Transforming Time Series into Complex Networks

Michael Small, Jie Zhang, and Xiaoke Xu

Department of Electronic and Information Engineering
Hong Kong Polytechnic University
ensmall@polyu.edu.hk
http://small.eie.polyu.edu.hk/

Abstract. We introduce transformations from time series data to the domain of complex networks which allow us to characterise the dynamics underlying the time series in terms of topological features of the complex network. We show that specific types of dynamics can be characterised by a specific prevalence in the complex network motifs. For example, low-dimensional chaotic flows with one positive Lyapunov exponent form a single family while noisy non-chaotic dynamics and hyper-chaos are both distinct. We find that the same phenomena is also true for discrete map-like data. These algorithms provide a new way of studying chaotic time series and equip us with a wide range of statistical measures previously not available in the field of nonlinear time series analysis.

Keywords: nonlinear time series, chaos, chaotic dynamics, complex networks.

1 Turning Time Series into Networks

The simplest method to transform a time series into a complex network is through the well established recurrence plot [2]. Recurrence techniques provide a recognised method for constructing a sparse binary matrix from a time series. That matrix may be (although to the best of our knowledge it never has been) interpreted as the adjacency matrix of a complex network, and one may then study that network to get an insight into features of the dynamics not apparent from the time series. Many measures have been associated with recurrence plots. Recurrence Quantification Analysis (RQA) introduces a host of such measures based on identifiable patterns within the recurrence plots [13,6]. Moreover, Theil and co-workers [11] have showed how the recurrence plot can be considered as a surrogate for the correlation integral and it can then be used to estimate versions of the usual dynamic invariants. However, RQA and the other measures derived from the recurrence matrix all treat properties of the matrix rather than properties of the network. For example, the temporal ordering of the rows and columns of the recurrence matrix are important, for an adjacency matrix the result is invariant under permutations of the rows and columns.

In this communication we do not consider recurrence plots. We consider an alternative method of constructing a complex network from a time series. The

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 2078–2089, 2009.

features we wish to examine in that complex network are features of the network rather than features of the matrix. That is, we are concerned primarily with path length and clustering of points within a network (and the prevalence of various motifs [8] within the network). The way in which we obtain the adjacency matrix of the network from the time series also differs from the approach used in recurrence plots.

In the remainder of this paper we consider two distinct approaches to the construction of complex networks from time series. The first scheme was introduced in 2006 [16] and provides a method to construct complex networks from pseudo-periodic time series. The second method is currently under development and provides a generic method to construct complex networks from any time series data [14]. In the following two sections we outline these methods.

2 Networks from Pseudo-periodic Time Series

In [16] we introduce a method to construct complex networks from time series data. The method assumes that the time series is approximately periodic and takes as the basic unit of the time series a single oscillation of that periodicity. Pseudo-periodic time series have been considered previously [10] and the same basic definition is the one we adopt here. Pseudo-periodic time series exhibit some oscillation, and that oscillation is rhythmically repeating. The object of interest is the inter-cycle variation in that rhythmic oscillation. Is the underlying system low dimensional chaos, or a periodic orbit [10]?

Networks are constructed from data such as these in the following way. First the time series is divided up into individual cycles and each cycle is then treated as a node on a network. Exactly how to divide the data into cycles is not stipulated. In fact, the method will depend on the data. Nonetheless, the objective is to deconstruct the time series in such a way that the individual cycles can be meaningfully compared to one another. Next, some metric and a suitable threshold are chosen. While the choice of metric and threshold will have some affect on the results we have found that the results are robust across the usual range of metrics and a wide range of threshold values. All cycles are compared with this metric, and those found to be closer than the threshold are said to be neighbours. Finally, we construct links on the network corresponding to cycles which are neighbours. That is, two nodes are linked if the corresponding two cycles are close to one another under the chosen metric.

In [16] we do this in the obvious way. First, cycles are split at local maxima. Second, cycles are compared with either linear correlation or Euclidean distance (after sliding the shorter cycle along the larger and finding a minimum That is, for two cycles $C_i = (x_1, x_2, x_3, \ldots, x_{n_i})$ and $C_j = (y_1, y_2, y_3, \ldots, y_{n_j})$ (with, without loss of generality $n_i \le n_j$) the distance between them is defined as

$$d(C_i, C_j) = \min_{1 \le i \le (n_j - n_i)} \phi\left((x_1, x_2, \ldots, x_{n_i}), (y_{i+1}, y_{i+2}, \ldots, y_{i+n_i})\right) \qquad (1)$$

where $\phi(\cdot, \cdot)$ is some measure of distance in $\mathbf{R}^{n_i}$ — typically either Euclidean distance of linear correlation. This scheme is depicted schematically in Fig. 1.

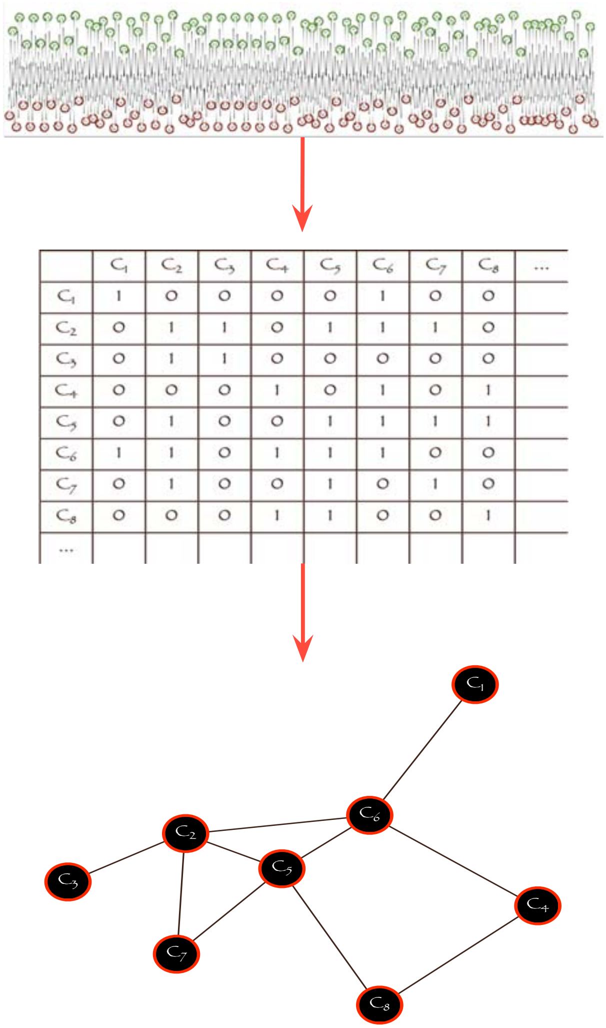


Fig. 1. This cartoon depicts the scheme utilised in [16] to construct a complex network from a time series. The pseudo-periodic time series is first divided into cycles and the distance between these cycles is measured and compared to some threshold. This yields a matrix which is treated as an adjacency matrix to construct a network.

Using linear correlation between cycles as the measure of closeness has several advantages when treating real data. In particular, the effect of (stationary) additive noise is minimised, and one can avoid the often difficult task of successfully embedding the data. Of course, for particular applications one may choose different measures of closeness. Moreover, the choice of measure may depend on the application one is considering. However, for the general problem of analysis of the attractor reconstructed from a time series, the measure we have chosen is appropriate. Using this algorithm we have constructed complex networks from various time series and have examined the gross measures of network structure: degree distribution and vertex strength [16]. We found that this simple measure allows one to differentiate between noisy periodic orbits and chaotic dynamics. In particular we observed peaks in the degree distribution of the complex network corresponding to the unstable periodic orbits of the underlying system. Moreover, when applied to experimental and clinically obtained Electrocardiograms (ECGs) we found fundamental differences in the structure of sinus rhythm ECG of healthy volunteers and of coronary care patients.

Recently, the results of [16] have been more thoroughly studied by Zhang and colleagues [17]. In this work, the network transform introduced by Zhang and Small [16] is exploited and the standard battery of network-based statistics are applied: degree, degree correlation, betweenness centrality and path lengths. In [16] we observed UPOs (Unstable Periodic Orbits) in the degree distribution. In [17] we go further and find that the joint degree distribution characterises the organisation of cycles in phase space, and of course the assortativity coefficient provides a succinct measure of this feature. Hence, the chaotic Rössler system is assortative (with a highly structured joint degree distribution) while noisy periodic signals are either uncorrelated or disassortative.

The technique of Zhang and Small has recently been extended by Yang and Yang [15] to the case of time series without obvious period. In the method of Yang and Yang all windows of a fixed length L along a time series are considered as nodes and links between them are drawn if the magnitude of the correlation coefficient exceeds some threshold. Of course, this is equivalent to performing an L-dimensional embedding with time lag 1 and allows one to construct networks from arbitrary time series. A further trivial generalisation of this method would be to allow an arbitrary embedding. Certainly, for particular types of dynamics a larger time delay may be preferable.

In [4], Lacasa and co-worker introduce another technique for constructing networks from time series. Unlike the method of Zhang and Small [16] this method does not require that the time series is pseudo-periodic. Rather, [4] maps each scalar time series point to points on a complex network. The nodes corresponding to two points (t_a, y_a) and (t_b, y_b) are then said to be connected if for all intermediate third points (t_c, y_c) with $t_a < t_c < t_b$ we have that

$$\frac{y_c - y_b}{t_b - t_c} < \frac{y_b - y_a}{t_b - t_a}. \tag{2}$$

This method is simple, and the test easy to perform. But, it is unclear (at least, it is unclear to the current authors) what feature of the dynamics is being measured by this convexity constraint. This "visibility" constraint is motivated by considering the points on the time series as mountain or building peaks. The aim of the criterion is to identify which peaks are visible from the current peak [3]. The criterion for this is (2). Nonetheless, Lacasa and colleagues apply this criterion and are able to distinguish between broad classes of dynamical systems. In particular, this constructions maps periodic signals to regular graphs and random signals to random graphs. Scale-free networks are obtained from fractal time series. We note that the emergence of scale-free networks from fractal time series is a trivial consequence of the choice of (2) and self-similarity.

3 Networks from Embedded Data

The main strength of the method introduced by Zhang and Small [16] is that it is fairly robust to noise and does not require embedding. The main weakness of that method is that it will only work provided one has some way of comparing orbits. In some instances (hyper-chaos, for example) it is not clear how to achieve this. In this section we describe a generic alternative [14].

The first step of this method is to embed the data in some suitable phase space. Problems of choosing embedding dimension, embedding lag, and selecting an appropriate embedding are not addressed here. Nonetheless, Fig. 2 depicts

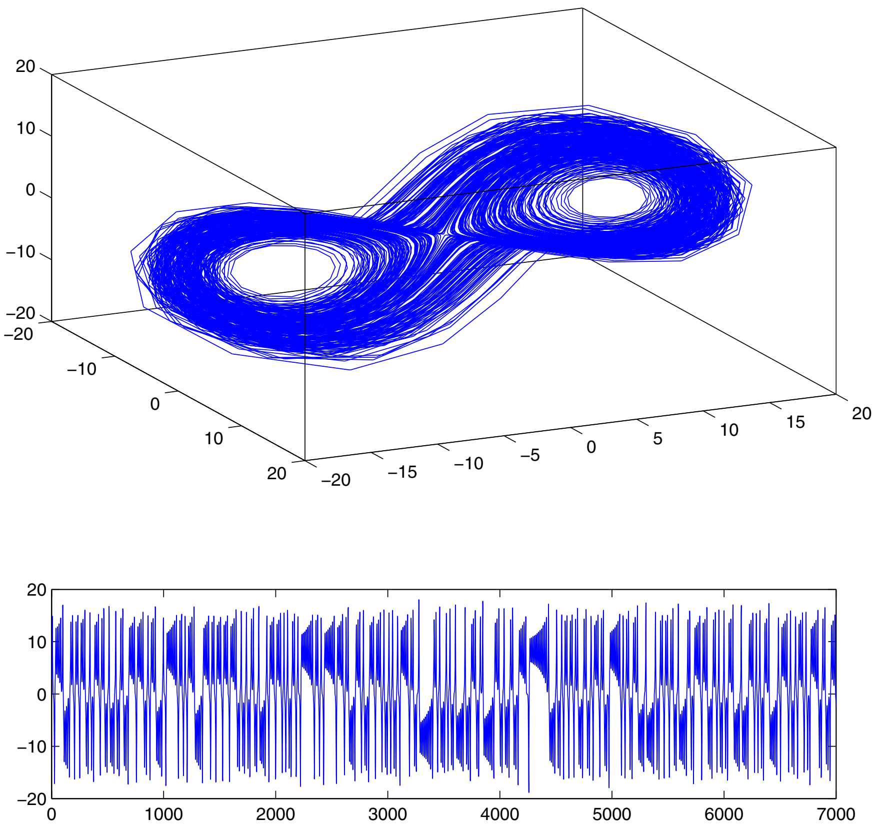


Fig. 2. The lower panel depicts the x-component of the Lorenz system and the upper panel a time delay embedding $d_e = 3$, $\tau = 2$ of that data. The embedded phase space points are used to generate the complex network in Fig. 3.

the result of this step for one archetypal model of chaos. Assuming a suitable embedding can be found, each embedded phase space point represents a node on the complex network. For some fixed k link each point x_i with its k nearest neighbours. For instances where two points are closest to one another, the same link is added only once (we not consider either directionality of multiplicity of links). In the event that two nodes are mutually closest to one another (that is, the same link could've been added twice), the next closest link to either node is added (one more link is added). Hence each node will contribute *on average* k links to the network. The mean degree of the network will be exactly $2k$, but some nodes will certainly be more highly connected than others.

Finally, in order to ensure that neighbours are spatial neighbours rather than merely temporal successors, we exclude points on the same strand of the trajectory from being neighbours. That is, for any i the neighbours of x_i are the points x_j such that $|j-i|>T$ and $\|x_i-x_j\|$ is one of the k smallest observed distances (T is chosen to be one quarter of the recurrence time).

The main strength of this procedure is that the transformation ignores the relative sparsity or density of points in phase space: each point will generate the same number of neighbours and correspond to a node with the same number of links. What is left when we examine the network structure is the local spatial ordering of points. Temporal effects have been removed (the rows and columns of the adjacency matrix can be shuffled without changing the network) and large scale variation in the density of points is not considered (each point will contribute the same number of neighbours).

The result of this construction for the chaotic Lorenz system is shown in Fig. 3. In this figure we depict the network constructed from 7000 points sampled from the x-component of the chaotic Lorenz system. The time series and the usual time delay embedding are shown in Fig. 2. Based on continuity of the embedded phase space it is easy to see that the two wings of the Lorenz attractor correspond to the two lobes of the network. We believe that the central region of the network corresponds to the small neighbourhood of the separatrix at the origin of the Lorenz system. The two arms corresponding to transmission between the two wings in either direction. The intricate web structure is a consequence of the fact that temporal successor are not allowed to be neighbours: x_i and x_{i+1} are not neighbours, but they may both be neighbours of x_j ($|j-i|>T$).

In contrast to the Lorenz system, a network constructed from an orbit of the Rössler much more closely resembles the original attractor. Figure 4 depicts the structure of the network derived from the x-component of the chaotic Rössler system. We emphasise here that neither Fig. 4 or Fig. 3 contains any temporal information. The Rössler attractor-like structure of Fig. 4 is a consequence of the proximity of points which are not temporally close. Despite this, even the folding mechanism which gives rise to chaos in the Rössler system is evident.

This strong determinism and the longer chains of these neighbours lead to the complex structure and even the local cycles observed in the network depicted in Fig. 3. Our aim is to find some way to quantify the length and frequency of these cycles in the network. As a first step to this goal we examine the motifs occurring

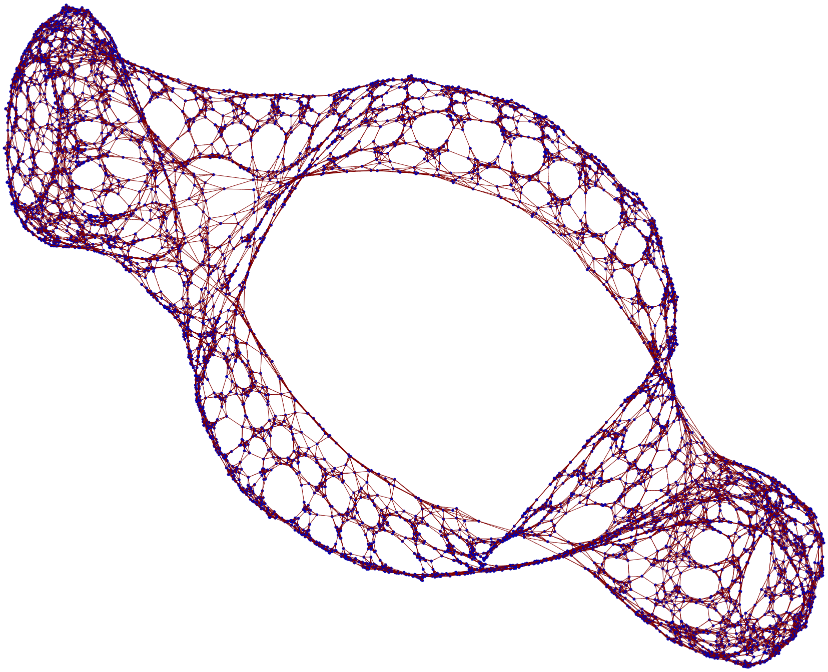


Fig. 3. The complex network constructed from a time series of the x-component of the chaotic Lorenz system. The figure depicts 7000 nodes of the complex network (this choice is merely a constraint of our computing resources). The local neighbourhood structure and the two wings of the Lorenz system, together with the dynamic of the central separatrix are all clearly depicted. The nodes of the network are distributed in $\mathbf{R}^3$ using a spring embedding which aims to places connected nodes close together and unconnected nodes far apart.

in the network and the relative frequency with which the occur [8,7]. A motif of size n is simply a subnetwork consisting of n nodes. For ease of computation, we consider only connected motifs of size 4. We compute the frequency with which each connected motif of size 4 occurs and compare the relative frequency of these motifs, ranking them from most to least frequent. The results of this computation are shown in Table 1. Details of the various systems we study are given in the Appendix.

Table 1 provides a summary of the relative frequency of various different local structures within the network. One observes, for example, a variation in the frequency of the fully connected motif: most common in regular periodic systems, less frequently observed in low-dimensional (one positive Lyapunov exponent) chaos and least frequent for noise and high dimensional chaos. Similarly the symmetric motif (the square) is more common in the high-dimensional systems and less common in the low-dimensional ones. Both these observations are natural: one would expect less mixing in the more regular (i.e. periodic systems) and progressively more mixing in the higher dimensional ones.

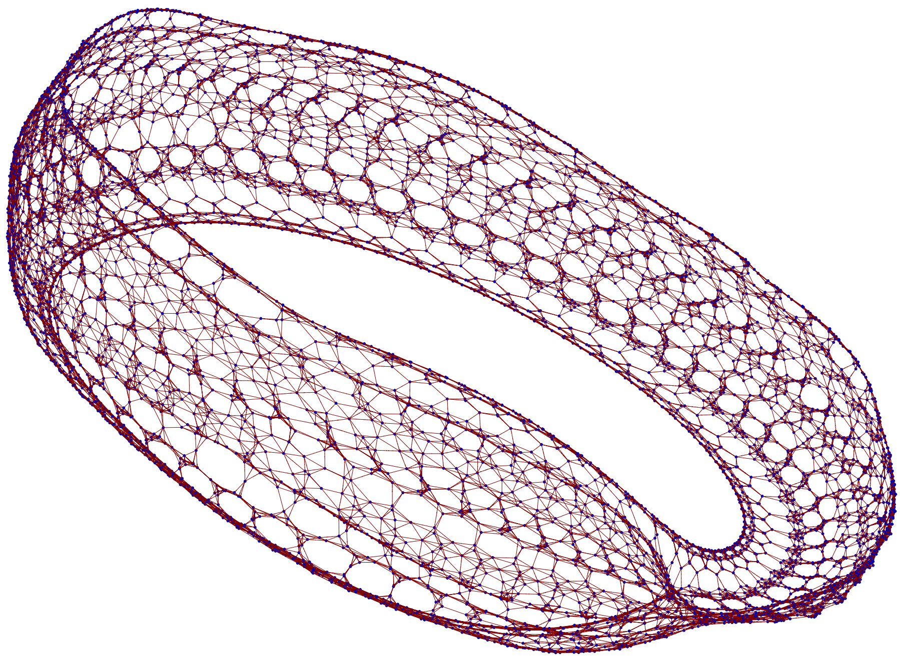


Fig. 4. The complex network constructed from a time series of the x-component of the chaotic Rössler system. The figure depicts 7000 nodes of the complex network. The nodes of the network are distributed in $\mathbf{R}^3$ using a spring electrical embedding which aims to places connected nodes close together and unconnected nodes far apart.

Similarly, the less transitive motifs are more common in the higher dimensional systems. For example the motif consisting of a single node fully connected to the remainder (i.e. a is connected to b, c and d but none of b, c or d are connected to one another) is most common in the systems which exhibit the most stretching. Sensitivity to initial conditions implies that points should spread apart so that the neighbours of a are not necessarily neighbours themselves. This is observed in the relative frequency of this motif for chaotic, hyper-chaotic and periodic system. For noisy signals this motif is most common, for this system points are arranged more randomly and it is therefore unlikely that neighbours will be transitive.

We now turn out attention to discrete chaotic and noise processes. We consider data generated by chaotic maps, hyper-chaotic maps and Gaussian and fractal noise sources. The results are depicted in Table 2 and the general observations that we can make are the same as those observed for flows. Details of the various systems we study are given in the Appendix.

We observe that non-transitive motifs (a connected to b, c, and d, but no other connections) are again more prevalent in the increasingly high-dimensional systems. Fully connected motif are more frequently observed in data from low-dimensional system. The reasons for this are exactly the same as we observed earlier.

Table 1. Relative motif frequency (Motifs of size 4) for a variety of different dynamical systems. Chaotic systems include the classic Lorenz and Rössler systems and Chua's chaotic circuit. In each case with a variety of different parameter values. The periodic Rössler (with period-2 up to period-8 motion) was also tested and found to exhibit distinct dynamics, as was noise contaminated Sine (0 dB to 30 dB) and the infinite-dimensional Mackey-Glass system in a chaotic regime. In each case, the most frequent motif is shown on the left, and the other motifs are then depicted in decreasing order of frequency.

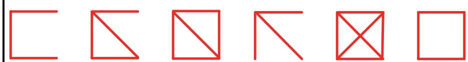
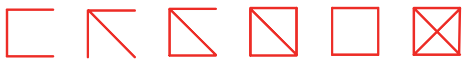
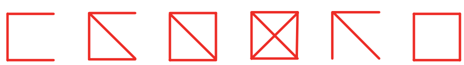
Data Source	Motif frequency
Chaotic Lorenz Chaotic Rössler Chaotic Chua's circuit	
Hyper-chaotic Mackey-Glass	
Periodic Rössler	
Noisy Sine	

Table 2. Relative motif frequency (Motifs of size 4) for a variety of different discrete systems. Chaotic systems are the three usual suspects: logistic map, Hénon map and Ikeda map. We also study two hyper-chaotic system and various noise sources (white noise and various Fractal processes). In each case, the most frequent motif is shown on the left, and the other motifs are then depicted in decreasing order of frequency.

Data Source	Motif frequency
Chaotic logistic map Chaotic Hénon map Chaotic Ikeda map	
Hyper-chaotic folded towel map Hyper-chaotic generalised Hénon map	
White noise Fractal noise	

4 Conclusion

Constructing a complex network from a recurrence plot would provide a method of reconstructing the attractor responsible for the recurrence plot. Essentially, as we see here, the network constructed from the Rössler system quite closely resembles the Rössler system itself. Similarly, if one was simply to use the recurrence matrix as an adjacency matrix one would obtain the generic attractor (close points would be linked, but those links would of course not follow the trajectory). To completely recover something homomorphic to the original attractor one could take the network constructed from the recurrence matrix and then remove all links while retaining the spatial arrangement of nodes on the network. One then reconnects the nodes according to the actual temporal order of the nodes in the recurrence matrix. The spatial adjacency of the points would have been obtained from the network structure. Hence, there is a genuine dualism between recurrence plot and attractor [12]. (In [12] Theil develops the bijection between recurrence plot and attractor slight differently, but the result is the same.) Provided one identifies sufficiently many neighbours to retain the vital information when constructing the recurrence matrix, one may infer the attractor from the recurrence plot.

However, in this paper we are interested in examining a slightly different set of properties. While we have briefly introduced the idea of generating networks from recurrence plots, that is not out main purpose. The adjacency matrix we use is not equivalent to the recurrence matrix and the properties we examine are topological features of the network rather than the temporal structure of the attractor.

We show that super-families exist among the complex networks generated from time series. That is, the relative frequency of occurrence of the various different motif structures is the same for various time series obtained from the same type of complex system: periodic, low-dimensional chaos, high-dimensional chaos, and noise. Nonetheless, despite this super-family structure individual difference exist between members of a single super-family (compare Fig. 3 and Fig. 4). In some sense, the network obtained in each case is a fingerprint of the particular dynamical system that generated the corresponding time series. The super-family structure is a crude counting of simple properties of those fingerprints, but much more complex structure exists. We speculate that the network itself contains detailed information concerning the stability of the underlying dynamics. In Fig. 4 we see the folding mechanism that is the signature of Rössler-type chaos and in Fig. 3 we see three distinct regions and the effect of the central separatrix which characterises the Lorenz system.

Essentially, viewing time series as networks provides a whole new arsenal of nonlinear statistics and measures which one may apply in the analysis of those data. It may also be instructive (for example) to look at the behaviour of the corresponding complex network as the original dynamical system undergoes a bifurcation: say the Rössler system bifurcating through multiple periodicities and into a variety of different chaotic regimes. It would also be instructive to apply this method to problems related to multivariate time series: either to look for coherence and organisation between multiple channels (by replacing the embedding step in the procedure we describe here with a single multivariate time sample) or to look for synchronisation between channels. In either case the change in behaviour would be readily apparent from the structural properties of the resultant network.

Acknowledgements

This work was supported by the University Grants Council of the Hong Kong government (No. PolyU 5268/07E).

References

1. Baier, G., Klein, M.: Maximum hyperchaos in generalized henon maps. Physics Letters A 151, 281–284
2. Eckmann, J.P., Kamphorst, S.O., Ruelle, D.: Recurrence plots of dynamical systems. Europhys. Lett. 4, 973–977 (1987)
3. Lacasa, L.: Private communication (2008)
4. Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuño, J.: From time series to complex networks: The visibility graph. Proc. Natl. Acad. Sci. USA 105(13), 4972–4975 (2008)
5. Mackey, M.C., Glass, L.: Oscillations and chaos in physiological control systems. Science 197, 287–289 (1977)
6. Marwan, N., Romano, M., Theil, M., Kurths, J.: Recurrence plots for the analysis of complex systems. Physics Reports 438, 237–329 (2007)
7. Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., Alon, U.: Superfamilies of evolved and designed networks. Science 303, 1538–1542 (2004)
8. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: Simple building blocks of complex networks. Science 298, 824–827 (2002)
9. Rössler, O.E.: Continuous chaos — four prototype equations. Annals of the New York Academy of Sciences 316, 376–392 (1979)
10. Small, M., Yu, D., Harrison, R.G.: Surrogate test for pseudo-periodic time series data. Physical Review Letters 87, 188101 (2001)
11. Theil, M., Romano, M., Read, P., Kurths, J.: Estimation of dynamical invariants withour embedding by recurrence plots. Chaos 14, 234–243 (2004)
12. Thiel, M., Ramano, M.C., Kurths, J.: How much information is contained in a recurrence plot? Physics Letters A 330, 343–349 (2004)
13. Webber Jr., C.L., Zbilut, J.P.: Dynamical assessment of physiological systems and states using recurrence plot strategies. Journal of Applied Physiology 76, 965–973 (1994)
14. Xu, X., Zhang, J., Small, M.: Superfamily phenomena and motifs of networks induced from time series. Proc. Natl. Acad. Sci. USA 105(50), 19601–19605 (2008)
15. Yang, Y., Yang, H.: Complex network-based time series analysis. Physica A 387, 1381–1386 (2008)
16. Zhang, J., Small, M.: Complex network from pseudoperiodic time series: Topology versus dynamics. Physical Review Letters 96, 238701 (2006)
17. Zhang, J., Sun, J., Luo, X., Zhang, K., Nakamura, T., Small, M.: Characterizing pseudoperiodic time series through the complex network approach. Physica D 237, 2856–2865 (2008)

Appendix

A High Dimensional Chaos

In this section we review the various dynamical systems used to generate the time series data used in this study. We assume that most readers are familiar with the canonical Rössler, Lorenz and Chua systems, and the Ikeda, Hénon and logistic maps (and for details of our particular interpretation we refer the interested reader to [14]). Our other examples need a little further explanation.

Chaotic Mackey-Glass delay system [5]: $a = 0.2$, $b = 0.1$, $d = 17$. The sampling interval $T = 0.25$ and the time delay $\tau = 40$.

$$\dot{x}(t) = \frac{ax(t-d)}{1 + x^{10}(t-d)} - bx(t) \tag{3}$$

Hyper-chaotic generalized Hénon map [1]: $a = 1.9$, $b = 0.03$.

$$\begin{cases} x_{n+1} = a - y_n^2 - bz_n \\ y_{n+1} = x_n \\ z_{n+1} = y_n \end{cases} \tag{4}$$

Hyper-chaotic folded-tower map [9]: $a = 3.8$, $b = 0.2$.

$$\begin{cases} x_{n+1} = ax_n(1 - x_n) - 0.05(y_n + 0.35)(1 - 2z_n) \\ y_{n+1} = 0.1((y_n + 0.35)(1 + 2z_n) - 1)(1 - 1.9x_n) \\ z_n = 3.78z_n(1 - z_n) + by_n \end{cases} \tag{5}$$