Characterizing Pseudoperiodic Time Series through Complex Network Approach

Jie Zhang [a] Junfeng Sun [a] Xiaodong Luo [b] Kai Zhang [c]
Tomomichi Nakamura [d] Michael Small [a]

[a] *Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

[b] *Oxford Center for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford, UK*

[c] *Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong*

[d] *Sony Computer Science Laboratories, Inc. Tokyo, Japan*

Abstract

Recently a new framework has been proposed to explore the dynamics of pseudoperiodic time series by constructing a complex network [Phys. Rev. Lett. 96, 238701 (2006)]. Essentially, this is a transformation from the time domain to the network domain, which allows for the dynamics of the time series to be studied via the organization of the network. In this paper, we focus on the deterministic chaotic Rössler time series and stochastic noisy periodic data that yield substantially different structures of the networks. In particular, we test an extensive range of network topology statistics, which have not be discussed in previous works, but which are capable of providing a comprehensive statistical characterization of the dynamics from different angles. Our goal is to find out how they reflect and quantify different aspects of specific dynamics, and how they can be used to distinguish different dynamical regimes. For example, we find that the joint degree distribution appears to fundamentally characterize the spatial organizations of the cycles in phase space, and this is quantified via assortativity coefficient. We applied the network statistics to the electrocardiograms of a healthy individual and an arrythmia patient. Such time series are typically pseudoperiodic, but are noisy and nonstationary and degrade traditional phase-space based methods. These time series are, however, better differentiated by our network-based statistics.

Key words: Nonlinear time series analysis; Pseudoperiodic signal; Complex network transformation; Unstable periodic orbits; Chaos

Email address: enzhangjie@eie.polyu.edu.hk (Jie Zhang).

1 INTRODUCTION

In the past few years substantial attention has been devoted to the topic of complex networks, which provides a new paradigm and profound insights into many of complicated systems in engineering, social and biological fields [1–3]. Recently a network approach to pseudoperiodic time series analysis has been proposed as a kind of transformation from the time domain to the complex network domain [4]. This method exploits the periodicity contained in the time series [5] and segments it into sequential cycles. By considering the individual cycles as the nodes and associating the network connectivity with the correlation among cycles, the time domain dynamics are naturally encoded into the network configuration.

Representing the time series through the corresponding complex network, we can then explore the dynamics of the time series from the network organization, which is quantified via a number of topological statistics. Such statistics can provide new information about the phase space geometry of the cycles within pseudoperiodic time series. For example, we have shown that noisy periodic signals correspond to random networks, and chaotic time series generate networks that exhibit small world and scale free features [4]. Furthermore, the validity of the network statistics from complex network to capture the cycles' spatial correlation intrinsic to the chaotic time series has been tested [6] using the PPS surrogate [7], and such statistics are shown to be more sensitive to the subtle changes in dynamics of the pseudoperiodic time series than conventional invariants such as correlation dimension [6]. However, only the basic statistics of the network are studied and utilized in previous works [4,6], an extensive set of characteristics, for example, those associated with local and global properties of the network, are left unchecked. In this paper, we quantitatively analyze different networks extracted from two typical pseudoperiodic time series with distinct dynamics in terms of the various network metrics. Our goal is to provide a comprehensive statistical characterization of the dynamics of the time series via an extensive set of network statistics. Specifically, we want to associate different aspects of the dynamics of the time series with the topological indices of the network, and demonstrate how such statistics can be used to quantify and distinguish different dynamical regimes.

By pseudoperiodic we mean the time series has an dominant frequency component (i.e., a predominant peak in its power spectrum) so that the data demonstrates a repetitive pattern. For example, the human ECG recording, with each cycle representing one heartbeat, and the human gait time series, with the cycles corresponding to the strides, see Fig. 1. Pseudoperiodic time series are also refereed to as (irregular) cyclic processes or almost periodic signals in the literature [8]. The dominant frequency in the data guarantees that it can be readily divided into individual cycles that follow a similar pattern

and are of similar length. Usually the segmentation is done by the local minimums (or maximums) for sinusoidal-like data (e.g., the phase-coherent chaotic Rössler data in Fig. 2 (a), where the local minimums are indicated by circles). For pseudoperiodic signals with complex waveforms, the segmentation can be done by utilizing the marker events such as the pulse-like R wave in ECG data (see Fig. 1(a)), or making use of the "repetitiveness" of the pattern in the time series (see Fig. 1(b)).

It should be noted that our method may not be suitable for those oscillatory data that have no dominant peak in their power spectrum, since it will be difficult to decide how to segment the time series into cycles. For example, Fig. 2 (b) and (c) give two possible segmentation schemes for the data from Rössler system in the funnel regime. The longer cycle denoted by thick line in Fig. 2 (c) may be further partitioned into two smaller cycles, which are indicated by dashed and dotted lines in Fig. 2 (b), respectively. In fact a thorough segmentation like Fig. 2 (c) may lead to cycles that are too different (especially in length) to be properly compared by the method described in this paper, thus an appropriate segmentation still needs to be defined for such non-pseudoperiodic but oscillatory time series. In this paper we confine ourself to the data of pseudoperiodic type, and two typical time series are chosen for analysis. These are the data from the x component of the chaotic Rössler system:

$$\begin{cases} x' = -(y+z) \\ y' = x + 0.398y \\ z' = 2 + z(x-4) \end{cases} \tag{1}$$

shown in Fig. 2 (a), and the noisy periodic signal $y_n = \sin(2\pi\omega n) + b\eta_n$ ($b = 0.2836$), where η is I.I.D noise following $\eta \sim N(0, \sigma^2)$, which are the same time series as were used in [4]. These two time series contain an obvious periodic component, and can be easily segmented into consecutive cycles. The sketch of the chaotic Rössler attractor is shown in Fig. 3.

It should be noted that we have chosen these two archetypal systems for the purpose of illustrating our methods. In Section 4 we will turn to the much more difficult problem of analyzing real-world data.

Both correlation coefficient and phase space distance can be used in comparing two cycles, and we have shown that these two measures are essentially equivalent [4]. In this paper we choose to use the phase space distance, since it is physically meaningful. The phase space distance between cycle C_i and C_j

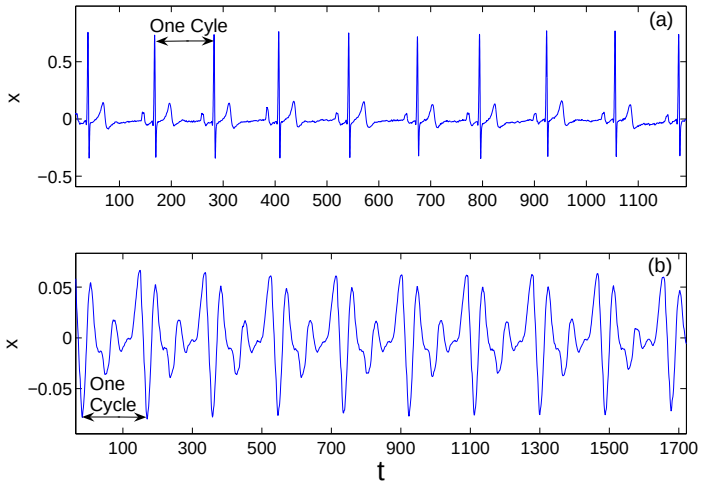

Fig. 1. Typical pseudoperiodic data from (a) human electrocardiogram and (b) human gait data.

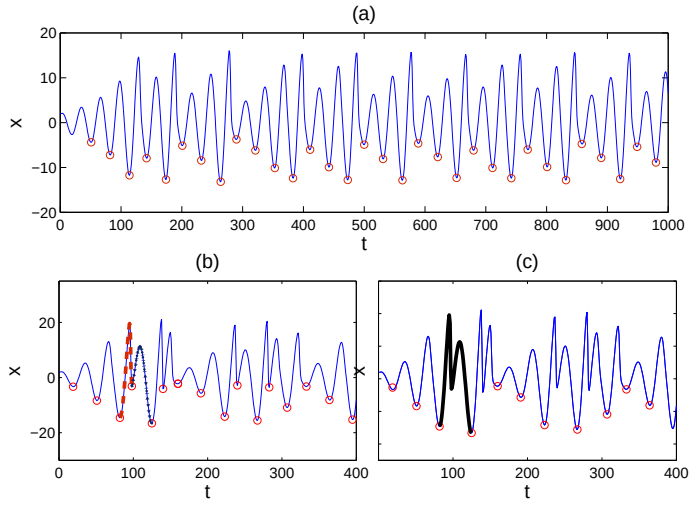

Fig. 2. Segmentation of the time series into cycles for the x component of (a) chaotic Rössler system, (b) and (c) Rössler system in the funnel regime. As is shown, the coherent chaotic Rössler time series can be readily divided into consecutive cycles of similar lengths by local extremes, while for the Rössler system in the funnel regime, the segmentation may not be unique, and needs to be taken with great care.

is defined as:

$$D_{ij} = \min_{l=0,1,\cdots,|l_j-l_i|} \frac{1}{\min(l_i, l_j)} \sum_{k=1}^{\min(l_i,l_j)} \|X_k - Y_{k+l}\| \tag{2}$$

where X_k and Y_k is the kth point of C_i and C_j in the reconstructed phase space, and l_i and l_j are the lengths of C_i and C_j, respectively [5].

The complex network constructed from the time series can be treated either as a weighted network, in which the weight is defined as the phase space distance between cycles, or as an unweighted (binary) one, and the former can be transformed into the latter by setting a threshold on the weight. In Section 2, we focus on the binary networks, and we first discuss how to choose a suitable threshold to obtain the binary network. We then explore the degree correlation (or associativity), betweenness distribution and motif structures of the two distinct networks from the two pseudoperiodic time series mentioned above. Finally, a framework to distinguish the different dynamical regimes of the time series is formed by exploiting the shortest path between nodes decided by the intrinsic dimension of the time series. In Section 3, we study the properties of the weighted networks in terms of subgraph coherence. The application of the proposed method to human ECG data and conclusions are made in Section 4 and Section 5, respectively.

2 BINARY NETWORK ANALYSIS

2.1 Graph representation of the complex network

We begin with a graphical representation of the binary network using the KK algorithm [9] (the choosing of the threshold is discussed in detail later in this section). This algorithm performs graph layout (in a two-dimensional plane) for undirected graphs. It operates by relating the layout of graphs to a dynamic spring system and minimizing the energy within that system. The strength of a spring between two vertices is inversely proportional to the square of the shortest distance (in graph terms) between those two vertices. Essentially, vertices that are closer in the graph-theoretic sense (i.e., by following edges) will have stronger springs and will therefore be placed closer together. The KK algorithm comes up with the geodesic distance between nodes given the binary adjacency matrix. For clarity of visualization and ease of reference in the following sections, we draw the networks for the Rössler system and the noisy periodic time series with only 60 nodes, see Fig. 4 and Fig. 6, respectively, using the software "Pajek". As can be clearly observed, these two networks demonstrate fundamentally different structures. In Fig. 4, the nodes lie on an elongated manifold, exhibiting a heterogeneous distribution. That is, the manifold has nodes congregating at different locations, so that some regions are highly clustered with nodes and others are rather sparse. In comparison, the network from the noisy periodic time series looks like a random network, with the nodes entangled with each other and their edges intersecting.

Before further analysis, we first give an intuitive description of the cycles-layout in phase space for chaotic system in terms of the unstable periodic

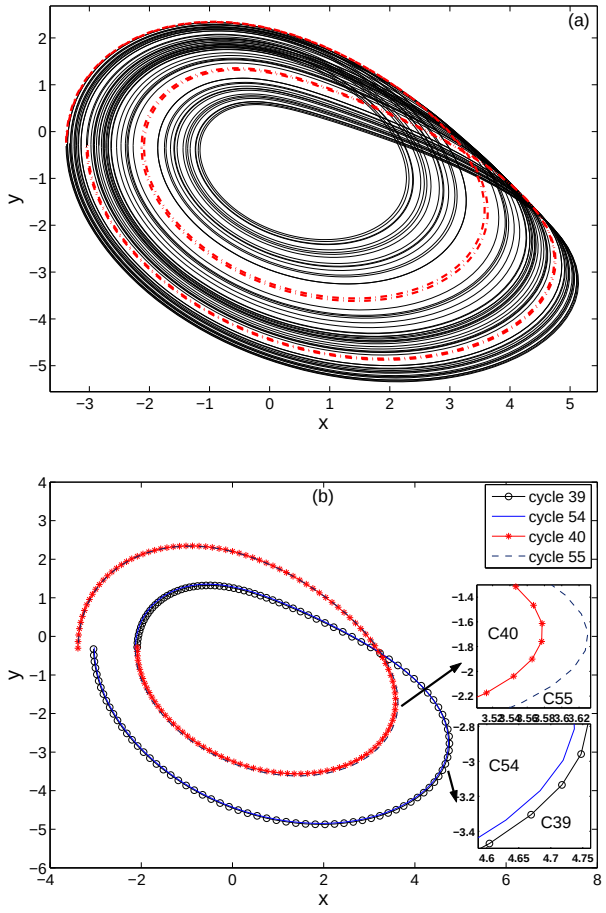

Fig. 3. Chaotic attractor of the Rössler system (a) the full attractor, with cycle 39, 40, 54, and 55 that lie on the sparse region of the attractor indicated by dash-dotted lines, which are further plotted out separately in (b).

orbits (UPOs). A chaotic attractor contains infinitely many UPOs, and for short data only the dominant UPOs (usually of low order) can be detected. Take the Rössler system for example, since the time series involves a strong periodic component, the dominant UPO-n (UPO of order n) will consist of n cycles. Each cycle that belongs to a certain UPO-n has many other cycles in its vicinity due to the attraction of the stable manifold associated with the UPO-n. It then becomes a center of a cluster and the number (or density) of the neighbors is related to its stability decided by the vector field along its trajectory. Since the stability of each center cycle may vary, we can see sparse as well as dense regions (see Fig. 3), where clusters are naturally formed in the latter. This is exactly captured in Fig. 4, where the "real" position of every cycle (node) is derived from the adjacency matrix via the KK algorithm. In Fig. 3 (b), cycles situated in the sparse regions are plotted out separately. These cycles also appear as nodes in the sparse regions of the corresponding network, see Fig. 4.

The above binary networks are achieved by setting a threshold to the original weighted network that is fully connected, where the weight W_{ij} represents the phase space distance between cycle i and cycle j. A pair of nodes whose distance is smaller than the threshold are considered as connected. For the study of the configuration of the cycles in phase space, the weighted network is fully connected and may contain redundant information. Therefore we first convert the weighted network to its binary counterpart.

An appropriate threshold therefore should be chosen to fully preserve the local clustering property of the network, but not to be too large that may obscure or conceal the local property by over connecting the nodes. Fig. 5 gives the density of the network (defined as the number of edges divided by the largest number of edges possible) versus the threshold d. As can be seen in Fig. 5(c), the increase of degree reaches the maximum rate at about $d_c = 0.08$ (which is only 3% of the network radius), where we set the threshold. We now explain why we set the threshold d at this critical point. It can be imagined that the degree increases more rapidly as the threshold changes within the cluster radius due to the adjacency of the nodes inside. For a network from a chaotic system that has many clusters differing in sizes, it can be further imagined that the edge increase will arrive at the maximum rate as the threshold approaches the critical point d_c, i.e., the "mean radius" of all the clusters. The binary network obtained at d_c will maintain the clustering property of the cycles, and thresholds beyond this value will result in a much slower edge increase, causing redundant connections among nodes. Consequently we choose d_c to derive the binary network. In fact we find that for networks derived at thresholds within a small range near d_c, the topology statistics assume qualitatively similar results. Finally it should be noted that the length of the time series, or, the size of the network, will influence d_c. We use a relatively large threshold $d_c = 0.3$ in Fig. 4 to preserve the clustering of the cycles in the relatively sparse phase space consisting of only 60 cycles, in contrast to $d_c = 0.08$ for 500 cycles shown in Fig. 5 .

For networks from noisy periodic data, as can be seen in Fig. 7, the weight follows a Gaussian distribution and the degree increase rate reaches the maximum at the mid-point of the weight distribution. This is obvious since the distance of any one cycle to the rest follows the same Gaussian distribution. We find that the nodes projected into a plane resemble a 2-D Gaussian distribution (see Fig. 6), which reflects the fact that the cycles are uncorrelated to one another. In fact, the decision of suitable threshold for the noisy periodic data may not appear as important as that for chaotic Rössler data. We have shown in [4] that binary networks derived at different thresholds are essentially random networks with different p (the probability that arbitrary two nodes are linked), which assumes a monotonic relationship with the threshold. In other words, the networks derived at different thresholds show no signs of clustering and take on similar random network characteristics. We use a

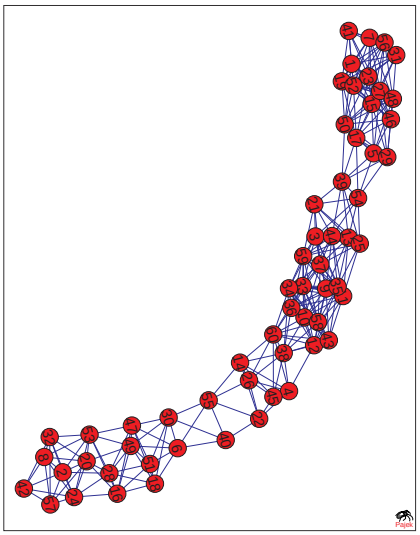

Fig. 4. Complex network of 60 nodes from the Rössler system with $d = 0.3$. The number attached to each node corresponds to the temporal sequence of the cycles.

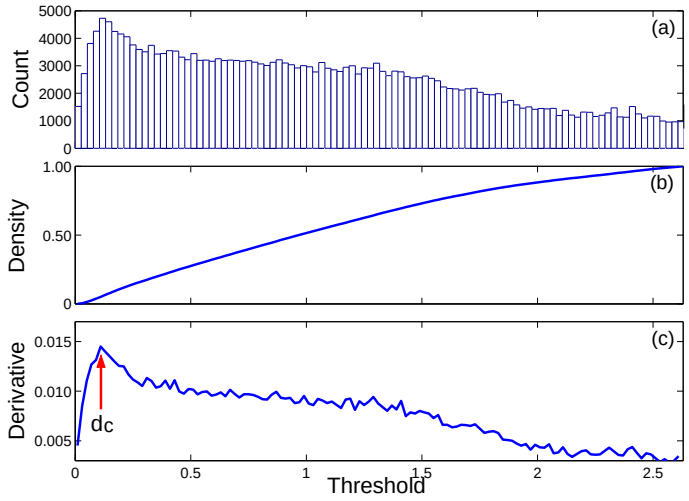

Fig. 5. Complex network from the Rössler system with 500 nodes. (a) Weight distribution (b) Density versus threshold (c) Derivative of density versus threshold, with $d_c = 0.08$.

smaller threshold 0.305 in Fig. 6 that brings on a low density network for ease of demonstration.

2.2 Degree-Degree Correlation

In our previous work [4], we showed that the degree distribution for chaotic systems demonstrate multiple peaks and that these reflect the various UPOs embedded in the chaotic attractor. For the noisy periodic data, the degree dis-

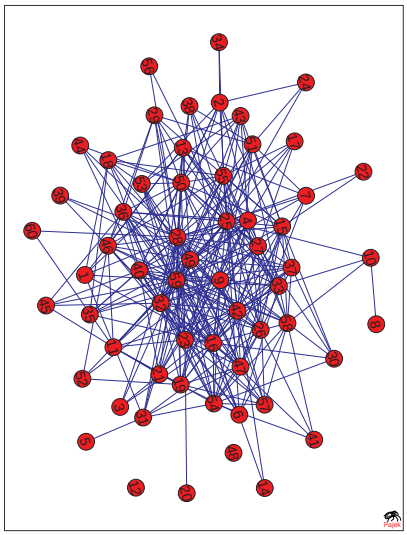

Fig. 6. Complex network of 60 nodes from the noisy sine signal with $d = 0.305$.

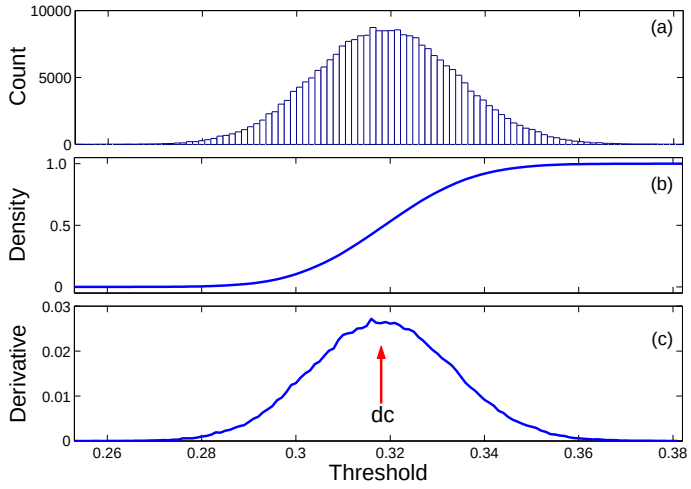

Fig. 7. Complex network from the noisy sine signal with 500 nodes. (a) Weight distribution (b) Density versus threshold (c) Derivative of density versus threshold, with $d_c = 0.3183$.

tribution is Gaussian. We now consider the joint degree distribution $P(k_1, k_2)$, which is a natural extension of the 1-D degree distribution. It gives the probability that a randomly selected edge has degrees of adjacent nodes equal to k_1 and k_2. $P(k_1, k_2)$ contains more information about the connectivity between nodes and is expected to capture the degree correlation, or alternatively, assortativity, among the nodes.

Many networks show assortative mixing on their degrees, i.e., a preference for high degree vertices to attach to other high-degree vertices, while others show dis-assortative mixing, i.e., high-degree vertices attach to low-degree ones [10]. Interestingly, we find that the network from the Rössler system also

demonstrates assortativity, due to the distinct distribution of cycles around the skeleton of UPOs specific to the chaotic nature of the system. For an appropriately chosen threshold, the clustering of the cycles is generally preserved, resulting in a complex network with multiple clusters. Note that these clusters are on a low dimensional manifold and are not overlapping, and most nodes connected to each other within the same cluster have a roughly similar number of connected neighbors or degrees (except for a few on the margin of the cluster). The common degree shared by the nodes from one cluster may differ from that of another because of the different stability of the center node (cycle). This has led to the fact that nodes with similar degrees are interconnected to each other, i.e., assortativity property, because they belong to the same cluster. Figure 8 depicts the overall assortativity trend, in which we find the grids near the main diagonal line have typically large values, indicating a high level of assortativity. The diffusion near the main diagonal corresponds to the cycles that lie on the edge of the clusters. The assortativity coefficient for this case is 0.7204.

Another way of capturing the degree correlation is to examine the average degree of neighbors of a node with degree k [11], which is defined as: $k_{nn} = \sum_{k'} k' P(k'|k)$, where $P(k'|k)$ denotes the conditional probability that an edge of degree k connects a node with degree k'. If this function is increasing, then the network is assortative, since it shows that nodes of high degree connect, on average, to nodes of high degree. Alternatively, if the function is decreasing, the network is dis-assortative. As can be seen in Fig. 9, k_{nn} increases with k for chaotic Rössler system, and we find this holds true for thresholds within a large range near d_c. For the network from the noisy periodic time series, it is interesting to note that the assortativity feature changes with the threshold, see Fig. 10. That is, no assortativity is observed at small thresholds, with the assortativity coefficient being zero. However, dis-assortativity arises as the threshold increases.

2.3 Betweenness Centrality

From the above analysis, we can see that the clustering property of the Rössler data can be well characterized by the degree correlation of the nodes (or by assortativity), and for networks from the noisy periodic signal, no assortativity is found at different thresholds. Assortativity provides the information about the interactions of all node pairs. Sometimes we may also want to know how important, or how central a single node is in a network. A concept that fulfills this requirement is betweenness, or betweenness centrality, which was first proposed in social network studies [12]. Take the actor network for example, the betweenness of an actor is an indicator of who are the most influential people in the network, the ones who control the flow of information between most

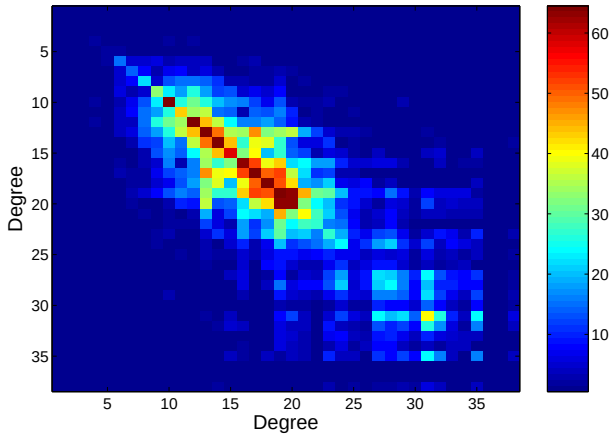

Fig. 8. Joint degree distribution for complex network from the Rössler system, with 500 cycles, $d = d_c = 0.08$. The shade of color corresponds to the number of node-pairs (see the color bar) that are connected.

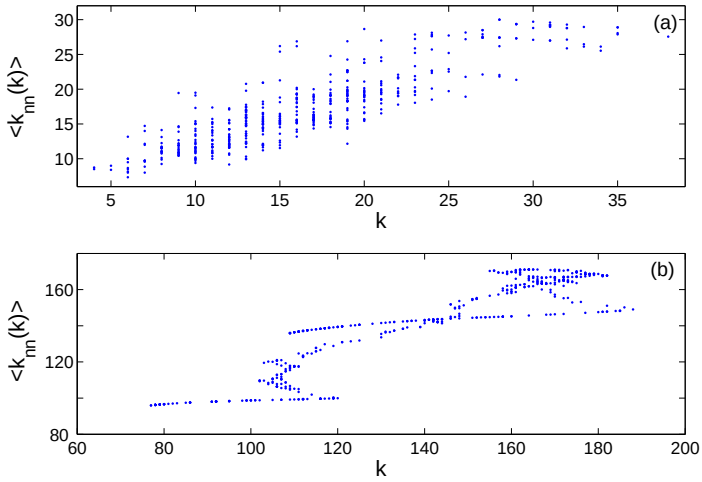

Fig. 9. Average degree of neighbors of k-nodes (nodes with degree k) for complex networks from the Rössler system with (a) $d = d_c = 0.08$ and (b) $d = 0.5$.

others. The vertices with highest betweenness also result in the largest increase in typical distance between others when they are removed. The betweenness centrality of a vertex v is defined as [12]:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3}$$

where σ_{st} denotes the number of shortest paths from node s to node t, and $\sigma_{st}(v)$ denotes the number of shortest paths from s to t that passes through v. A high centrality score thus indicates that a vertex can reach others on relatively short paths, or that a vertex lies on a considerable fraction of shortest

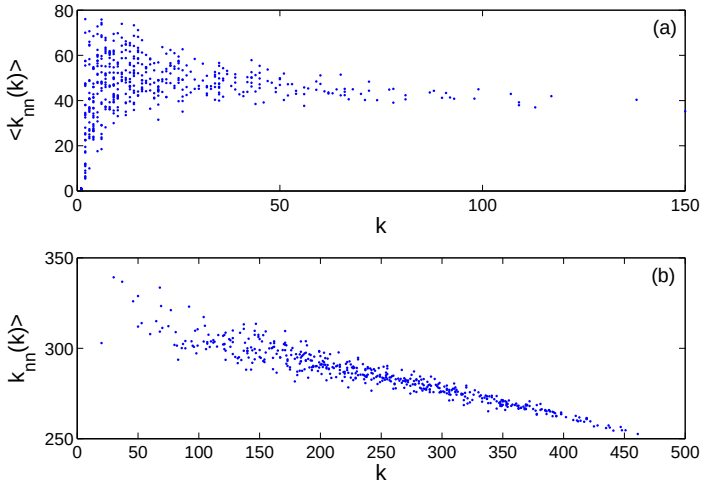

Fig. 10. Average degree of neighbors of k-nodes (nodes with degree k) for complex network from the noisy periodic data with (a) $d = 0.2930$ and (b) $d = d_c = 0.3183$.

paths connecting others.

For the network from chaotic time series, we find that the nodes that lie in the "sparse area" in phase space (e.g., nodes 39, 40, 54, and 55, see Fig. 3) have high betweenness centrality. Such nodes act as joints that connect adjacent clusters, and therefore play an central role in information transmission and have high C_B. The low C_B cycles are those lying on the marginal area of the manifold, like nodes $1, 7, 31, 41, 42, 46, 56$. We find that the overall betweenness from the chaotic Rössler data follows a power law distribution, in contrast to the exponential distribution for the noisy periodic data, see Fig. 11. The distinct distributions show that the cycles are structured with different mechanisms in phase space. The PDF of a power law type usually decreases much more slowly than an exponential one, which indicates that for the chaotic time series, the number of nodes with high betweenness much exceeds that from the periodic signal. This is essentially a reflection of the clustering property associated with the UPOs embedded in the chaotic attractor. The high betweenness nodes, as have been pointed out, correspond to cycles in between adjacent clusters that act as bridges. Since a chaotic attractor contains infinitely many UPOs, there will be numerous clusters in the corresponding network, such intermediate cycles will also be large in number.

The clustering property can also be quantified in terms of the number of motifs of different order. A motif [13,14] M is a pattern of interconnections occurring either in a undirected or in a directed graph G at a number that is significantly higher than in the randomized counterparts of the graph, i.e., in graphs with the same number of nodes, links and degree distribution as the original one, but where the links are distributed at random. As a pattern of interconnections, M is usually a connected graph with n-nodes, which is a subgraph of

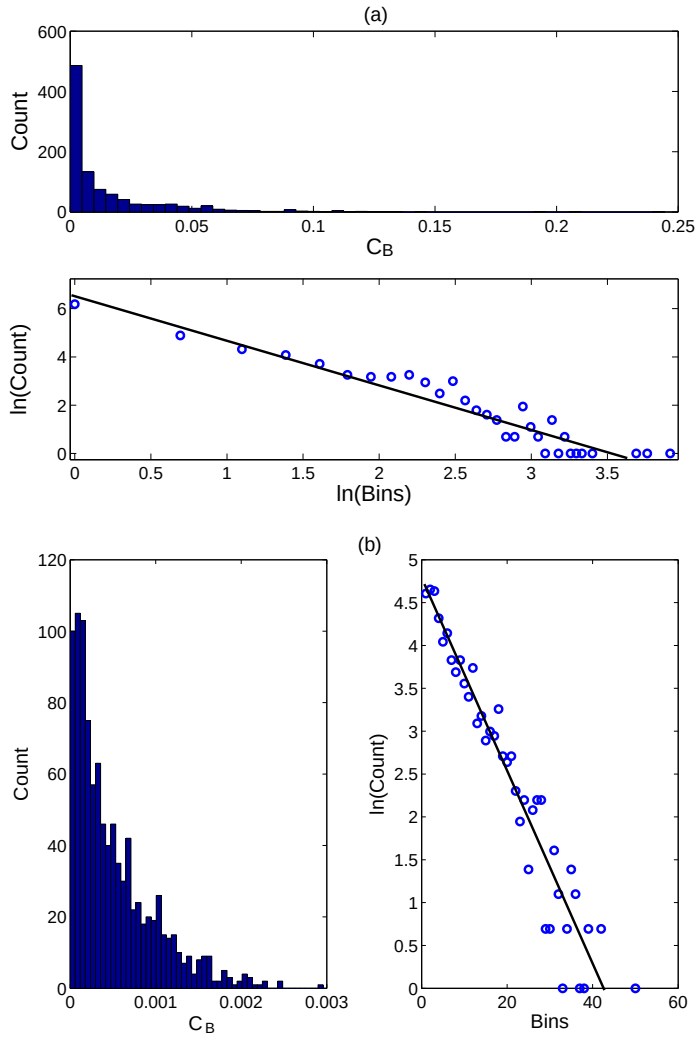

Fig. 11. Betweenness distribution for complex network of 1000 nodes from (a) the Rössler system, with $d = 0.04$. The original range $[0, 0.25]$ in the upper panel of (a) is rescaled to $[0, 50]$, i.e., 50 bins in the lower panel. (b) the noisy periodic data, with $d = d_c = 0.3183$. The original range $[0, 0.003]$ in the left panel of (b) is also rescaled to $[0, 50]$ in the right panel.

G. For example, triangles are among the simplest nontrivial motifs. For the network from chaotic time series which displays many clusters, the motifs (e.g, fully connected subgraphs of order n) will appear more frequently than in the random graph from the noisy periodic signal, where there are no obvious patterns. As is shown in Fig. 12, the number of fully connected subgraph-4s in the network from the chaotic system significantly exceeds that from the noisy periodic data, especially at low network density, where the former is almost four to six times larger than the latter.

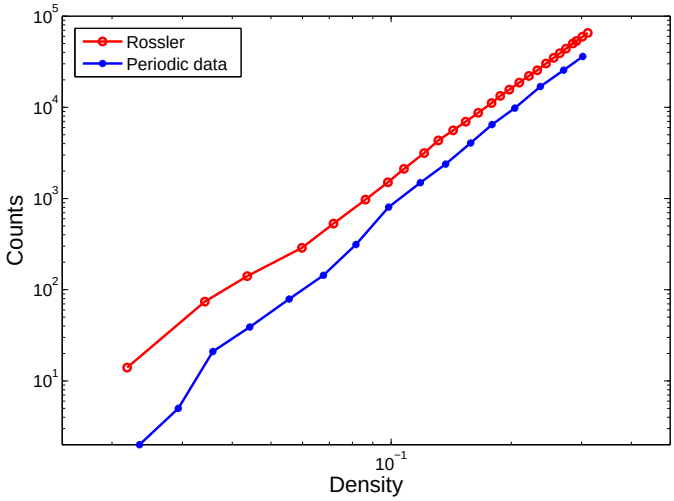

Fig. 12. Number of motif (subgraph-4 that has six edges, i.e., fully connected) for networks from the Rössler system and the noisy periodic data. The Binary networks are obtained at a series of thresholds, or correspondingly, network densities.

2.4 Phase space distance versus Shorest Path

Finally, we have developed a network-based framework to test if the underlying system is consistent with a low dimensional deterministic one. The elongated shape of the node distribution, obviously in Fig. 3, indicates that the cycles from the chaotic Rössler system lie on a low dimensional manifold that is smooth, so that the projection of the cycles onto the 2-D plane leads to a curve with certain thickness. This is due to the smoothness and continuity of the vector field of the underlying deterministic system. In comparison, the nodes from the network for the noisy periodic signal show totally random structure resembling the high dimensional case. Therefore the nodes will entangle with each other when projected to a plane, causing the edges to intersect. The basic idea of the method to distinguish these two types of networks is to check the relation between of the shortest path and the real phase space distance of all pairs of cycles. For low-dimensional deterministic systems (e.g., chaotic Rössler system), two cycles far away in phase space will also have a longer shortest path in the corresponding network space, i.e., it takes a longer path to connect cycles from one end to the other on the manifold, see Fig. 13 (a), where the shortest path grows almost linearly with the phase space distance. Here the phase space distance is computed for every two cycles which have the shortest path i. Both the mean value and standard deviation of the distance between two cycles with shortest path i are shown in this figure.

For noisy periodic data, the shortest path assumes similar values for most pairs of cycles because of the interwoven nature of the nodes when projected from the originally high to the current lower dimensions, see Fig. 13(b). The

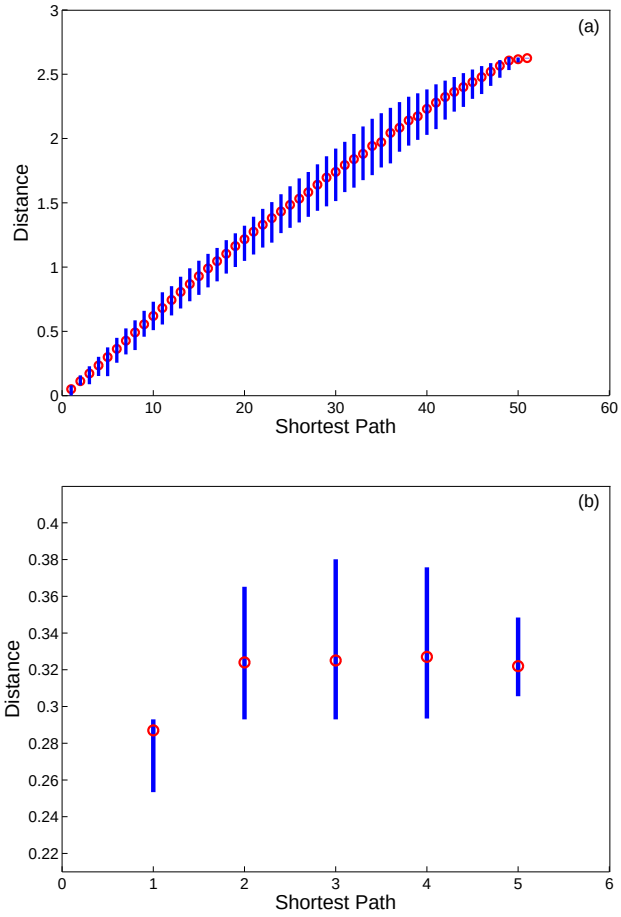

Fig. 13. Shortest path versus phase space distance for complex network from (a) the Rössler system, with $d = d_c = 0.08$ and (b) the noisy periodic data, with $d = 0.293$. Both networks have 500 nodes. Here we do not use d_c for the noisy periodic data because we want to derive a binary network with comparable density to that from the Rössler system, so that the shortest path from the two networks can be compared.

exception at shortest path equal one arises from the trivial correlation among cycles associated with the mutual sinusoidal waveform. In addition, the intrinsic high dimensional layout of the cycles in the noisy periodic data has rendered the average path length quite small (about 2.41). For the network from a Rössler system of comparable density (about 3%) the average path length is much larger (18.01).

3 WEIGHTED NETWORK ANALYSIS

Up to now we have focused only on the binary network derived with an appropriate threshold from the original weighted network. In this section, we

treat the network as a fully connected one with the weight between two nodes being the phase space distance between corresponding cycles. In particular, we consider the subgraph intensity and coherence of the weighted network.

Onnela et al. have recently proposed a systematic way to extend motifs to a weighted graph $G_W = (N, L, W)$ [15]. A motif M is defined as a set of the topological equivalent subgraphs of G_W. The intensity $I(g)$ and the coherence $Q(g)$ of a given subgraph $g = (N_g, L_g, W_g)$ with n nodes and l links are defined as:

$$I(g) = W^{\frac{1}{|l_g|}}, Q(g) = I(g) \cdot |l_g|/W \tag{4}$$

respectively, where $W = \prod_{(i,j)\in l_g} w_{ij}$, and l_g is the links in subgraph g with $|l_g|$ being the number of links. We note that Q for noisy periodic data is distributed in a very narrow range close to unity (see Fig. 14 (a)), this is because the weights do not differ much (which is obvious since the weight follows a Gaussian distribution and 95% of the samples are within the range of $[\mu - 2\sigma, \mu + 2\sigma]$), and that the weighted network is internally very coherent. Moreover, the subgraph coherence is found to follow an exponential distribution. In comparison, the weighted network from the Rössler system is rather non-coherent, with Q distributed within $[0.2, 1]$, see Fig. 14 (b). This reflects the non-coherence of the cycles distributed in the chaotic attractor, which are typically grouped into clusters.

4 APPLICATION TO ECG DATA

In order to show the effectiveness of the network-based statistics, in this section we apply them to the human electrocardiogram (ECG), which is one of the most prominent clinical tools to monitor the activity of the heart. The ECGs here are recorded from a healthy volunteer and an arrythmia patient, and the time series are typically pseudoperiodic, due to the continuous generation of action potentials and the fixed pattern according to which the electrical activity spreads out over the cardiac muscle. The two ECG time series we choose seem to be morphologically similar, but are noisy and nonstationary, see Fig. 15 and Fig. 16.

Figure 17 gives the network representation for the two ECGs. As can be seen, the two networks take on significantly different structures. The network from the healthy subject assumes a elongated shape, in which the nodes distribute on a low dimensional manifold, forming clusters over different regions. In comparison, the network from the patient seems to be a random network. Though the distribution of the betweenness are similar (both power-law distribution,

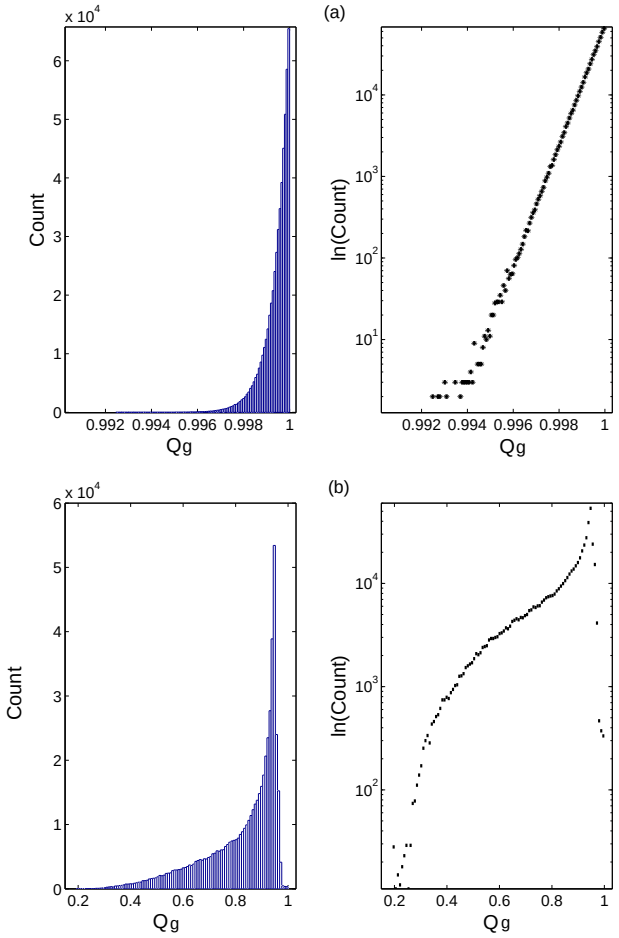

Fig. 14. Distribution of subgraph coherence (triangle) for network from (a) the noisy periodic data (b) Rössler system. The right panel shows the same distribution as in the left panel but is in *log* scale.

see Fig. 18), the network betweenness centrality does show some difference (0.124 for the healthy and 0.049 for the patient). Moreover, the two networks also display different extent of assortativity. This is clearly illustrated in Fig. 19, where the network from the healthy subject demonstrates obvious assortativity, with the corresponding assortativity coefficient being 0.674, much larger than that of the arrythmia patient (0.208).

Now we compare our method to the traditional phase space embedding based techniques, in particular, the correlation dimension, which is widely used to characterize the geometry of the attractor in phase space. We find that the correlation dimensions are 1.903 and 1.845 for the ECGs of the arrythmia patient and the healthy subject after appropriate phase space embedding, which do not seem to demonstrate significant difference. Though phase space embedding has been proved to be applicable for ECG signal [16–19], traditional methods based on phase space reconstruction are always seriously hampered by noise corruption. The ECG data shown in Fig. 15 are typically very noisy,

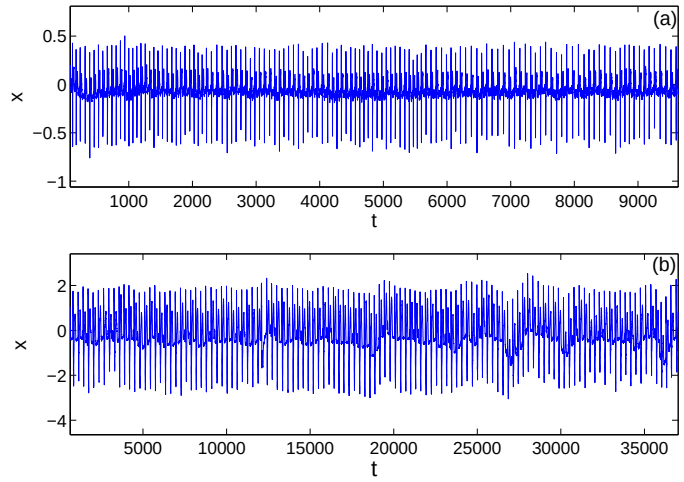

Fig. 15. Time series from ECGs of (a) Healthy subject (female, age 18) and (b) Arrythmia patient, ECG 107 from MIT arrythmia in Physionet.

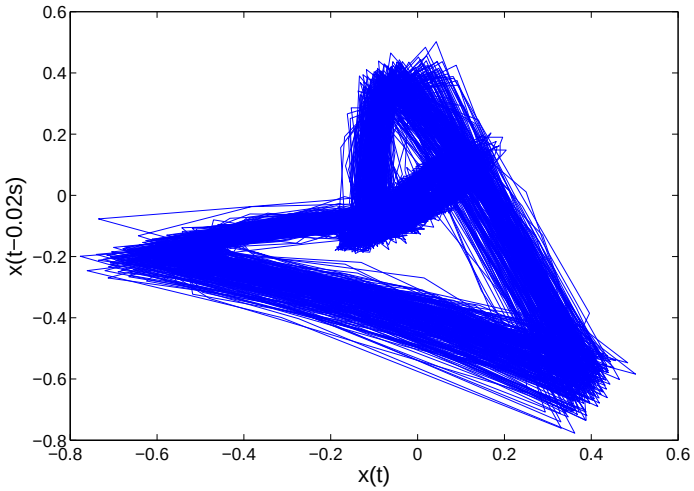

Fig. 16. Phase space reconstruction of the ECG from the healthy volunteer.

and the corresponding attractors in phase space seem to be dominated by the noise component that smears the fine structure of the ECG data (see Fig. 16). From this comparison we can see that the network representation of the ECG on the cycle scale helps to reveal the essential difference between the data structures of the healthy subject and the patient, which tends to be masked by the noise and is invisible to conventional methods that calculate on the point scale. Application of the proposed network transformation to larger and different datasets, however, is needed to test its performance more objectively.

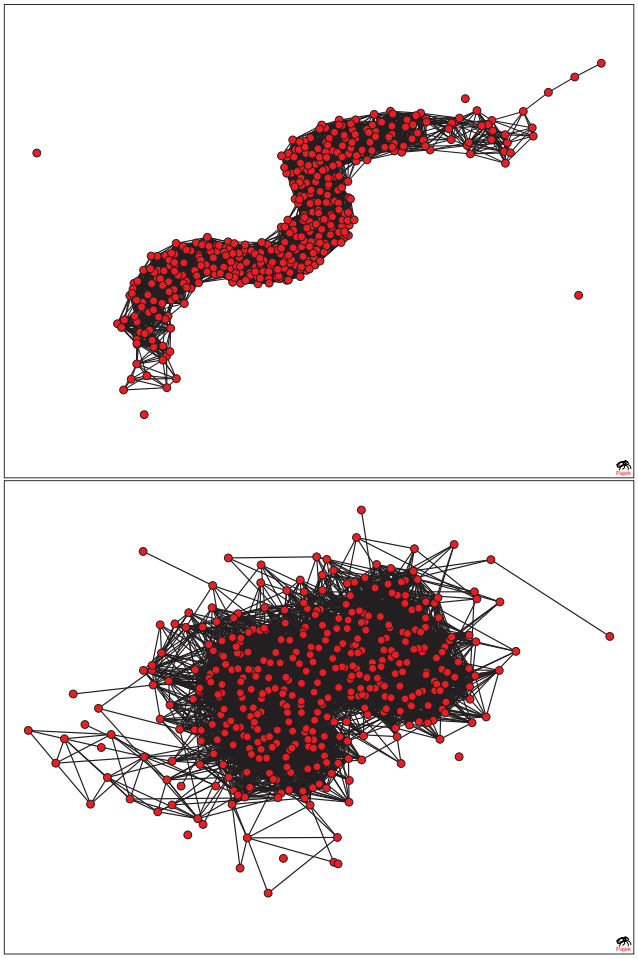

Fig. 17. Complex network constructed form ECG of a healthy subject (the upper panel), with the threshold $\rho_{th} = 0.87$ and an arrythmia patient (the lower pannel), with $\rho_{th} = 0.969$. The thresholds are set such that both networks (with 400 nodes) have the same number of edges.

5 DISCUSSION

The pseudoperiodic signals are naturally simplified by treating each cycle, rather than each sampling point, as the basic unit. Another technique that has similar effect is the Poincare section method. The Poincare section is the intersection of a flow data in the state space with a certain lower dimensional subspace, or hyperplane, which is transversal to the flow of the system. The highly sampled data representing the continuous time of a dynamical system are reduced to a new vector series, which each vector representing the corresponding cycle in the pseudoperiodic time series. Usually, the placement of the Poincare surface is of high relevance for the usefulness of the result. Our construction of the complex network relies on the segmentation of the pseudoperiodic time series, and this is equivalent to choosing the location of the Poincare section.

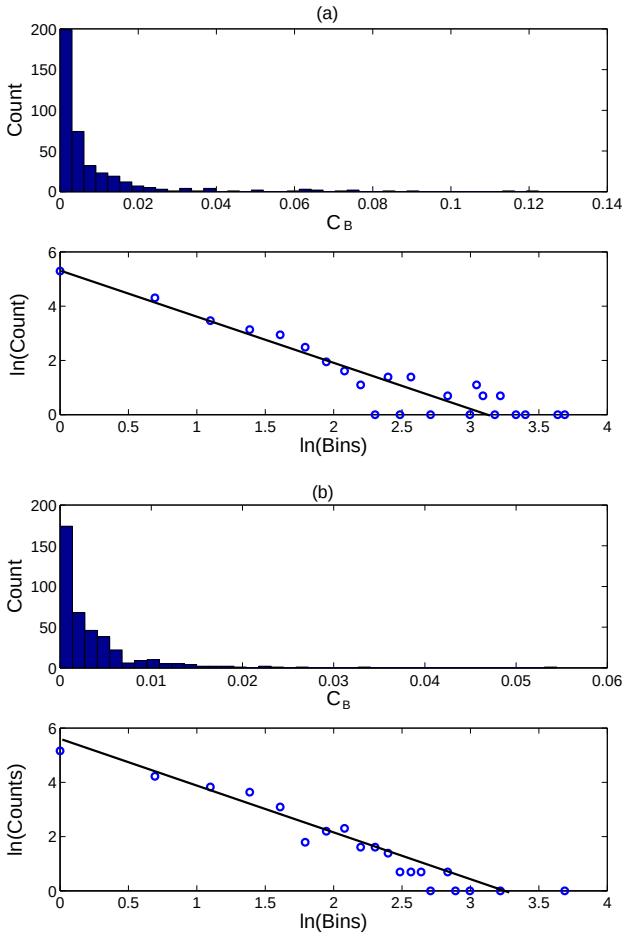

Fig. 18. Betweenness distribution for the network from (a) Healthy subject and (b)Arrythmia patient. Similar to Fig. 11, both the lower panels of (a) and (b) rescale the original ranges to [0, 50], i.e., 50 bins, and plot in a *log-log* scale.

Both methods depend on an appropriate segmentation of the time series, if one defines a Poincare surface as the zero crossings of the temporal derivative of the signal, then this is synonymous with collecting all maxima or all minima, which is the procedure we have used to divide the time series into cycles. Nevertheless, these two methods reflect different sources of information from the underlying system and possess different level of robustness. The discrete points on Poincare section only specify the dynamics of the map at that section and the information regarding the trajectories that connect the Poincare section points are inevitably lost, and such dynamics are liable to change with the section and are vulnerable to noise. Comparatively, we utilize all the points associated with a cycle and characterize the correlation among the cycles. Thus not only the full information of the trajectory is kept, but the method becomes more robust to noise. Although our method also depends on the segmentation scheme, the correlation structures of the networks derived from different segmentations are essentially very similar.

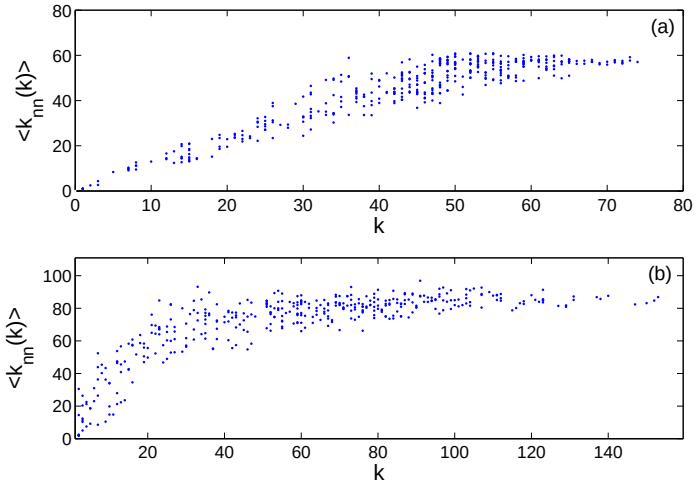

Fig. 19. Average degree of neighbors of k-nodes (nodes with degree k) for complex network from (a) Healthy subject. The ascending trend spaning over almost the entire range indicates that nodes with high degree tend to connect to those with high degrees, i.e, strong assortativity. (b) Arrythmia patient. The uprising trend is only obvious for $k < 30$, and the network shows week assortativity.

The connection of our method to the Poincare section reminds one of the symbolic dynamics analysis, by which the Rössler system can also be analyzed. It embeds as a one hump map that gives rise to a full-shift on two symbols. Though it would be interesting to associate our framework with the traditional symbolic dynamics, this is not the focus of this paper. Indeed we note that a large variety of real-world signals are too noisy and nonstationary to be effectively explored and characterized via the traditional methods as have been previously described. The motivation of this paper therefore is to reliably characterize the correlation structure among the cycles first, before attempting to understand the mathematical origin of the dynamics.

In summary, we have explored an extensive set of topological statistics for the networks constructed from two pseudoperiodic time series with distinct dynamics, i.e., the deterministic chaotic Rössler time series and the stochastic noisy periodic data. The essential features of these two kinds of time series are that the cycles from the former lie on a low dimensional manifold, with their correlation manifested in the strong local clustering, while the cycles from the latter are randomly correlated and cannot be effectively represented in a low dimensional space. We find that the joint degree distribution appears to fundamentally characterize the clustering property of the cycles induced by the UPOs intrinsic to the chaotic attractor. The extent of assortativity, as a single-value quantification of joint degree distribution, appears to be efficient in distinguishing different sorts of networks derived from various time series. In addition, the betweenness centrality and the motif also achieve different extent of success in discriminating different dynamical regimes in the

pseudoperiodic time series. We demonstrate that the topological statistics discussed in this paper combine to provide an all-around characterization on how cycles are structured in phase space, and furthermore a comprehensive quantification of the dynamics of the pseudoperiodic time series. Application to the human electrocardiograms shows that this alternative approach is effective in differentiating healthy subject from the arrythmia patient.

ACKNOWLEDGEMENT

This research is funded by the Hong Kong Polytechnic University Postdoctoral Fellowships Scheme 2007-08 (G-YX0N).

References

[1] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. 74 (2002) 47-97.

[2] M. E. J. Newman, SIAM Review 45(2) (2003) 167-256.

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Physics Reports 424 (2006) 175-308.

[4] J. Zhang and M. Small, Phys. Rev. Lett. 96, 238701 (2006).

[5] J. Zhang, X. Luo and M. Small, Phys. Rev. E 73 (2006) 016216 .

[6] J. Zhang, X. Luo, N, Tomomichi, J. Sun and M. Small, Phys. Rev. E 73 (2006) 016218.

[7] M. Small, D.J. Yu, R.G. Harrison, Phys. Rev. Lett. 87 (2001) 188101.

[8] P. P. Kanjilal, J. Bhattacharya, G. Saha, Phys. Rev. E 59 (1999) 4013.

[9] T. Kamada, S.Kawai, Information Processing Letters 31(1) (1989) 7-15.

[10] M. E. J. Newman, Phys. Rev. Lett. 89 (2002) 208701.

[11] R. Pastor-Satorras, A. Vazquez, A. Vespignani, Phys. Rev. Lett. 87 (2001) 258701.

[12] L. C. Freeman, Sociometry 40 (1977) 35-41.

[13] S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Nature Gen. 31 (2002) 64-68.

[14] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashan, D. Chklovskii, U. Alon, Science 298 (2002) 824.

[15] O. Jukka-Pekka, S. Jari, K. Janos, K. Kimmo, K. Kimmo, Phys. Rev. E 71 (2005) 065103(R).

[16] M. Richter, T. Schreiber, Phys. Rev. E 58 (1998) 6392-6398.

[17] M. Small, D.J. Yu, R.G. Harrison, C. Robertson, G. Clegg, M. Holzer and F. Sterz, Chaos, 10 (2000): 268-277.

[18] M. Small, D. J. Yu, R. Clayton, T. Eftestol, K. Sunde, P. A. Steen and R.G. Harrison, International Journal of Bifurcations and Chaos, 11 (2001): 2531-2548.

[19] M. Small, D. J. Yu, J. Simonotto, R. G. Harrison, N. Grubb and K. A. A. Fox, Chaos, Solitons and Fractals, 13 (2002): 1755-1762.