2008 International Symposium on Nonlinear Theory and its Applications
NOLTA'08, Budapest, Hungary, September 7-10, 2008

Time series classification by complex network transformation

J. Zhang†, J. Sun, X. Xu and M. Small

Department of Electronic and Information Engineering,
Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
Email: †enzhangjie@eie.polyu,edu.hk

Abstract—The nonlinear transformation from time domain to complex network domain recently introduced for pseudoperiodic time series has been shown to be a powerful tool in characterizing the complex dynamics of time series via the organization of the corresponding complex networks. In this paper, we test an extensive range of network topological statistics for time series from two archetypal systems and show that they are capable of providing a comprehensive statistical characterization of the dynamics from different aspects, and can be used to distinguish different dynamical regimes. Application of such network statistics to the human electrocardiograms reveals significant difference between the healthy individual and arrhythmia patients.

1. Introduction

The past few years have witnessed great advances in the field of complex networks, which provides a new paradigm and profound insights into many of complicated systems in engineering, social and biological fields [1, 2, 3]. We propose a network approach to pseudoperiodic time series analysis recently 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, the dynamics of the time series can then be explored 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. 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.

2. Chaotic vs. stochastic time series

We use the chaotic time series from the x component of the well know Rössler system given by:

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

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.

We first segment the pseudoperiodic time series into m consecutive cycles according to the local minimum (or maximum), denoted as $\{C_1, C_2, ..., C_m\}$. For each pair of cycles C_i and C_j ($i, j = 1, 2, ..., m, i \neq j$) with length l_i and l_j, respectively, 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.

We begin with a graphical representation of the binary network using the KK algorithm [6] (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. 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. As can be clearly observed in Fig.1, the networks from the two time series demonstrate fundamentally different structures. In the network from chaotic data, 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. A chaotic attractor contains infinitely many UPOs, 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.

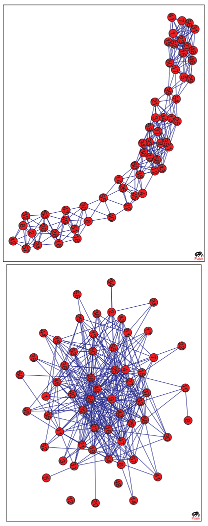

Figure 1: Complex network from (a) x component of Rössler system (upper panel) (b) noisy sine signal (lower panel).

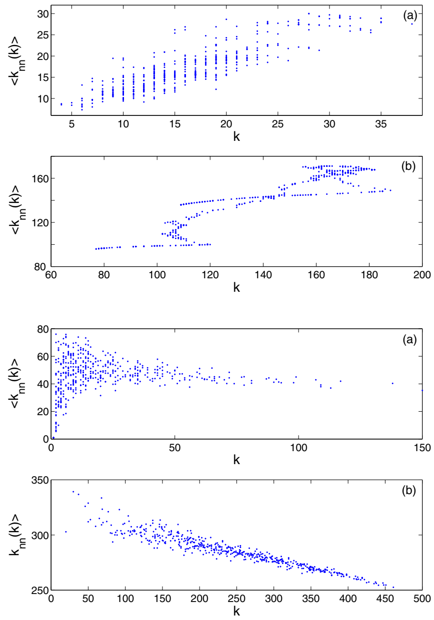

Figure 2: Assortativity property for networks from (a) x component of Rössler system (b) noisy sine signal.

Degree distribution is a basic statistic of the network. We 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 [7]. Interestingly, we find that the network from the Rössler system also demonstrates assortativity. Note that the clusters in Fig. 1 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. 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. This has led to the fact that nodes with similar degrees are interconnected to each other, i.e., assortativity property.

One way of capturing the degree correlation is to examine the average degree of neighbors of a node with degree k, 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. From Fig. 2, we can see that the network from chaotic time series shows a high degree of assortativity in sharp contrast to that from the noisy sine signal, which has no assortativity or disassortativity property.

Now we check the betweenness centrality, which was first proposed in social network studies [8]. 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 others. 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. 3. 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 [9] 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. 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

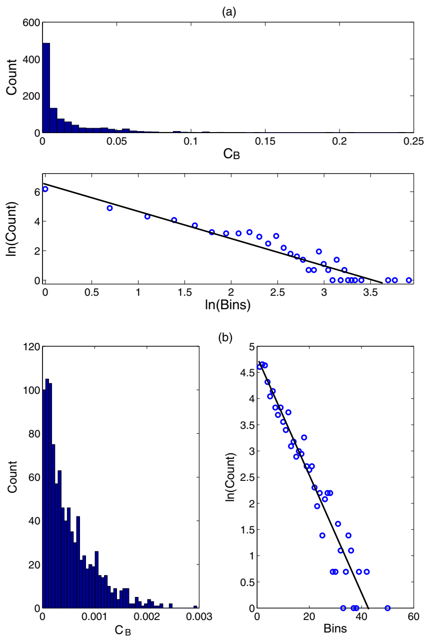

Figure 3: Between centrality for complex networks from (a) x component of Rössler system (b) noisy sine signal.

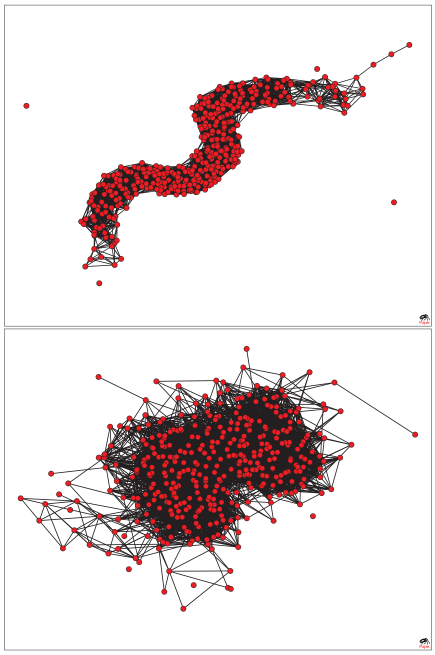

Figure 5: Complex networks from a healthy subject (upper panel) and an arrythmia patient (lower panel).

are no obvious patterns. As is shown in Fig. 4, 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.

3. Application to human electrocardiograms

In this section, we calculate the network statistics for time series from various ECGs. Figure 5 gives the network representation for the two ECGs. As can be seen, the two networks take on significantly different structures.

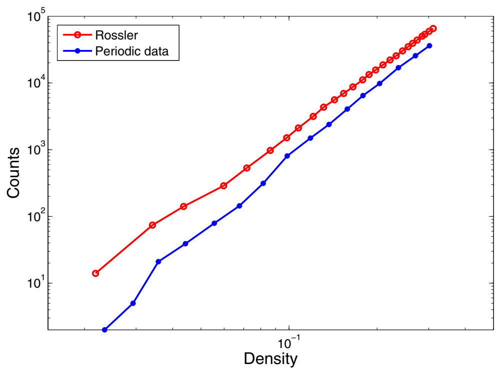

Figure 4: Motif numbers for complex networks from x component of Rössler system and noisy sine signal.

The network from the healthy subject assumes a elongated shape. 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, 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. 6, 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).

4. Conclusion

We have tested an extensive range of statistics of the networks constructed from archetypal and real-world time series. We finds that these statistics can provide useful characterization of the dynamics of the data from different perspectives. Application to human ECG data shows that the method can effectively distinguish healthy and arrythmia patients.

Acknowledgments

This research was funded by a Hong Kong University Grants Council Grant Competitive Earmarked Research Grant (CERG) number PolyU 5235/03E.

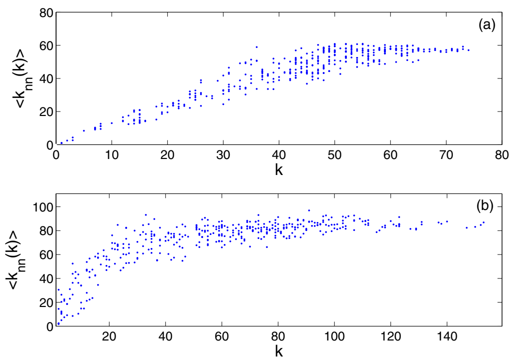

Figure 6: assortativity for complex networks from a healthy subject and an arrythmia patient.

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

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

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

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