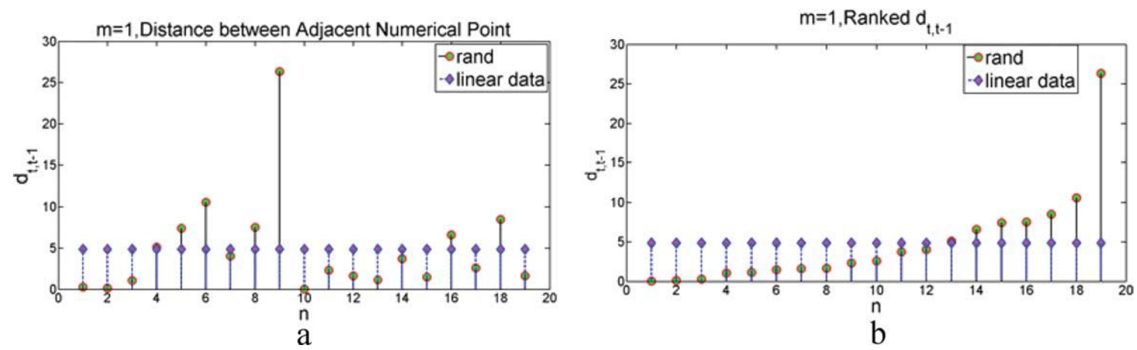
FIGURE 3



Random series and linear series. (a) Original random sequence a linear sequence and (b) ranked sequence a linear sequence.

DOI 10.1002/cplx

FIGURE 4



Distance between two adjacent points. (a) Original distance and (b) ranked distance.

speed. For any node, if the least distance is larger than d_0, other distances must be larger than d_0. Of course, no connection exists then.

Therefore, the method in this article suggests a linear division of the space between the maximum and the minimum of any series. That is to say, this algorithm makes a linear partition in the phase space. If a series is linear divergent, the smallest connectivity network—a tree (see endnote on page xx)—will appear. If a system diverges greater than a linear divergence, some unconnected sections will emerge. For a time series with a constant value, both d_{ij} and Δ are zero; hence, a fully connected network graph will result. This approach indicates that we can use different network graphs to represent different time series.

4. RESULTS OF EXPERIMENTAL DATA

The properties of a time series are varied, and include such attributes as convergent or divergent, periodic or nonperiodic, regular or irregular, and chaotic or purely random and so on. In this article, the method mapping a time series into a network is primarily useful for identifying the divergent property (divergent speed), periodicity, and randomness of a time series. Accordingly, we selected for our study the constant series, the periodic series of sin function, the linear divergent series, the chaotic series (logistic series and Henon series), and the pure random series. The constant series with a fixed value is selected to observe the effect caused by change of value, by comparing the network generated by a time series with value changing, to that generated by a constant series. The periodic series of sin function is selected to observe the effect caused by periodicity, by comparing the network generated by a nonperiodic time series to that generated by a periodic series. A linear divergent series is selected to enable observation of the effect caused by divergent speed (faster than linear divergent time series), by comparing the network generated by a linear time series to that generated by a nonlinear series. Two chaotic series, as well as the logistic series and Henon series, are selected to observe the difference in networks between chaotic series and random series. Because chaotic series are produced by a deterministic mapping, chaotic series is a time series that has internal certainty but external randomness. It is difficult to distinguish a chaotic series from a random series, but it can be solved by generating a network. A random series is selected so that we can observe the results caused by complete randomness, and the difference between the complete random and the random including some certainty. We can, in this way, distinguish pure random series and chaotic series, by comparing the significantly different topology structure between pure random series and chaotic series.

As an additional factor, there is no relationship between the properties of a series and the size of that series. For our analysis, we research time series of length $n = 1000$. If the size of the data is too large, a higher performance of computing device is required, because in the process of generating network, the adjacency matrix of a network will be calculated, which will occupy a large computer memory. Therefore, when the series data is very large, we can apply a sliding window method to deal with the data. We can then judge the features of the time series by observing the characteristics within every window.

We apply our generating network method to the constant series, periodic series, linear divergent series, logistic

DOI 10.1002/cplx

series, Henon series, and random series. The six time series are as follows:

constant series: $x(t) = 1$, $t = 1, 2, \ldots, n$;
periodic series: $x(t) = \sin(a{\cdot}\pi{\cdot}t)$, $t = 1, 2, \ldots, n$; in this article $a = 0.25$;
linear divergent series: $x(t) = \mu{\cdot}t$, $t = 1, 2, \ldots, n$; in this article, $\mu = 1$;
logistic series: $x(t) = \mu{\cdot}x(t-1)(1 - x(t-1))$, $t = 1, 2, \ldots, n$; $0 < \mu \leq 4$; in this article, $\mu = 4$, $x(1) = 0.55$;
Henon series:

$$\begin{cases} x(t) = 1 - a \cdot x(t-1)^2 + y(t) \\ y(t) = b \cdot x(t-1) \end{cases}; a = 1.4 \text{ and } b = 0.3;$$

random series: 0–1 random number with uniformly distribution.

These series with 1000 data points are drawn in Figure 5. The corresponding networks from these series (with n = 25 and 1000) are shown in Table 1. The numerical experiment is for series with $n = 1000$. But when $n = 1000$, the total edges of the generated network from the constant series and periodic series is very large. The network graph that results is a large black round shape, and the internal structure can be clearly observed. Also, the linear series will generate a black circle when $n = 1000$. In fact, the topology structure of network generated from the constant series, periodic series, and linear series is the same for different n (which we will explain in the latter content of our article). To show the internal structure more clearly, however, when $m = 2$ and 3, we use 25 data points for the constant series, periodic series, and linear series. It can be observed that the structures of these three series do not change significantly (Figure 5).

The graphs in Table 1 show the network graphs at different embedded dimension m. The number of edges and nodes are indicated at the bottom of each figure. All nodes of the network are placed at a ring. When the number of nodes is large, the nodes will overlap each other. Also, when the number of edges is large, the edges will overlap. Hence, when both the number of nodes and number of edges are large, the result will be a graph that is unclear. To enable readers to see, visually, the difference between the generated network and the different time series, we first plot the case of small data—that is, $n = 25$, as shown in the first line of Table 1. In this line, we can clearly see the fully connected graph, regular graph, tree, and the irregular graphs generated by the other three corresponding series. But the actual study is for relatively large-scale data of $n = 1000$, so we next plot the generated networks at $m = 1$ and $n = 1000$ on the second line. For this case, the networks generated by a constant series, the periodic series becomes two black round shapes because there are too many links. The network generated by a linear series becomes a black circle because there are too many nodes. The internal structure is completely blind. To enable the reader to see clearly, we provide the sketch map at n = 25—but not $n = 1000$—for the constant series, periodic series, and linear series in the case of $m = 2$ and $m = 3$.

The graphs in Table 2 show the distribution of degree of the six networks, according to different values of m (m = 1, 2, 3). Degree refers to the number of edges of one node. The first column shows the degree distribution at linear scale, and the second column shows the degree distribution at logarithm scale. The abscissa represents degree, and the ordinate represents the probability density of degree $p(d)$. For every value of m, the degree distributions of networks from the six time series are plotted, with the name of the series noted at the top of each figure. Also, the maximum value and the minimum of degree are marked in the degree distribution graph at linear scale.

The figures in Table 3 show the clustering coefficient distribution of the generated network at different embedded dimension m. The average clustering coefficient of each network is indicated at the bottom of each figure. The names of all six time series are noted at the top of every figure.

To compare the clustering coefficient of six series, we plot all average clustering coefficients of the six series in Table 4.

Figure 6 shows the total edges of the six series at n = 1000 for different embedded dimension m. The horizontal axis is the time series. The first series is the constant series; the second is the sin series; the third is the linear series; the fourth is the logistic series; the fifth is the Henon series; and the sixth is the random series. The vertical axis is the total edges. Because the value, the last four series are unclear, so for the last series we plot the total edges in a small graph.

5. SUMMARY AND DISCUSSION

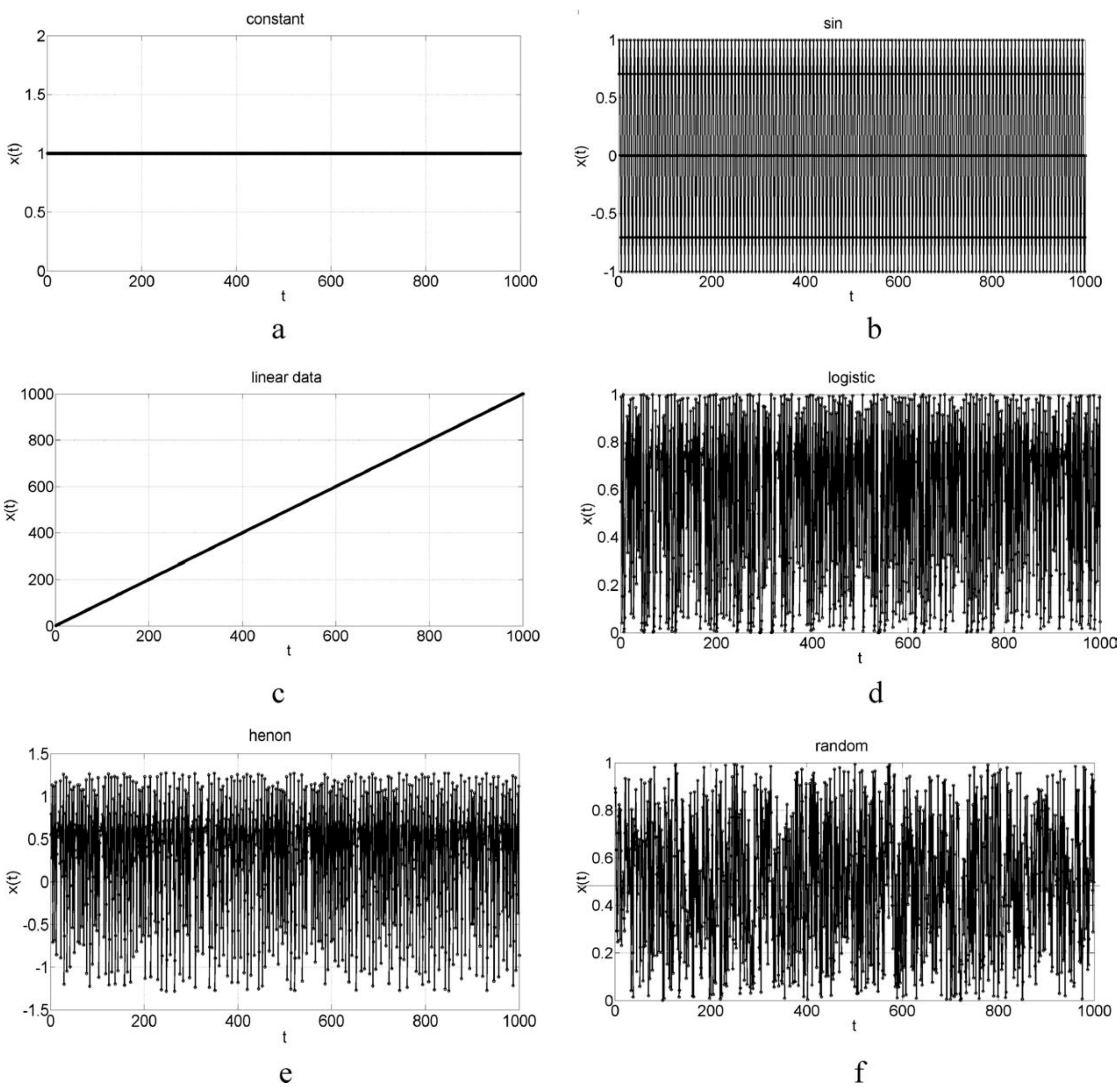
5.1. Summary of Observation

From Tables 1–4, and from Figure 6, we make the following observations.

1. The constant series can be expressed as a network where each point is connected with all other points—that is, a fully connected graph will emerge. When m increases, the network remains fully connected and its characteristics are unchanged.
2. The periodic series (sin) results in a regular network. The degree of each node is large. In other words, the sum of edges of the network is large. The degree distribution is two concentrated degrees. When m increases, the topological characteristics of the network graph

DOI 10.1002/cplx

FIGURE 5



Six time series. The t denotes time, and x(t) denotes the value at t. (a) Constant series; (b) sin function; (c) linear series; (d) logistic series; (e) Henon series; and (f) random series.

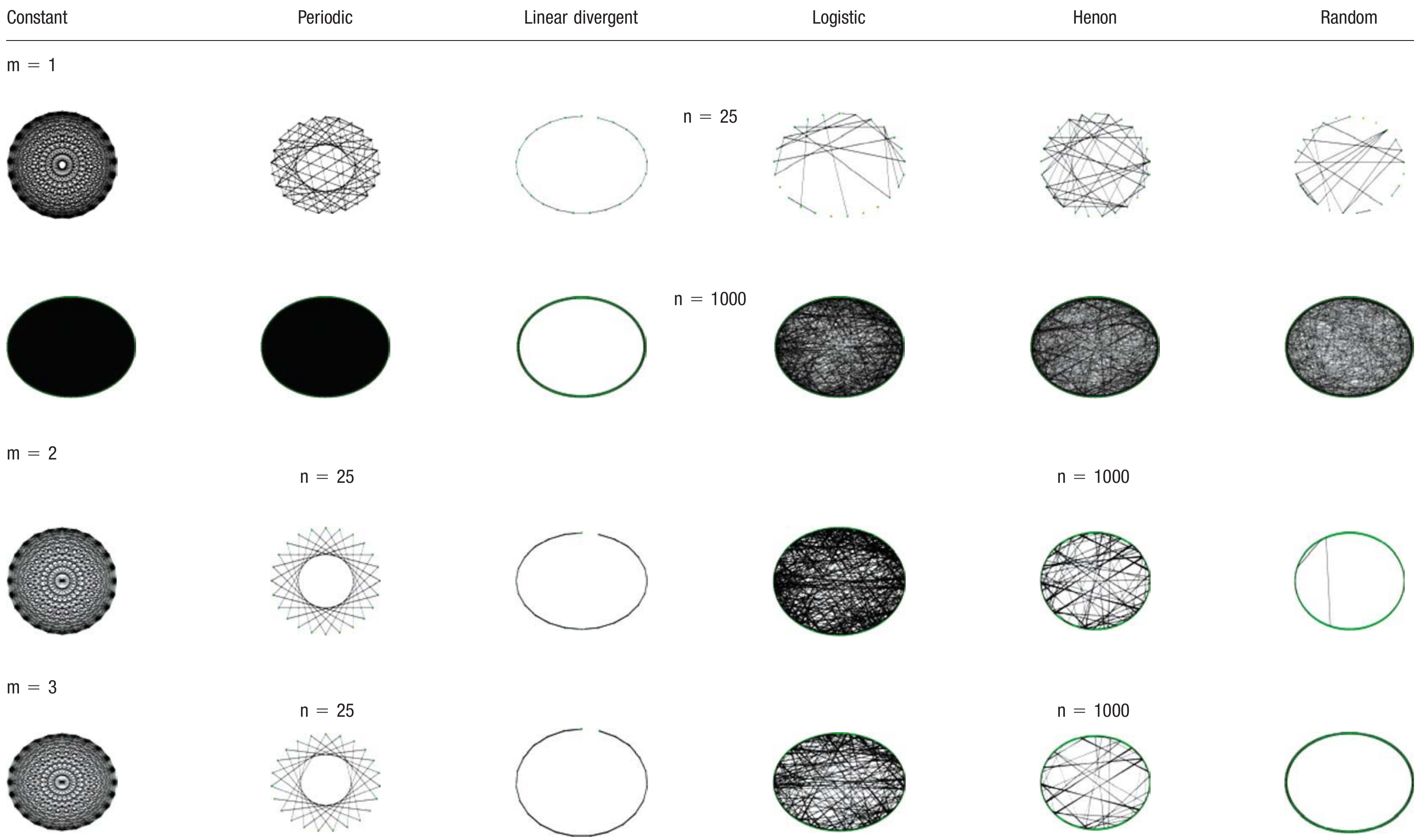
remain unchanged. It is still a regular network, concentrated degrees, and high edges. But the concentrated degrees may change in value, for example, when m increases from 1 to 2.

3. The linear divergent time series is transformed into a minimum connected network graph, namely a tree. The degrees of the starting and end nodes are one; the other nodes have a degree of 2. When m increases, the topological characteristics of the network graph remain unchanged—it is still a tree. But compared with the periodic series, the average degree is small and the total number of edges is very small (for example, for $m = 1$, the two concentration degrees of sin series are 124 and 249, whereas those of the linear divergent series are 1 and 2).
4. The greatest tolerated divergent level is linear divergent. When the global or local divergent motive force is greater than a linear divergent level, some nonconnected parts will appear.

DOI 10.1002/cplx

TABLE 1

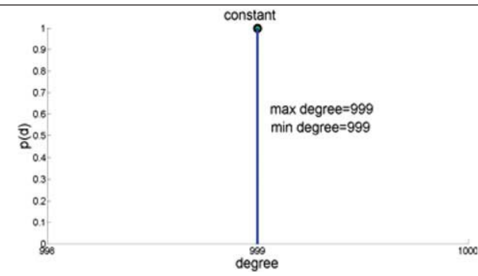
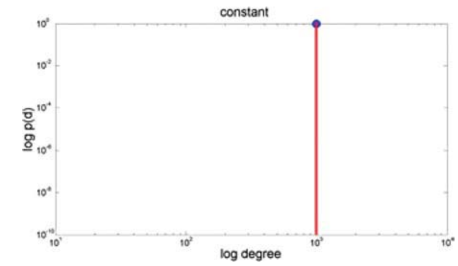
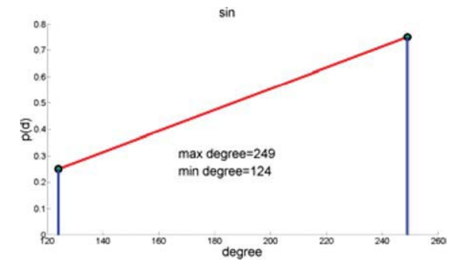
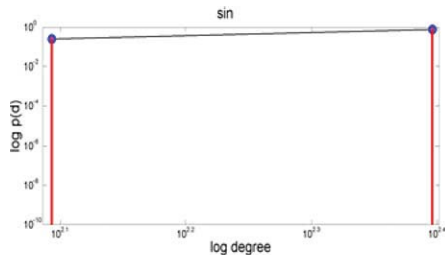
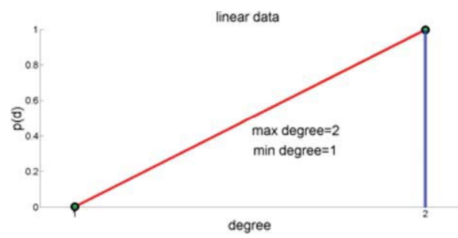
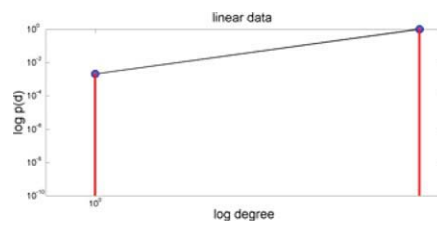
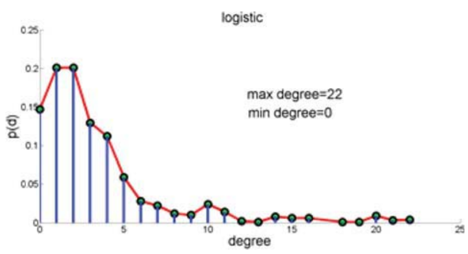
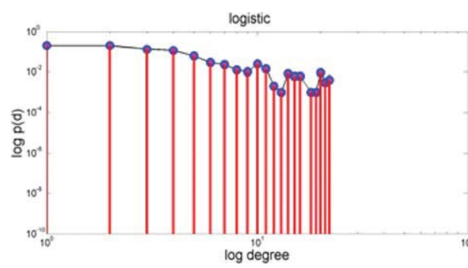
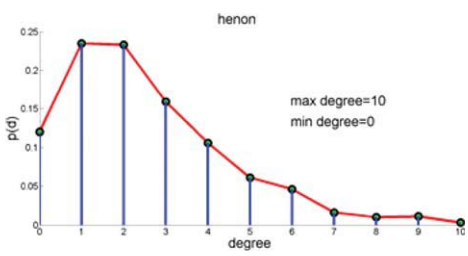
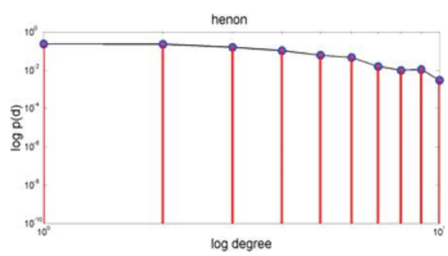
Evolution of Networks from Six Time Series when m Increases



DOI 10.1002/cplx

TABLE 2

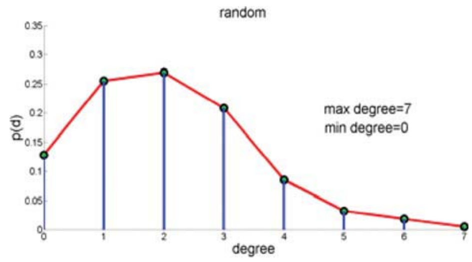
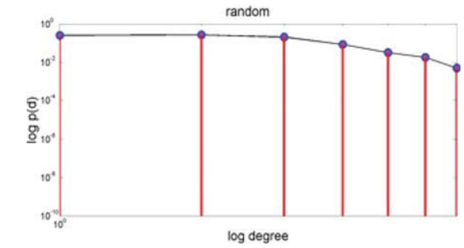
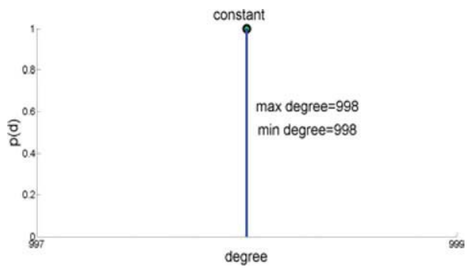
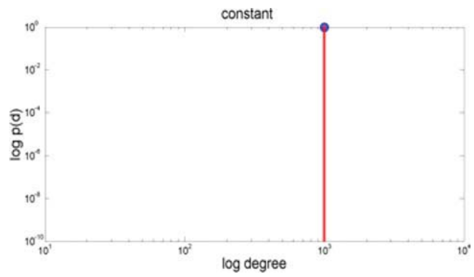
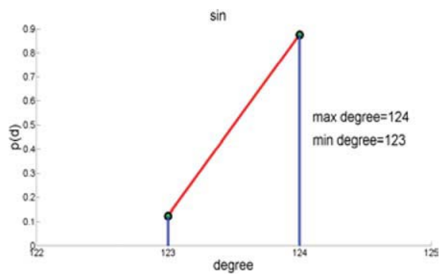
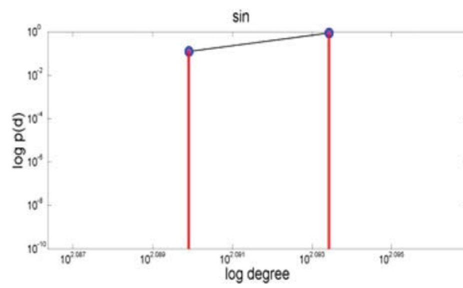
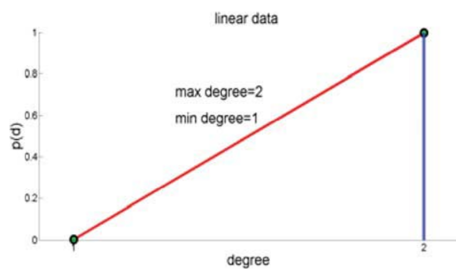
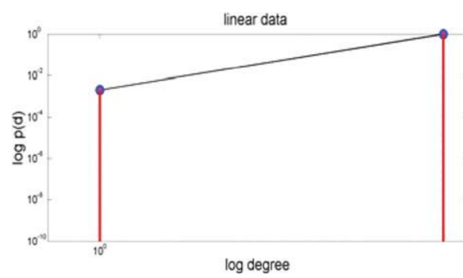
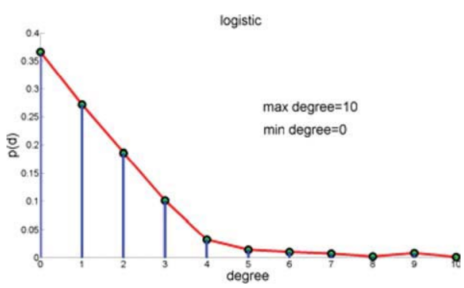
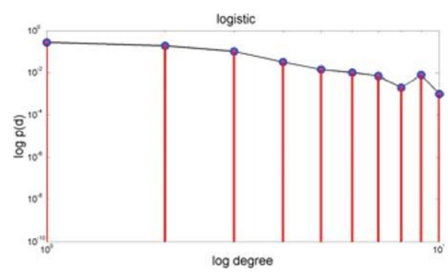
Distribution of Degree

	Degree Distribution on Linear Scales	Degree Distribution on Logarithm Scales
m = 1		
		
		
		
		

DOI 10.1002/cplx

TABLE 2

Continued

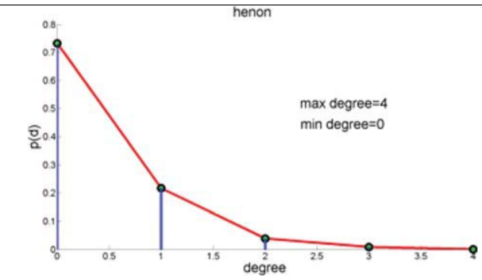
	Degree Distribution on Linear Scales	Degree Distribution on Logarithm Scales
		
m = 2		
		
		
		

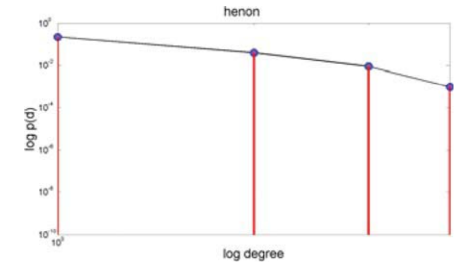
DOI 10.1002/cplx

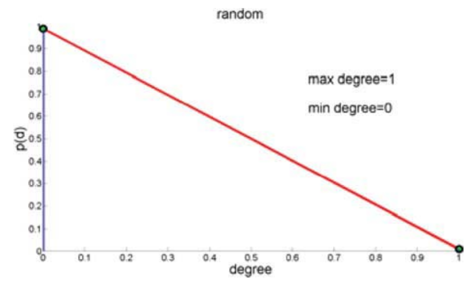
TABLE 2

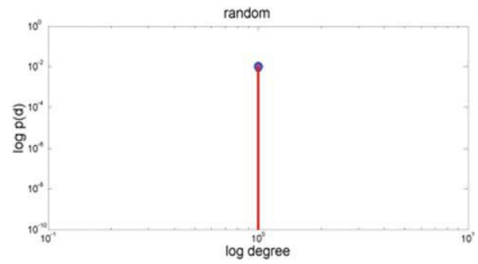
Continued

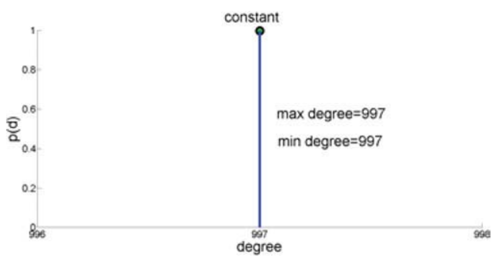
	Degree Distribution on Linear Scales	Degree Distribution on Logarithm Scales
m = 3		

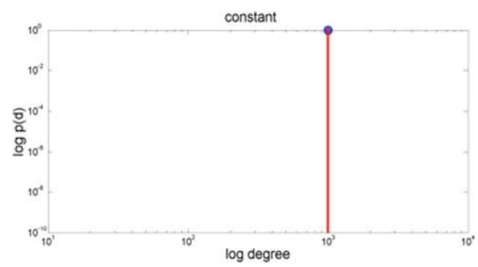


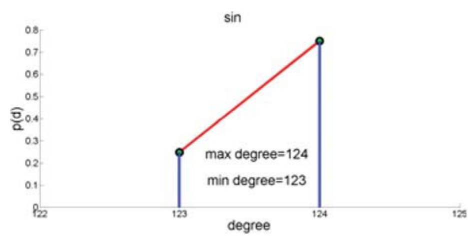


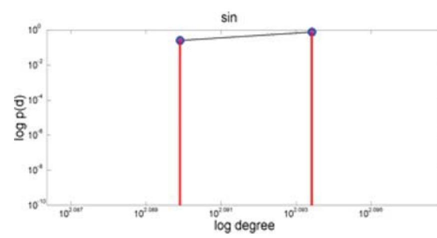


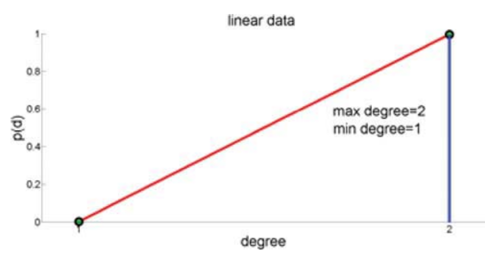


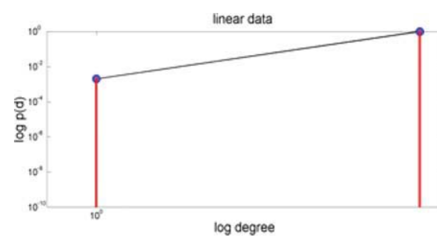








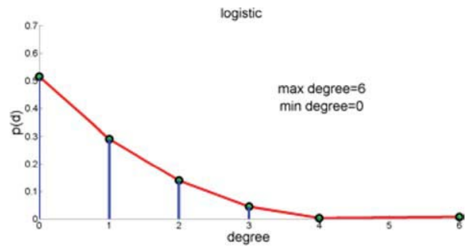
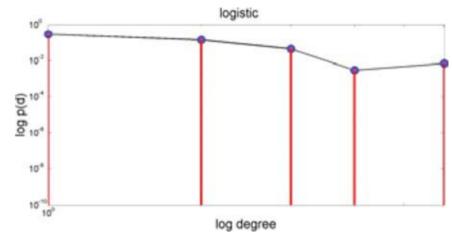
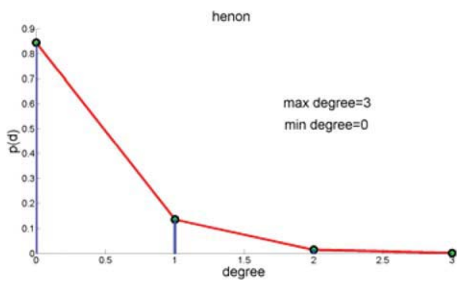
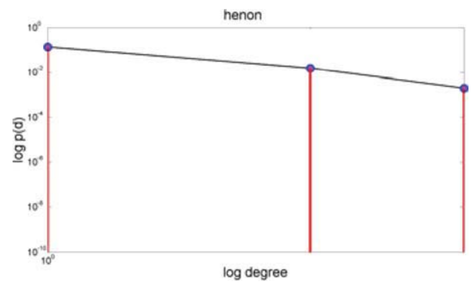
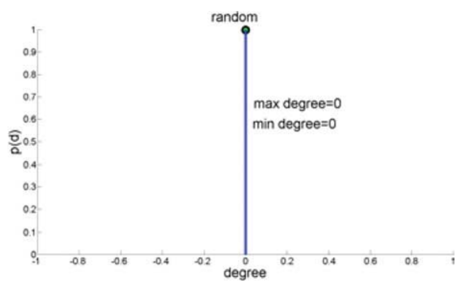
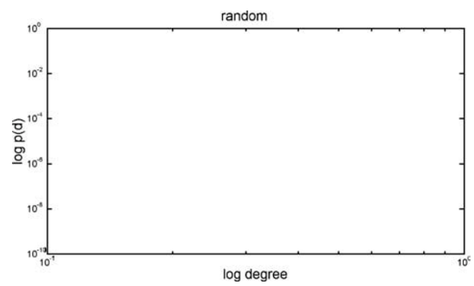




DOI 10.1002/cplx

TABLE 2

Continued

Degree Distribution on Linear Scales	Degree Distribution on Logarithm Scales
	
	
	

5. The chaos system driven by nonlinear chaotic forces displays nondeterministic and stochasticlike behavior. The driving force is nonlinear and partly greater than linear force. Therefore, the graph from a chaotic series starts to include isolated nodes and cannot form a connected graph. The degree distribution is approximately power law. When m increases, the topological characteristics of the network graph is unstable. The greater m is, the worse the connectivity becomes, and more isolated nodes will emerge. Compared with the random series, the graph from the chaotic series remains partly connected when m grows. This is due to the internal certainty of the chaotic time series and the concentrated area of the chaotic attractor.
6. For the random series, when $m = 1$, some connected regions exist in the graph. As m increases, the connectivity of the graph degrades and the number of isolated nodes increases rapidly. When $m = 2$, only two nodes are connected and their degree is one. When $m = 3$, all nodes become isolated. The graph becomes a completely unconnected graph.

5.2. Changing Trend of d_{max} and Δ

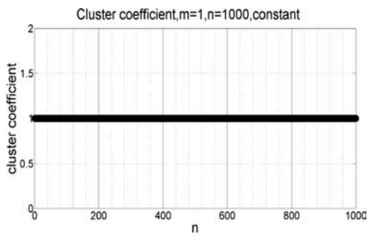
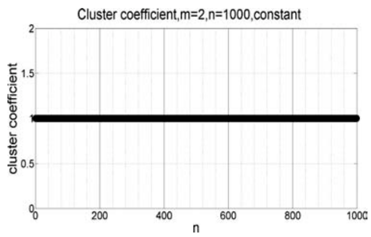
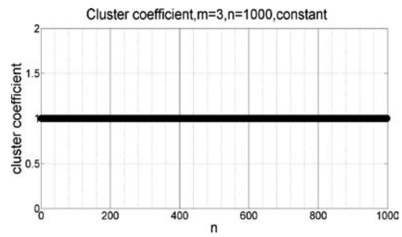
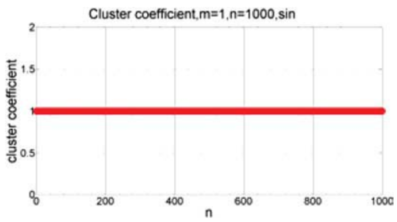
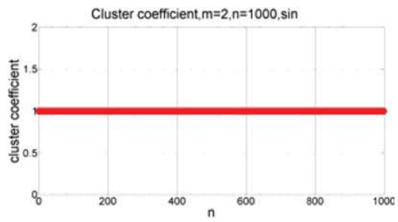
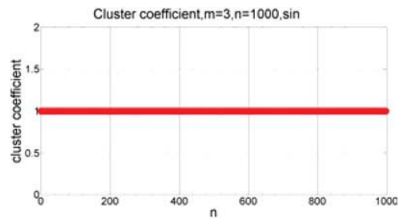
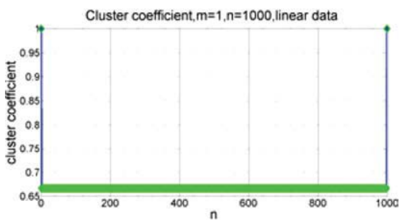
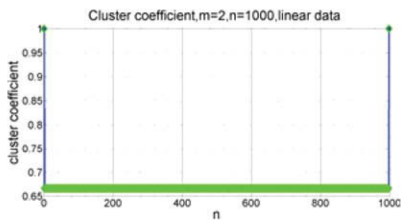
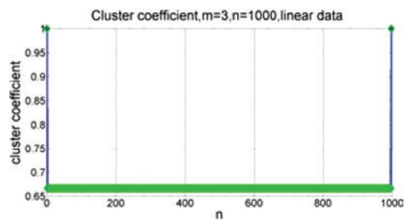
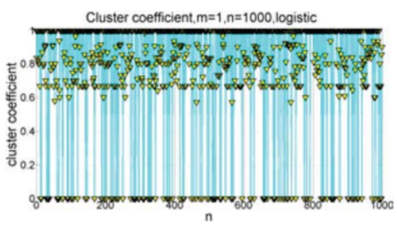
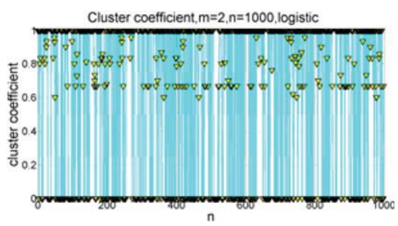
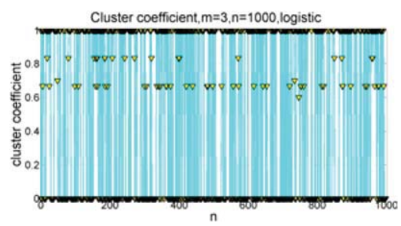
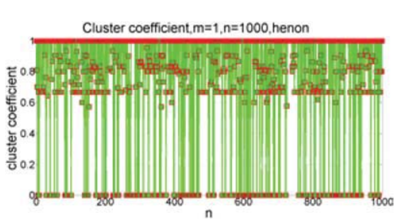
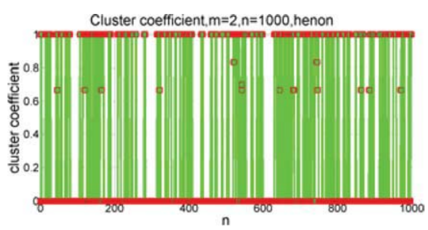
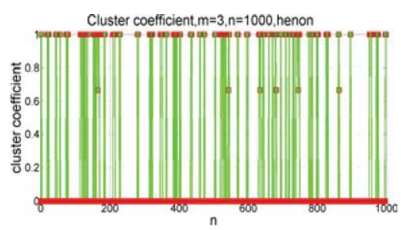
The parameter of d_{max} represents the size of the entire distance space, and the changing trend of d_{max} with n increasing represents the changing trend of time series. The parameter of Δ represents the linear partition to the distance between the minimum and maximum of time series.

As mentioned in Section 3.2, the judge distance Δ depends on d_{max} and n, but d_{max} also changes when n changes. The value of d_{max} represents the size of covering space of the time series. Because d_{max} is the maximum of

DOI 10.1002/cplx

TABLE 3

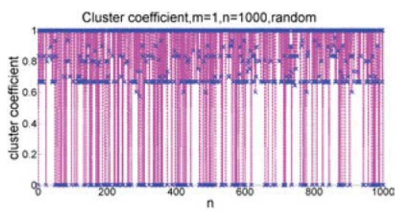
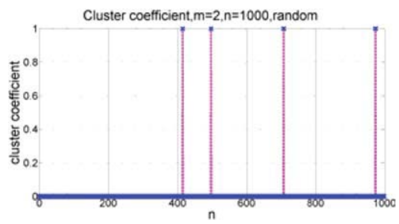
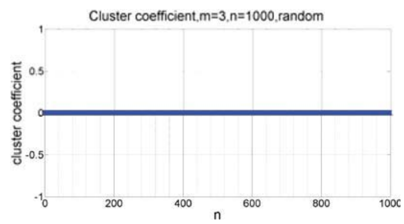
Clustering Coefficient of Six Time Series when m Increases

m = 1	m = 2	m = 3
		
		
		
		
		

DOI 10.1002/cplx

TABLE 3

Continued

m = 1	m = 2	m = 3
		

the entire distance set, it is a measure at a macrolevel. The change path of d_{max} with n increasing is given in Figure 7. In fact, the changing trend of d_{max} reflects the moving property of time series.

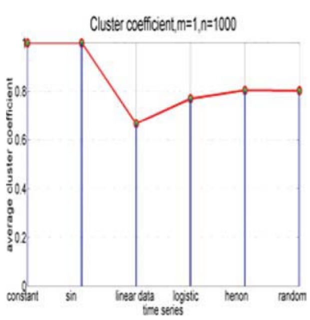
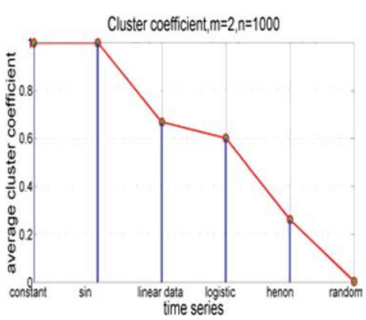
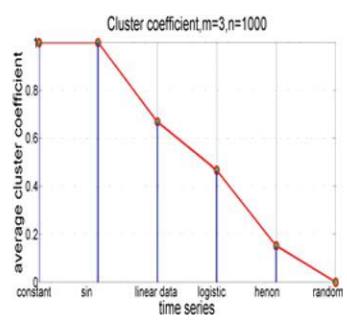
In Figure 7, when n increases, d_{max} of the constant series keeps an unchanged value of 0; d_{max} of the periodic series remains unchanged; d_{max} of the linear time series synchronously increases with and increasing n; d_{max} of the other three series, including logistic, Henon, and random series, increase within a small range. The reason for this small increase is that the sample space increasing lead to the distance space $\{d_{ij}|i = 1, 2, \ldots, k; j = 1, 2, \ldots, k; i \neq j\}$ increasing. Therefore, there are some opportunities for d_{max} to increase. But after n increasing to enough large value, d_{max} trends to a stable value. The result is just consistent with our analysis result in Section 3.2.

The relation between Δ and the size n of the series is shown in Figure 8. Among the six series, at a fixed embedded dimension m, when n increases, Δ of constant series retains an unchanged value of 0; Δ of linear series keeps an unchanged value; Δ of the rest four series decline. In the formula of $\Delta = d_{max}/(k - 1)$, where $k = n - m + 1$, n is in denominator. When the increasing speed of the numerator of d_{max} is slower than that of n, Δ will decline, as shown in Figure 7.

To show clearly, the y-axis is at logarithm in Figure 8. Because Δ equals to 0 for the constant series, the Δ of the constant series cannot be reflected in Figure 8.

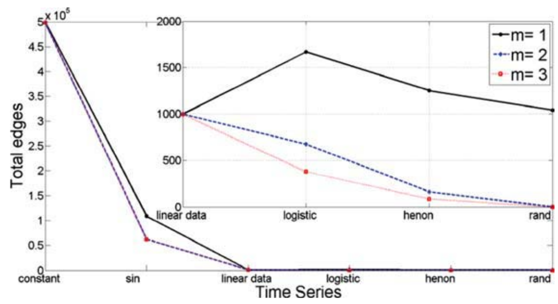
TABLE 4

Average Clustering Coefficient

m = 1	m = 2	m = 3
		

DOI 10.1002/cplx

FIGURE 6



Total edges of six time series.

The relation between Δ and m is shown in Figure 9. At the fixed size n, the Δ is consistent with the trend of $d_{\max}$. The Δ of the other series increases, except Δ of the constant series, which remains an unchanged value of 0. The reason for increase is that when increasing items in the formula $d_{ij} = \sqrt{|x_{i_1} - x_{j_1}|^2 + |x_{i_2} - x_{j_2}|^2 + |x_{i_3} - x_{j_3}|^2 + \cdots |x_{i_m} - x_{j_m}|^2}$ lead to d_{ij} increasing, $d_{\max}$ also increases, and then Δ increases.

5.3. Discussion of $p(d \leq \Delta)$

We analyze the changing trends of $d_{\max}$ and Δ with m and n for the six series. However, the number of edges of the network depends on how many distances are less than or equal to Δ—that is, $N(d \leq \Delta)$. The value of $N(d \leq \Delta)$ is directly related to the probability density $p(d \leq \Delta)$. Figure 10 is the probability density distribution graph for the six series.

Figure 10 gives the distribution of $p(d)$, which has a different shape from one series to another. For the constant series, the distance between any two nodes and Δ both are 0, then $p(d \leq \Delta) = 1$. For the periodic series, $p(d)$ does not equal 0 at several discrete locations, and the remainder of the locations all equal 0. For the linear series, the probability density distribution is a straight line with a negative slope. For the logistic, Henon, and random series, when m increases, the left-most value of $p(d)$ quickly decreases. That is to say, the probability of the small d decreases rapidly, so that $p(d \leq \Delta)$ decreases very significantly which, in turn, leads to a rapid decline of the total edges.

When $m = 1$, the distributions of $p(d)$ of logistic and Henon series are very similar to that of the random series, and even with the linear series. The same point of these four series is that the value of $p(d)$ of the small d is relatively large, the value of $p(d)$ of the large d is relatively small, and the distribution is an approximate straight line. At this time, the chaotic series cannot be distinguished from the random series. The situation of $m = 1$ for the logistic and Henon series is different from m equaling to other values. When m increases, for logistic, Henon, and random series, the value of $p(d)$ of the small d rapidly decline. Within that, the probability of the small d of the random series declines fastest. With m increasing, the distribution shapes of $p(d)$, and of the chaotic (logistic and Henon) and random series, begin to separate. The distribution shape of the $p(d)$ of the random series is clearly close to a normal distribution, but the distribution shape of the $p(d)$ of the logistic and Henon series shows a peak form. The peak means the value near the median (or mean) greater than the theoretical estimate of the normal distribution.

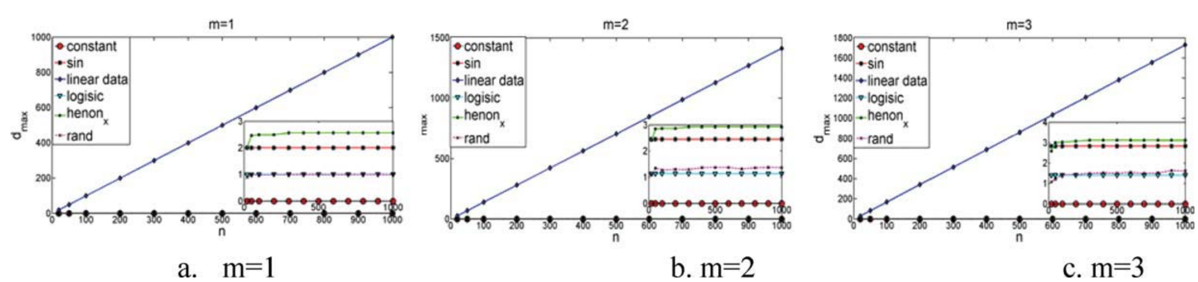
Figure 11 shows the different values $p(d < \Delta)$ at different m for the six series. From Figure 11, with m changing, $p(d < \Delta)$ maintains the value of 1 for the constant series; keeps the value of 2×10^{-3} for the linear series; keeps a stable value after a slight drop from the initial value of 2×10^{-2} for the periodic series; but declines for the chaotic and random series.

5.4. Interpretation for the Structure of Generated Network from Six Time Series

5.4.1. Constant Series

Because the value of the constant series is fixed, that is, $x(i) = a$, $i = 1, 2, \ldots, n$, all points in the reconstructed phase space also equal to the same value. For example,

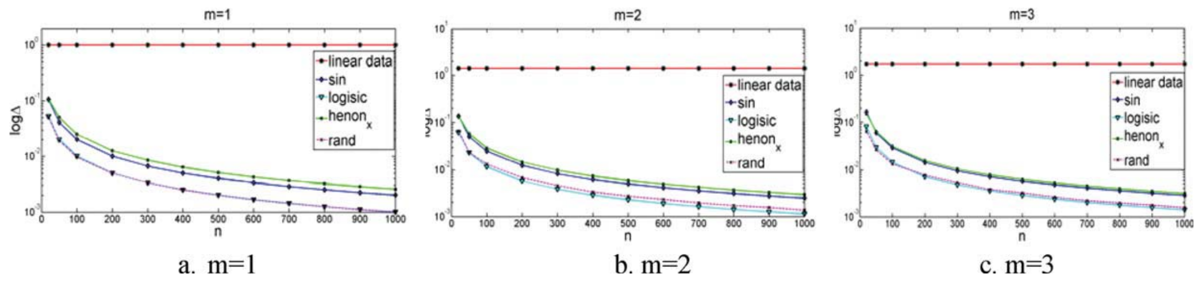
FIGURE 7



Change path of $d_{\max}$ when n increases. (a) m = 1; (b) m = 2; and (c) m = 3.

DOI 10.1002/cplx

FIGURE 8



Change path of Δ when n increases. (a) m = 1; (b) m = 2; and (c) m = 3.

when $m = 3$, all points are as follows:

$$\begin{matrix} a & a & & a \\ a & a & \ldots & a. \\ a & a & & a \end{matrix}$$

Therefore, the distance between any two points is 0—that is, $d = 0$. In addition, $\Delta = 0$, so that $p(d \leq \Delta) = 1$. Consequently, no matter what the value of n and m are, the generated network is always a fully connected graph.

5.4.2. Periodic Series

Because the periodic series is a series with some repeated elements, the distance between any two points in phase space is repeated. Then, the distance set is a finite set including several discrete values. In the distance set, the number of distances that are larger than Δ and smaller than Δ are both certain. Therefore, the degree distribution is a discrete distribution of finite elements. The $p(d \leq \Delta)$ graph shows, when m equals to a different value, the value of $p(d \leq \Delta)$ keeps a stable value after a slight decrease from the initial value of 2×10^{-2}. At different m, the characteristics of topology structure of generated network are similar, and all are regular graphs.

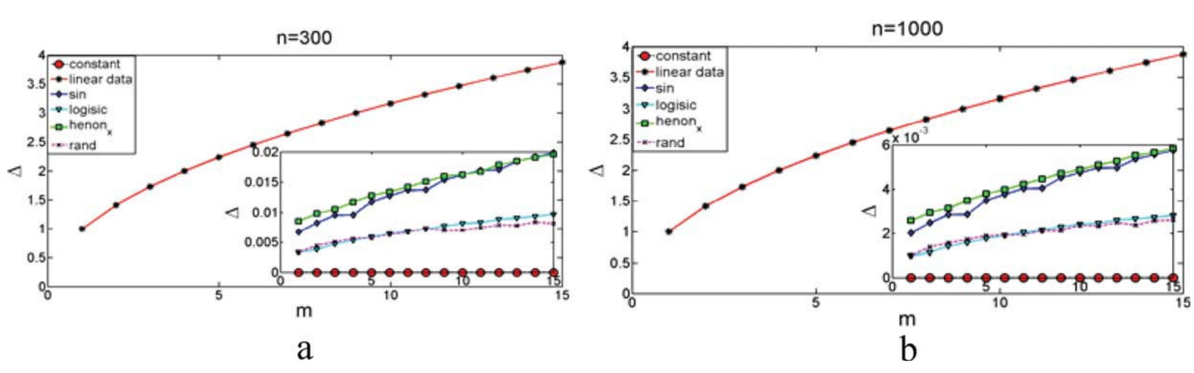
5.4.3. Linear Series

No matter what value n and m are, the distance between adjacent points in phase space is equal to the judgment distance Δ. As well, any other distances are an integer multiple of Δ, which is larger than Δ. Because only the adjacent nodes are connected, a tree is created from the start node to the end node. At the same time, $p(d \leq \Delta) = \frac{k-1}{C_k^2} = \frac{n-m+1-1}{C_{n-m+1}^2} = \frac{n-m}{C_{n-m+1}^2} = \frac{2}{n-m+1}$. The $p(d \leq \Delta)$ graph shows, when $n = 1000$, all $p(d \leq \Delta)$ approximate to 2×10^{-3}, which equals to above theoretical value. The generated network is always a tree.

5.4.4. Chaotic Series

Because of its randomness, when the dimension m of phase space increases, the randomness leads to $p(d \leq \Delta)$ gradually decreasing. Because the chaotic series contains determinacy internal cause, there is strange attractor. The

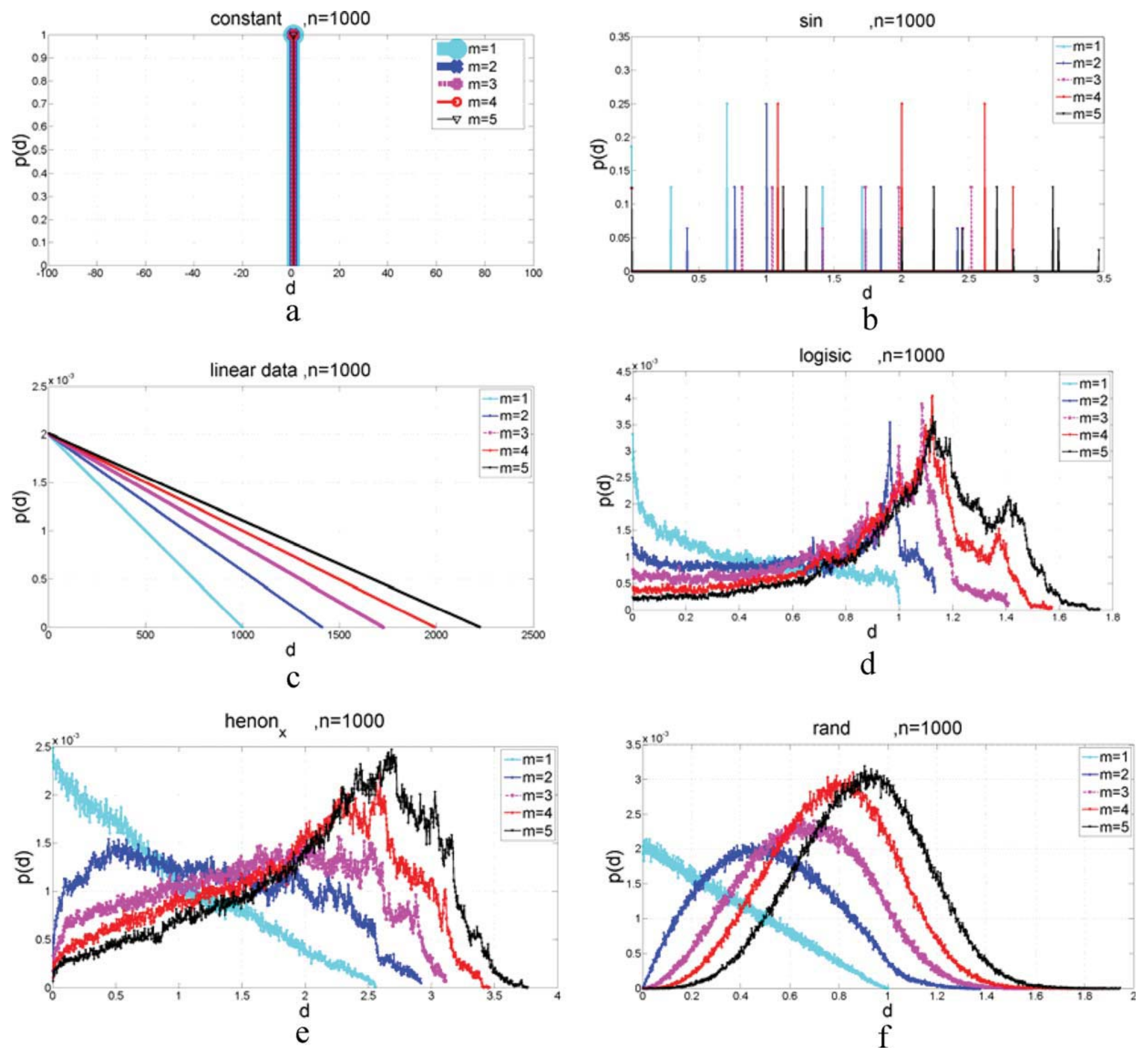
FIGURE 9



Change path of Δ, when embedded dimension m increases. (a) $n = 300$ and (b) $n = 1000$.

DOI 10.1002/cplx

FIGURE 10



Distribution of $p(d)$ for six series at different m. (a) Constant series; (b) periodic series; (c) linear series; (d) logistic series; (e) Henon series; and (f) random series.

series is density near the attractor in the phase space. Therefore, when $m = 3$, some distances between two node is smaller than Δ, namely $p(d \leq \Delta) \neq 0$. There are a few connections in the generated networks.

5.4.5. Random Series

Due to its complete randomness, when the dimension m of phase space increases, the strong random driving force causes the $p(d \leq \Delta)$ to rapidly decrease. The decreasing speed is faster than that of a chaotic series containing a determinacy internal cause. When $m = 3$, the strength of randomness has led to the distances between any two points in the phase space larger than a linear dividing distance Δ. Hence, $p(d \leq \Delta) = 0$. No edges exist in the generated network.

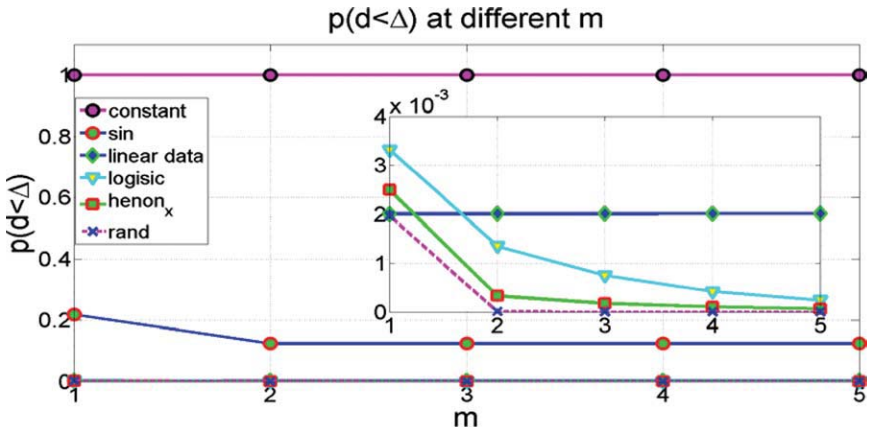
5.5. Summary of Results and Identification Method for Time Series

On the basis of the above observations, we can draw the following conclusions.

1. Our algorithm can display the divergent characteristics of a dynamic system. When the divergent speed is faster than a linear divergence, the nodes are not connected.
2. Our algorithm can reflect the random nature of a dynamic system. When the dimension of the reconstructed phase space m increases, the randomness becomes stronger, the network is more separated, or there are more isolated nodes.

DOI 10.1002/cplx

FIGURE 11



$p(d < \Delta)$ for six series.

3. Our algorithm can depict the certainty and randomness of a chaotic time series. The certainty of a chaotic time series is shown by the relative stability of the generated graph when m increases. The connectivity drops slowly, and the characteristics of the degree distribution maintain consistency. There always are some connected regions. The randomness of a chaotic time series is shown by the fact that the connectivity gradually deteriorates and the number of isolated nodes gradually increase with m.
4. Our algorithm can reflect the randomness of a random series. The randomness of a random series is manifested by the fact that the relation between nodes rapidly disappears when m increases.

Therefore, we have found a simple method to distinguish the nature of time series based on network graph. We can also calculate the degree distribution of network graphs generated by our algorithm, and can distinguish a time series based on the following rules when m increases.

TABLE 5

Identification Method for Time Series

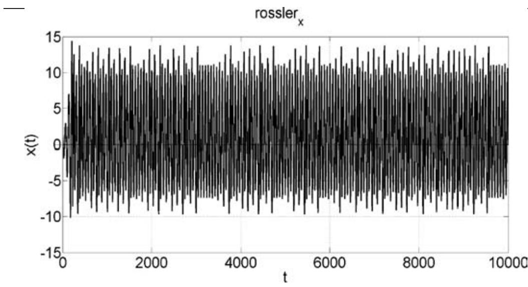
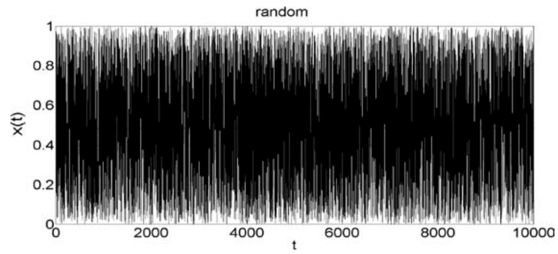
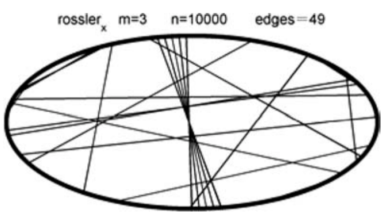
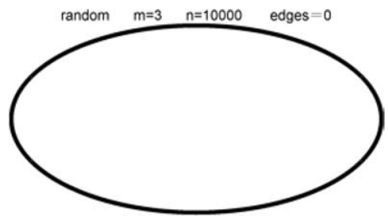
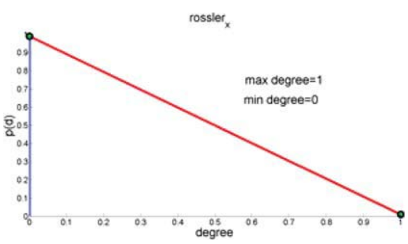
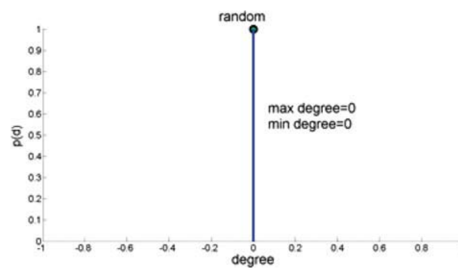
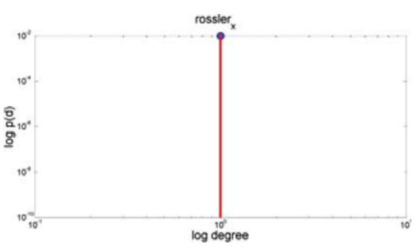
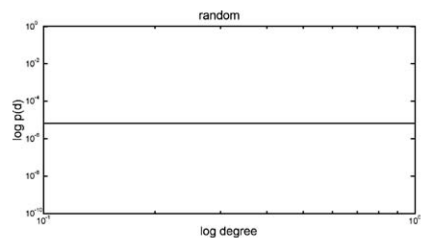
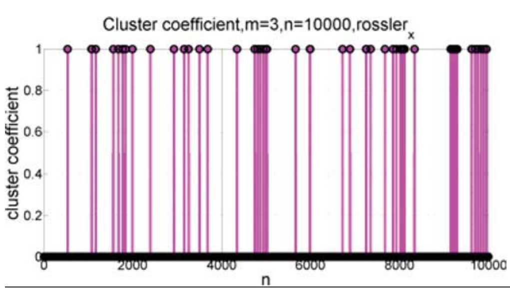
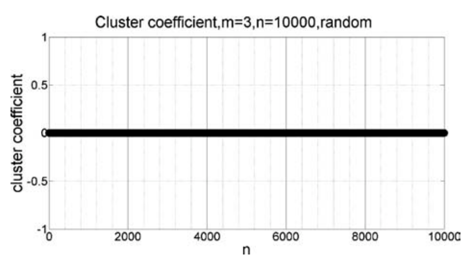
When $m \geq 3$	
Basic principles: The number of edges shows the certainty features of the time series. More edges indicate more certainty, and fewer edges indicate more randomness.	
Property of network	Characteristics of time series
1. Fully connected graph	Constant series
1. Stable discrete degree distribution 2. Regular discrete clustering coefficient distribution 3. Large number of total edges	Periodic series
1. Discrete degree distribution 2. Regular discrete clustering coefficient distribution 3. Small number of total edges	Linear divergent series
1. Approximate power-law degree distribution 2. Few clustering coefficient unequal to zero when $m = 3$ 3. Total edges decrease rapidly with increasing m, but at $m = 3$, unequals to 0	Chaotic series
1. At $m = 3$, degree decreases rapidly to zero 2. All clustering coefficient equal to zero	Random series

DOI 10.1002/cplx

TABLE 6

Comparison between Rössler Series and Random Series

n= 10,000, m = 3

Rössler Series	Random Series
	
	
	
	
	

DOI 10.1002/cplx

1. It is a constant series if the network graph remains fully connected.
2. It is a periodic series if the degree distribution of the network is stable and discrete (regular graph) and the average degree is high.
3. It is a linear divergent series if the degree distribution of the network is stable and discrete (regular graph) and the average degree is low.
4. It is a chaotic series if the degree distribution of the network is approximately power law.
5. It is a random series if the degree of every node decreases rapidly to zero (completely unconnected graph).

As m increases, the speed of a node's degree decreasing to zero reflects the randomness strength of a time series. The change in m can cause the graphs of different time series to respond in different ways, which helps to uncover the information of state variables hidden in a one-dimension time series. For example, as is shown in Table 1, when $m = 1$, $n = 25$, the number of edges of the constructed network graph from a random time series is greater than that from a chaotic time series. Therefore, in the case where $m = 1$, a chaotic time series cannot be distinguished from a random series. In other words, when $m = 1$, it is difficult to tell which one is more random. However, when m is increased to 2, the structural characteristics of the graphs from a chaotic time series and a random time series are clearly differentiable.

So, we determine the identification method as appropriate for the time series, as shown in Table 5.

6. TEST FOR IDENTIFICATION METHOD

We apply the identification method to test a chaotic time series and a random time series with a length of 10^4 to determine whether they can be distinguished. The Rössler system [28], as a chaotic system, has a simple structure, contains only one nonlinear item, and has characteristics of chaotic dynamics under the following parameters:

$$\begin{cases} \dot{x} = -(y+z) \\ \dot{y} = x + ay \\ \dot{z} = b + xz - cz. \end{cases}$$

Using the parameters $a = 0.2$, $b = 0.2$, and $c = 5.7$, we then can achieve a chaotic system. We compared the Rössler system and the random series with $n = 10{,}000$. The generated network, degree distribution, and clustering coefficient distribution are shown in Table 6. The results show that when $m = 3$, the total edges of completely random series turns into zero. But at this time, the Rössler series is different. There are a few connected groups in the network (their clustering coefficient unequal to zero).

From Table 6, we can observe the differences between the Rössler series and the random series at $m = 3$. When m equals 3 in the network generated from the Rössler series, a few edges still exist and the cluster coefficient of some nodes do not equal to zero. At the same time, however, the graph from the random series shows no edge, and all cluster coefficients equal to zero. From above results, we can conclude that the Rössler series is a Chaotic series which includes certainty, rather than being a random series, so that the Rössler series is identifiable from the random series.

7. CONCLUSION

This article proposes a method for transforming a time series into a network graph. Our algorithm relies on the space distance to construct a network. By controlling the dimensions of reconstructed phase space, we can distinguish different types of time series. Chaotic and random time series, particularly, can be very easily identified. Our method can visually display system divergence and convergence, as well as certainty and randomness. Notably, a constant series yields a fully connected graph, whereas a random time series generates a completely unconnected graph. Maybe, this method will be useful for the analysis of complex nonlinear systems such as chaos and random systems; the value lies in distinguishing the outcome—the time series—of the system, as a means of determining whether the system is a chaos system or a random system.

Acknowledgments

This work was supported, in part, by the National Natural Science Foundation of China (Grant Nos. 70801066, 71071167, and 71071168), and by a grant from Sun Yat-sen University Basic Research Funding (Grant No. 1009028).

REFERENCES

1. Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. Rev Mod Phys 2002, 74, 47.
2. Albert, R.; Jeong, H.; Barabási, A.L. Diameter of the World-Wide Web. Nature 1999, 401, 130.
3. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. Science 1999, 286, 509.
4. Newman, M.E.J. Networks: An Introduction; Oxford University Press: Oxford, 2010.
5. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.U. Complex networks: Structure and dynamics. Phys Rep 2006, 424, 175.

DOI 10.1002/cplx

6. Sinatra, R.; Condorelli, D.; Latora, V. Networks of motifs from sequences of symbols. Phys Rev Lett 2010, 105, 178702.
7. Yule, G.U. On a method of investigation periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. Philos Trans R Soc Lond A 1927, 226, 267.
8. Walker, G. On periodicity in series of related terms. Proc R Soc, Ser A 1931, 131, 518.
9. Box, G.E.P.; Jenkins, G.M. Time Series Analysis: Forecasting and Control. Holden-Day: San Francisco, 1970.
10. Box, G.E.P.; Pierce, D.A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J Am Stat Assoc 1970, 65, 1509.
11. Kim, C.-J.; Nelson, C.R. State-Space Models With Regime Switching. MIT Press: Cambridge, MA, 1999.
12. Tong, H. On a threshold model. In: Pattern Recognition and Signal Processing; Chen, C.H., Ed.; Sijthoff and Noordhoff: Amsterdam, 1978.
13. Terasvirta, T. Specification, estimation, and evaluation of smooth transition autoregressive models. J Am Stat Assoc 1994, 89, 208.
14. Lundberg, S.; Teräsvirta, T. Evaluating GARCH Models. J Econom 2002, 110, 417.
15. Grassberger, P.; Procaccia, I. Characterization of strange attractors. Phys Rev Lett, 1983, 50, 346.
16. Cohen, A.; Procaccia, I. Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems. Phys Rev A 1985, 31, 1872.
17. Schreiber, T. Interdisciplinary application of nonlinear time series methods. Phys Rep 1999, 308, 1.
18. Packard, N.H.; Crutchfield, J.P.; Farmer, J.D.; Shaw, R.S. Dynamical systems collective, geometry from a time series. Phys Rev Lett 1980, 45, 712.
19. Takens, F. Detecting strange attractors in turbulence. In: Lecture Notes in Mathematics; Springer: Berlin, 1981; p 366.
20. Zhang, J.; Small, M. Complex network from pseudoperiodic time series: Topology versus dynamics. Phys Rev Lett 2006, 96, 1.
21. Lacasa, L.; Luque, B.; Ballesteros, F.; Luque, J.; Nuno, J.C. From time series to complex networks: The visibility graph. Proc Natl Acad Sci 2008, 105, 4972.
22. Lacasa, L.; Luque, B.; Luque, J.; Nuno, J.C. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. Eur Phys Lett 2009, 86, 30001.
23. Luque, B.; Lacasa, L.; Ballestero, F.; Luque, J. Horizontal visibility graphs: Exact results for random time series. Phys Rev E 2009, 80, 046103.
24. Hirata, Y.; Horai, S.; Aihara, K. Reproduction of distance matrices and original time series from recurrence plots and their applications. Eur Phys J: Spec Top 2008, 164, 13.
25. Xu, X.; Zhang, J.; Small, M. Superfamily phenomena and motifs of networks induced from time series. Proc Natl Acad Sci, 2008, 105, 19601.
26. Li, Y.; Cao, H.D.; Tan, Y. A comparison of two methods for modeling large-scale data from time series as complex networks. AIP Adv 2011, 3, 012103–1.
27. Diestel, R. Graph Theory, 3rd ed.; Springer-Verlag: Berlin, 2005.
28. Rössler, O.E. An equation for continuous chaos. Phys Lett A 1976, 57, 397.

DOI 10.1002/cplx