Superfamily phenomena and motifs of networks induced from time series

Xiaoke Xu, Jie Zhang, and Michael Small[1]

Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

We introduce a transformation from time series to complex networks and then study the relative frequency of different subgraphs within that network. The distribution of subgraphs can be used to distinguish between and to characterize different types of continuous dynamics: periodic, chaotic, and periodic with noise. Moreover, although the general types of dynamics generate networks belonging to the same superfamily of networks, specific dynamical systems generate characteristic dynamics. When applied to discrete (map-like) data this technique distinguishes chaotic maps, hyperchaotic maps, and noise data.

chaos | complex networks | subgraphs | embedding | dimension

Recently, a bridge between time series analysis and complex networks has emerged (1, 2). Zhang and Small (1) first introduced a transformation from pseudoperiodic (that is, oscillatory) time series to complex networks. By connecting those nodes whose corresponding cycles are morphologically similar, the dynamics of time series are encoded into the topology of the corresponding network. Lacasa *et al.* (2) have proposed an alternative algorithm to characterize periodic, random, and fractal time series based on a similar philosophy. In their scheme, successive scalar time series points are mapped to nodes of the network with links between nodes for which the corresponding points satisfy a condition on their relative magnitudes. By exploiting the fundamental properties of time series that manifest clearly in the corresponding networks, they are able to distinguish between broad classes of dynamical systems.

Although both these schemes have been successfully applied to generate complex networks from time series, the authors of each algorithm have only explored the basic global statistics of the network, such as degree distribution and average path length (3, 4). We note that many networks that have the same basic global properties, such as small-world character (5) and scale-free distribution (6), may have wildly different local structures (7). Conversely, networks with different global properties may demonstrate similar local structures (8). Actually, mounting evidence suggests that there might be strong ties between the global topological properties and key local patterns of networks (9).

In contrast to the degree distribution and the clustering coefficient, the relative frequency of small subgraphs (or motifs) can describe the local characteristics of complex networks (7). The rank distributions of these motifs can reflect the local structural properties and thus can be used to classify networks (8). To understand the transformation mechanism between time series and complex networks, it is important to make a comparison between the local structures of networks from different time series. Whereas the previous works (1, 2) focused on macroscopic properties of the dynamics evident in the network, we turn our attention to the fine features of the dynamics that are *only* evident on examination of the corresponding network.

In this article, we will discuss complex networks constructed from a comprehensive battery of flow and map time series, including periodic signals, periodic signals with noise, chaotic signals, hyperchaotic signals, and white and fractal noise. We focus more closely on the local properties of the networks and, in particular, on the distribution of subgraphs within the net-

works. We examine the frequency of occurrence of various subgraphs within the networks. Furthermore, we find that different types of time series belong to different superfamilies (that is, the set of networks with the same relative abundance of the different subgraphs), and the fine local structures in the complex network domain reflect the state recurrence properties of these time series.

A subgraph of size N is the network formed by examining only N specific nodes from the entire network. We only consider subgraphs in which any 2 nodes are connected either directly or indirectly. This leads to the sequence of hieroglyphs shown in Fig. 1 where we have listed all of the 6 different undirected subgraphs of size 4. By using the scheme described here, we generate a complex network and observe the relative frequency of different subgraphs for networks derived from time series realizations of different dynamical systems.

Results

After building the complex network by using the simple scheme described below, we calculate the number of occurrences of each subgraph and rank them in descending order. The relative frequency with which the different subgraphs occur is shown to be a sensitive measure of the underlying dynamics. We find that time series derived from different classes of dynamics have distinct local characteristics and therefore different subgraph ranks. Hence, data generated by different types of dynamical systems belong to different superfamilies. As we shall see later, our method can be applied not only to flow data, but also to data from maps.

Flow Data. Fig. 2 plots the subgraph ranks of the complex network constructed from different types of flow data, including periodic, chaotic, and noisy periodic data, each with 10^4 points. Here, the time delay τ is determined by the mutual information method (10) and a large embedding dimension ($d_e = 10$) is chosen to reliably unfold the fine structure. We find that the different complex networks from the same type of flow data (e.g., low-dimensional chaos) show the same rank ordering of subgraph frequencies. When this occurs we say that the networks belong to the same superfamiliy and we call this behavior the superfamily phenomenon of the time series. The observation can be explained when referring to the original dynamics of these distinct systems.

Time series derived from different dynamical systems demonstrate distinct local structures in phase space. For a chaotic signal, the reconstructed phase space will generally (although neither necessarily nor sufficiently) generate fractal self-
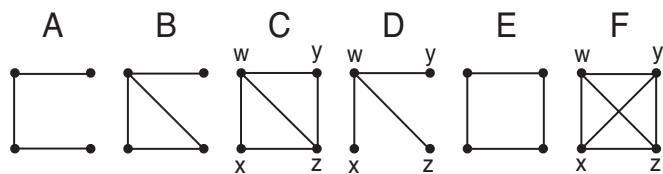
APPLIED PHYSICAL SCIENCES

Fig. 1. All subgraphs of size 4 in undirected networks. We arbitrarily label these subgraphs A, B, C, D, E, and F as shown.

similarity and heterogeneity. Points on the attractor are sparse in some areas, and dense in others. After being transformed into the complex network domain, the points in dense regions are found to have higher degrees than those in sparse regions. This
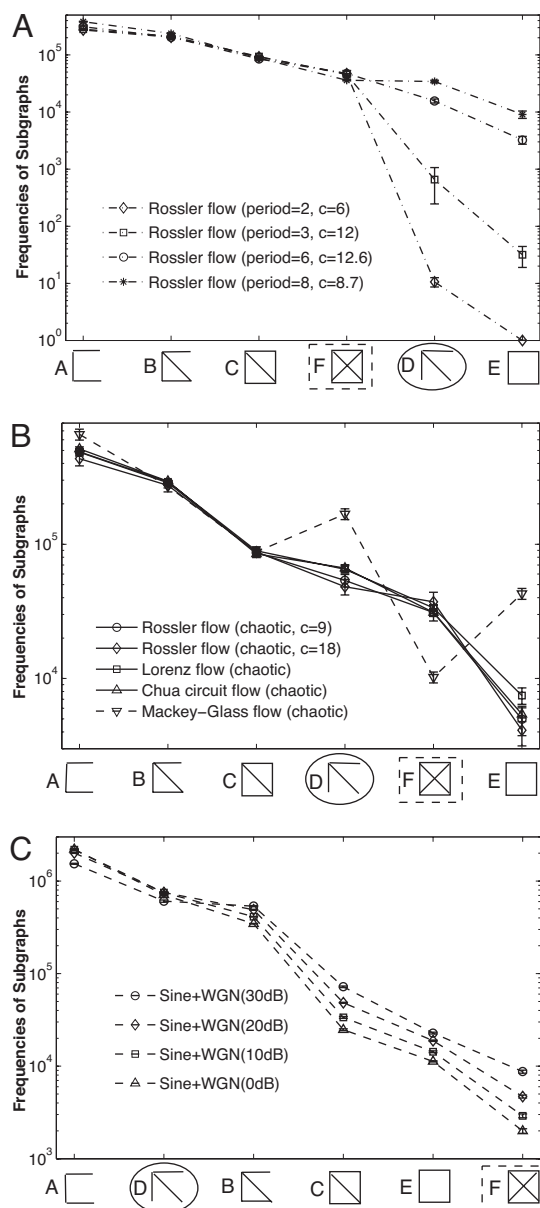

Fig. 2. Subgraph ranks of different types of time series. (*A*) Periodic flow data. (*B*) Chaotic flow data. (*C*) Periodic flow data with white Gaussian noise. Note that in *B*, the Mackey–Glass flow is a high-dimensional chaotic system and does not follow the same subgraph rank ordering as the other systems in this plot.

cannot be simply attributed to the fact that the density of points is higher, but rather that the interconnection among them is stronger. In a sparse region of phase space, the 4 nearest neighbors are less likely to be mutual (i.e., they are nontransitive), and therefore the corresponding node will only have 4 links. Conversely, the points in a dense region of phase space are more likely to be mutual (fully transitive) and therefore have a higher degree. In comparison, periodic flow signals show more regular local structures and a more homogeneous distribution of points on the attractive regions of phase space. Periodic flow signals with different noise levels produce more random structures and hence the region of phase space that they occupy becomes thicker (high dimensional or perhaps more diffuse) than purely periodic flow data.

Note that the distinction among periodic, chaotic, and periodic noisy flow data can be described with reference to the relative frequencies of 2 particular motifs (D and F). The rank of motif D is increased but motif F occurs less frequently from Fig. 2*A* (periodic) to Fig. 2*B* (chaotic) and further to Fig. 2*C* (noisy periodic). This is essentially due to the heterogeneity of the attractor and related to the intrinsic dimension of the system. Motif D will generally occur frequently if it is likely that node w is close to x, y, and z, but x, y, and z are not close to each other (for any 4 nodes w, x, y, and z forming a single motif). Conversely, motif F will occur more frequently if w is close to x, y, and z only when x, y, and z are also close to one another. As we would expect, the former structure D is more likely to appear only in higher dimensions, or when the distribution of points is heterogeneous; whereas F will occur more often when the points are evenly distributed in a low-dimensional (that is, linear or planar) attractor.

We assert that the motif F is more common for stable flow data, and less common for transitive dynamics because of the distribution of embedded points within phase space. Recall that for strictly periodic flow the points will be evenly distributed and, therefore, the strong mutual coupling implied by motif F will be common. For chaotic systems, the distribution is structured but nonuniform (a consequence of the self-similar fractal structure of points on the attractor). Hence, nontransitive coupling structures (such as motif D) will be more common. A careful examination of Fig. 2*A* and a comparison with Fig. 2*B* bears out the above observation; not only through the increasing frequencies of D and E, but also through the change of the relative frequencies of D and F. That is, when the periodicity increases (from period 2 up to period 8; see Fig. 2*A*), D and E increase because the attractor becomes more heterogeneous. Moreover, as the dynamics cease to be periodic and transit to chaos (Fig. 2*B*), the order of occurrence of the 2 subgraphs (D and F) changes too.

For periodic dynamics with significant noise (30 dB to 0 dB), the distinction of subgraph ranks is more obvious in Fig. 2*C*. The last 3 subgraphs (C, E, and F) occur less often when the noise level increases. Adding noise to periodic dynamics increases the dimension of the dynamics while retaining the homogeneity of the distribution of points. This causes the relative frequency of the nontransitive motif D to increase, whereas the frequency of fully transitive motif F decreases further. Moreover, the frequency with which motif C occurs decreases. This motif C will occur if w and z are connected to x and y, but x and y are not connected, so x and y must be close to both w and z, but not to one another. This configuration can frequently occur only if the distance between the points is irregular, i.e., $d(w, z) \ll d(x, y)$, that is, the distribution of points on the attractor is heterogeneous. Again, this is reflected not only by the relative frequency of the various subgraphs, but also by the increasing abundance with which they occur as a function of noise levels. We have repeated this analysis with a correlated noise contamination [an AR(3) process $s_n = 0.8s_{n-1} - 0.5s_{n-2} + 0.6s_{n-3} + \varepsilon_n$], and found
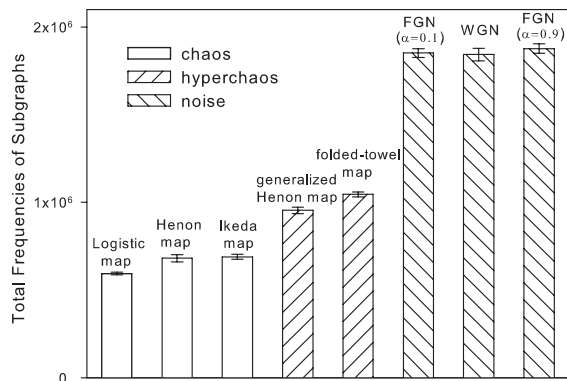
Xu *et al.*

Fig. 3. Subgraph ranks of different types of time series: chaotic map data (*A*), hyperchaotic map data (*B*), and noise data (*C*).

hyperchaotic generalized Henon map and folded-tower map data have the same subgraph ranks and hence belong to 1 superfamily. White Gaussian noise and fractal Gaussian noise ($\alpha = 0.1$ and $\alpha = 0.9$, generated by the method in ref. 11) belong to another superfamily. Moreover, similar to the above flow data, the frequency of occurrence of subgraph D also increases and that of subgraph F decreases from chaos (Fig. 3*A*) to hyperchaos (Fig. 3*B*) and finally to noise (Fig. 3*C*). Motif F will appear more in the homogenous and low-dimensional chaotic structure, whereas, as the data become hyperchaotic and high-dimensional, motif D becomes more prominent.

Conclusions and Discussions

The complex networks, built by our single simple scheme, group different periodic dynamics, chaos (one positive Lyapunov exponent), hyperchaos (multiple positive Lyapunov exponents), random noise and noisy periodic signals into separate super-families. That is, all of the same types of system exhibit one particular relative frequency of subgraphs. For example, all of the chaotic map systems tested exhibit the same ranks of relative frequency of subgraphs. Moreover, within each superfamily, the networks corresponding to time series from different specific dynamical systems exhibit a unique fingerprint specific to that system.

The results reported here were obtained from 10^4 data. For maps we found that a minimum of 5,000 points is required to obtain consistent results, for flow data 9,000 points are required. However, in all cases the method we present assumes that the underlying system is stationary. For time series from a nonstationary (nonautonomous or sufficiently high dimensional) system the theoretical foundation required for time-delay embedding is absent. Nonetheless, in such cases, this method may still be useful as a tool for data analysis. For example, in the case of a system undergoing abrupt parameter change one could imagine that this method would yield disjoint network components.

Compared with the algorithm we present here, we note that the method in ref. 1 does not require phase space reconstruction to build the network, because the correlation coefficient (12) is used to measure the distance between individual cycles (which are taken as nodes in the network). Of course, the focus on cyclic time series in ref. 1 is mainly motivated by the specific research interests of those authors, and the extension of that scheme to nonoscillatory signals is trivial. Comparatively, the mapping mechanism in ref. 2 depends on the statistical persistence of the time series. Two arbitrary points $A(t_a, y_a)$ and $B(t_b, y_b)$ will become connected nodes, if any other data C (t_c, y_c) between them ($t_a < t_c < t_b$) fulfills:

$$\frac{y_b - y_c}{t_b - t_c} > \frac{y_b - y_a}{t_b - t_a} \qquad [1]$$

This means that the linear interpolation between A and B should be larger than the value of all intermediate points. Although this method is simple, it is somewhat unclear what precise aspect of the dynamics is being exploited. Moreover, application of this method to chaotic data remains untested (2).

Distinct from our current algorithm, the complex networks in refs. 1 and 2 are built directly from the time domain, and therefore the construction algorithms have the advantage of being simple: essentially, the embedding step is avoided. The reason why we choose to perform an embedding here is that the basic unit is now points rather than cycles [compared with Zhang's method (1)] and that a phase space reconstruction allows for a deeper understanding of the topology of the data by recovering the inherent structure of the data, possibly high dimensional. For a high-dimensional system, such as various hyperchaotic systems, it is difficult to get enough information merely from 1 dimension (2). An important advantage of using

the motif rank distribution [see supporting information (SI) Fig. S1] to be identical to Fig. 2*C*. In the case of correlated noise, the only distinction we observe is that the variation of frequency of motif E with noise level is substantially reduced.

In addition to several low-dimensional chaotic flows, Fig. 2*B* also shows the result for a high-dimensional chaotic system (the Mackey–Glass system). In this one system the rank ordering of motifs is different from the low-dimensional chaotic flow cases. The Mackey–Glass flow shows an increasing prevalence of less ordered (connected) motifs (motif D vs. C and E vs. F) when compared with examples of low-dimensional chaos. The motif distribution is actually similar to that exhibited by noise-contaminated periodic systems, and again this is due to the relative high dimensionality of the test systems.

Map Data. We have also observed similar phenomena when we apply this method to discrete map data. Fig. 3 depicts the subgraph ranks of chaotic maps, hyperchaotic maps, and white and fractal Gaussian noise where we again build the complex network with 10^4 points. The time delay τ is 1 and the embedding dimension d_e is 5. We find that the chaotic logistic map, Henon map, and Ikeda map have the same subgraph ranks. The

Fig. 4. Total frequencies of all 6 subgraphs of size 4.

a phase space reconstruction is that if the embedding is chosen appropriately, the topological distribution of the set of points in phase space will accurately reflect the underlying dynamics of the original system. Therefore, the network inferred from that phase space reconstruction can be related directly back to the evolution operator of the underlying dynamical system. Moreover, connecting the fixed number of neighbors according to their distance essentially reflects the local recurrence features of the time series (13). For example, random noise has no state recurrence, periodic signals show trivial recurrence, and chaotic data demonstrate nonperiodic state recurrence related to a stretching and squeezing mechanism (14). The subgraph patterns revealed by the local (neighborhood) configuration of nodes essentially relate the state recurrence mechanisms of different systems with the local structures of the complex network, and therefore provide more detailed information than traditional measures (15, 16). Moreover, the recurrence information present in the network is intimately related to the distribution of unstable periodic orbits (UPOs). Indeed, in ref. 1 we saw that the cycles close to the low-order UPOs were located at peaks of the probability density of degree distribution. In the current method, a similar, but more subtle phenomenon occurs. Certainly, the presence or absence of a dense collection of UPOs strongly influences the result, and the motif distributions highlight this fact. However, a more subtle effect is hidden in the temporal organisation of motifs within the network. By emphasizing the commonality of motif distribution for data originating from the same general type of dynamical system, we have suppressed these more subtle characteristics.

In fact the subgraph ranks of time series characterizes the dynamics on a microscopic scale, i.e., it reflects the local properties of time series in phase space. The total frequency of subgraphs is a macroscopic measure similar to entropy (17), which reflects the global statistical characteristics of the time series. We have explored the overall frequencies of all map subgraphs of size 4, which we also find to be directly related to the distinct dynamics of the data. As can be seen in Fig. 4, random noise has the largest number of subgraphs. Next come hyperchaotic and chaotic signals. Note that motif D only consumes 3 links from the overall network degree but the motif F requires 6 links. All of the complex networks from time series have the same nodes and average degrees, so the different ranks of subgraphs will give different corresponding total frequencies

of subgraphs. Whereas the points from a deterministic process (here, we consider chaos) will typically be confined to a low-dimensional manifold, the fully transitive motif F is more common than the nontransitive motif D, which leads to the lower total frequency of these subgraphs. For random noise, the corresponding phase space points exhibit a high-dimensional structure that naturally gives rise to more of motif D but less of motif F, which leads to the higher total frequency of smaller (in terms of degree) subgraphs. The observation that different types of data exhibit different total frequencies of subgraphs is consistent with entropy measures (17) and complexity statistics (18), which indicates that our method links the macroscopic property of time series and microscopic properties of complex networks.

Finally, we note that there is a strong similarity between the method of this article (and also ref. 1, and even ref. 2) and the idea of recurrence plots (19, 20). However, the usual construction used in the 2 methods is different. In this article, we examine the few closest neighbors; a recurrence plot usually is constructed with proximity measured against some fixed threshold. Although it is possible to view the recurrence plot as an adjacency matrix of a complex network, to do so would suppress the most important feature of recurrence plots—temporal ordering. Conversely, the most powerful features in complex network analysis are properties of paths between nodes—that is, chains of connected nodes (and not simply neighboring dynamical strands). These are not evident from any of the usual recurrence measures. For example, it is not clear how to extract motifs from a recurrence plot. At present it is more appropriate to view both methods as complementary tools to analyze different aspects of topological information extracted from the time series.

Methods

The scheme we employ to generate complex networks is described in this section. The dynamical systems and the process of generating time series data are described in *SI Text*. From a given time series we generate a complex network representation as follows.

Step 1. Obtain the time series of a given length.

Step 2. Embed this time series in an appropriate phase space (21), and take each phase space point as a node in the network.

Step 3. Select a fixed number (in what follows, we choose 4) of nearest neighbors for each point (node) and connect each point with its neighbors to form a complex network.

Eligible neighbors should have a temporal separation greater than the mean period of the data, and thereby do not inhabit the same "strand" (13). At each step, each node will be assigned 4 new nearest neighbors irrespective of whether it has been connected before. Nonetheless, we do not allow multiple links between 2 nodes. Hence, on average, each node will have 8 links with 8 other nodes (directionality of links is not considered), and the complex networks from different time series will have the same size and average degrees ($\langle k \rangle = 8$). Because we can ensure this uniformity between the networks we generate for different time series, we do not need to adopt randomized networks (7) and a significance profile (8) when comparing these networks. Here, selecting neighbors with a fixed number is more flexible than simply setting a distance threshold. By doing this we are able to enforce a threshold adaptively according to the point density in different regions of phase space. This could be of great advantage for the analysis of chaotic data that always show similar geometry at different scales in phase space. Nonetheless, the choice of this number of neighbors (here 4) is important. We find that for both flows and maps the results are robust across a range of values. For maps a value in the range 3–8 produces identical results. For the flows described above, 3–4 works best.

1. Zhang J, Small M (2006) Complex network from pseudoperiodic time series: Topology versus dynamics. *Phys Rev Lett* 96:238701.
2. Lacasa L, Luque B, Bllesteros F, Luque J, Nuño JC (2008) From time series to complex networks: The visibility graph. *Proc Natl Acad Sci USA* 105:4972–4975.
3. Albert R, Barabási AL (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74:47–97.
4. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45:167–256.
5. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393:440–442.
6. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512.
7. Milo R, *et al.* (2002) Network motifs: Simple building blocks of complex networks. *Science* 298:824–827.

8. Milo R, *et al.* (2004) Superfamilies of designed and evolved networks. *Science* 303:1538–1542.
9. Vázquez A, *et al.* (2004) The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proc Natl Acad Sci USA* 101:17940–17945.
10. Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. *Phys Rev A* 33:1134–1140.
11. Makse HA, Havlin S, Schwartz M, Stanley HE (1996) Method for generating long-range correlations for large systems. *Phys Rev E* 53:5445–5449.
12. Zhang J, Luo X, Nakamura T. Sun J, Small M (2007) Detecting temporal and spatial correlations in pseudoperiodic time series. *Phys Rev E* 75:016218.
13. Gao JB (1999) Recurrence time statistics for chaotic systems and their applications. *Phys Rev Lett* 83:3178–3181.
14. Gilmore R (2006) Topological analysis of chaotic dynamical systems. *Rev Mod Phys* 70:1455–1529.
15. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. *Physica D* 9:189–208.
16. Wolf A, Swift JB, Swinney HL, Vastano JA (1985) Determining Lyapunov exponents from a time series. *Physica D* 16:285–317.
17. Christoph B, Pompe B (2002) Permutation entropy: A natural complexity measure for time series. *Phys Rev Lett* 88:174102.
18. Rosso OA, Larrondo HA, Martin MT, Plastino A, Fuentes MA (2007) Distinguishing noise from chaos. *Phys Rev Lett* 99:154102.
19. Eckmann JP, Kamphorst SO, Ruelle D (1987) Recurrence plots of dynamical systems. *Europhys Lett* 4:973–977.
20. Marwan N, Romano MC, Thiel M, Kurths J (2007) Recurrence plots for the analysis of complex systems. *Phys Rep* 438:237–329.
21. Packard N, Chrutchfield JI, Farmer D, Shaw R (1980) Geometry from a time series. *Phys Rev Lett* 45:712–716.

APPLIED PHYSICAL SCIENCES