Indonesian Journal of Electrical Engineering and Computer Science
Vol. 3, No. 2, August 2016, pp. 350 ~ 355
DOI: 10.11591/ijeecs.v3.i2.pp350-355

Finding Hidden Communities in Complex Networks from Chaotic Time Series

Jiancheng Sun
Faculty of School of Software and Communication Engineering,
Jiangxi University of Finance and Economics, Maple Road, Nanchang, 330013, China
*Corresponding author, e-mail: sunjc@jxufe.edu.cn

Abstract

Recent works show that complex network theory may be another powerful tool in time series analysis. In this paper, we construct complex networks from the chaotic time series with Maximal Information Coefficient (MIC). Each vector point in the reconstructed phase space is represented by a single vertex and edge determined by MIC. By using the Chua's circuit system, we illustrate the potential of these complex network measures for the detection of the topology structure of the network. Comparing with the linear relationship measure, we find that the topology structure of the community with MIC reveals the hidden or implied correlation of the network.

Keywords: *community structure, complex network, time series*

1. Introduction

In common cases, the sensor data are measured over time that can be called time series. Modern sensor networks are collecting enormous volumes of observational time series in many different applications, such as environmental sensing [1], traffic incident monitoring [2], robot manipulator [3], etc. Generally, it can be safely assumed that the sensors measure variables that are explicitly or implicitly linked, namely, that can be converted into a network version of a time series. We motivate our approach with a theory method in this communication: finding the implicit structure in the network which is transformed from a time series.

The primary purpose of nonlinear time series analysis is to character complicated dynamics for a wide variety of natural and experimental dynamical systems. Similarly, during the last decade, complex network theory has been proposed for understanding the structure and dynamics of complex systems in many areas of science, e.g., in ecology [4], social systems [5], neuroscience[6], or atmospheric dynamics [7]. Based on the reconstruction of phase space, a time series can be mapped into a complex network by measuring the similarity relationships between state vectors in phase space[8]. It therefore provides a novel method to analyze the time series by using the theories and strategies in complex networks fields.

One of the most interesting problems of a network is the community structure detection [9, 10]. Communities are groups of nodes that have more strongly relationship among themselves than with the other nodes of the network. To find the appropriate communities present in a network is an open problem. A related question is still unsolved that is how to find a hidden community. Hidden communities refer to groups of nodes having implied or non-intuitive correlation. For example, an egg and a chicken don't share common features in the sight of body; however, an egg can transform to a chicken by a "complex" way which is an implied relationship between them. Comparing with the classical community detection method, finding a hidden community is a much harder work. In this paper, we investigate the nonlinear relationship among the state point in the phase space, and then map a time series to a network form. Due to the nonlinear relationship, the hidden community can be found which provide a generic way for analyzing phase space properties in terms of network topology. Despite recent works on analyzing time series by complex networks [8], [11-12], to our knowledge, the problem of detecting hidden communities in reconstructed phase space has not been addressed.

Received March 2, 2016; Revised June 25, 2016; Accepted July 10, 2016

2. System Model and Hidden Community Structure Detection

In this section, we will present the method to detect the hidden communities in the context of time series. The first thing to be solved is the reconstructed embedding phase space of the time series, and then the network is mapped from the time series by measuring the correlation among the points in the phase space. Finally the hidden communities are detected with a detection algorithm.

2.1. Reconstruction of Embedding Phase Space

The chaotic system describes the time evolution of a system in some phase space $\Gamma \subset \boldsymbol{R}^m$. A state is specified by a vector $\boldsymbol{x} \in \Gamma$ and a map$\boldsymbol{F}: \boldsymbol{R}^{m+1} \rightarrow \boldsymbol{R}^{m+1}$ exists that transforms the current reconstructed state point $\boldsymbol{x}_i$ to the next state point $\boldsymbol{x}_{i+1}$, i.e.

$$\boldsymbol{x}_{i+1} = \boldsymbol{F}(\boldsymbol{x}_i) \quad (1)$$

Takens's embedding theorem ensures that there is a smooth map $h: \boldsymbol{R}^m \rightarrow \boldsymbol{R}$ such that [13]:

$$x_{i+1} = h(\boldsymbol{x}_i) \quad (2)$$

In Equation (1) and Equation (2), The function $\boldsymbol{F}(\cdot)$ represents the state change of the dynamical system. x_{i+1} is the observed value at time $i+1$ while $h(\cdot)$ represents the observation function.

In common cases, a discrete scalar time series$\{x_i, i = 1,2, \ldots, N\}$ will be observed on the chaotic dynamical system instead of knowing the exact system model. Based on Takens' Embedding theorem, the phase space can be reconstructed based on the appropriate embedding dimension m and the time delay, and the point(reconstructed vector) in the reconstructed phase space can be written as follows:

$$\boldsymbol{x}_i = \left[x_{i-(m-1)\tau}, x_{i-(m-2)\tau}, x_i\right] \quad (3)$$

For simply expression, we assume $\tau = 1$ in following discussion.

2.2. Construction of Complex Network from Time Series

Based on the reconstructed phase space from a given time series, the complex network can be constructed by measuring the correlation among the points in phase space. The points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ denoted in Equation (1) are considered as the nodes in the network, and the link between two vertices is determined by their relationship. In general cases, an undirected network structure can be represented by a adjacent matrix $\boldsymbol{A}$ and the element $a_{i,j}$ of $\boldsymbol{A}$ can be defined as:

$$a_{i,j} = \Theta\left(\varepsilon - d\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)\right) \quad (4)$$

Where $\Theta(\cdot)$ is the Heaviside function, ε a threshold used for defining the neighborhood of a state vector $\boldsymbol{x}_i$, and $d(,)$ some correlation measure. $\boldsymbol{A}$ is a binary matrix that encodes whether or not the phase space distance between two observed "state points" $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is smaller than a certain threshold ε. A network$G=(V,E)$ formally consists of a set of vertices V and a set of edges E between them. An edge $e_{i,j}$ connects vertex v_i with vertex v_j. In this study, vertices represent state vectors in phase space, while edges indicate a close proximity between two state vectors. There is an edge between two vertices unless $a_{i,j} = 0$. In Equation (4), correlation measure $d(,)$ play a very important role to determine the topology of the network. Despite recent works on mapping the time series to the network, the $d(,)$ is proposed based on the linear relationship, e.g, linear correlation coefficient [11, 14], Euclidean distance [12, 15], and ε-recurrence networks [16]. In general cases, a linear correlation cannot fully discover the implied or complex relationship among the variables, and its usefulness is greatly reduced when associations are nonlinear. Thus the linear correlation is not appropriate for revealing the hidden relationship. In this study, we proposed a novel strategy to find the hidden community by using the MIC (Maximal Information Coefficient) which is a nonlinear correlation measure presented recently [17].

The MIC is a powerful metric for measuring the dependence between two variables, and it has been celebrated as a "A Correlation for the 21st century" [18]. If two variables have: a) any association, b) a functional relationship, and c) a non-linear relationship, the MIC can capture it. To calculate the MIC value, several grids at different resolutions on the scatter plot of the two variables is built and the largest possible mutual information is computed by any grid applied to the data, and then the MIC is normalized to the [0,1] range. Considering the MIC, thus Equation (4) can be formatted as:

$$a_{i,j} = \Theta\left(\varepsilon - I_{MIC}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)\right) \tag{5}$$

Where $I_{MIC}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$ is the MIC between the variables $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

After reconstructing network from a given time series, we proposed a novel concept noted as R_{G_i,G_j} on the basis of clustering coefficient to evaluate the change among different topology structure for a network, which is defined as:

$$R_{G_i,G_j} = \frac{1}{n}\sum_{k=1}^{n}\left|C_{i,k} - C_{j,k}\right| \tag{6}$$

Where n is the number of the network G_i which represent the i_{th} topology structure for network G, $C_{i,k}$ is the clustering coefficient of the k_{th} vertex v_k for G_i. For a vertex v_k, the local clustering coefficient for undirected networks can be defined as [19]:

$$C_k = \frac{E_k}{(l_k(l_k-1))/2} = \frac{2E_k}{l_k(l_k-1)} \tag{7}$$

Where l_k is the number of neighbors of v_k and E_k is the number of connected pairs between all neighbors of v_k. It can be seen that a clustering coefficient is a measure of degree to which vertices in a network tend to community together. Consequently, the R_{G_i,G_j} can quantifies the dissimilarity between different topology structures for same network from the point of view of community structure.

Based on the restructured network with the MIC, the hidden communities can be found by a common community detection algorithm. Here we should note that the hidden characteristic of a community is not revealed by the community detection algorithm but rather by the performance of MIC. The reason is plain: the network restructured by the MIC is already a hidden or implied one due to the ability of MIC to capture the nonlinear or complex relationships. Thus the detected community in the restructured network is ought to be a hidden one even if a special step for exploring the hidden factor is not considered in the community detection procedure. In this study, we utilized a fast community detection algorithm proposed by Clauset, Newman and Moore (CNM algorithm) [20] to find the hidden community.

3. Results and Analysis

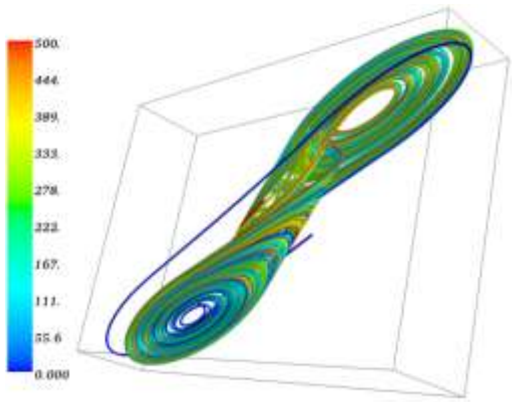
The Results and discussions may be presented individually or combined in a single section with short and informative headings.

In the following, we use the method presented in this paper to generate networks from Chua's circuit system and illustrate our previous considerations by numerical simulation. All three state variables are described by three ordinary differential equations as in Equation (8).

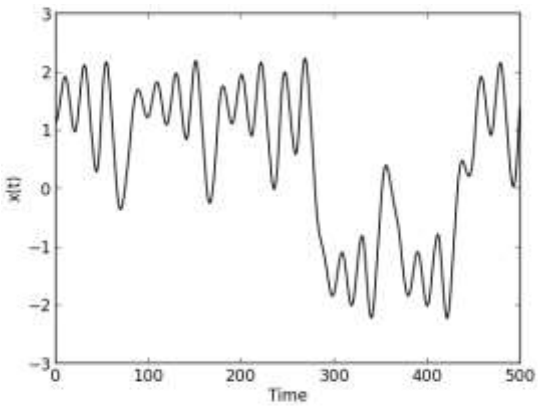
$$C_1\frac{dx}{dt}=\frac{y-x}{R}-g(x)$$
$$C_2\frac{dy}{dt}=\frac{x-y}{R}+z \quad (8)$$
$$L\frac{dz}{dt}=-y$$

Where $g(x)=K_b x+\frac{1}{2}(K_a-K_b)|x+E|+\frac{1}{2}(K_a-K_b)|x-E|$, The parameters used where $1/C_1=9$, $1/C_2=1$, $1/L=14+2/7$, $R=1$, $K_a=-8/7$, $K_b=-5/7$, and $E=1$, with the initial state $[0.\ -0.8\ -0.]^T$. Based on the Runge-Kutta integration method (scipy.integrate method in Python), the time series data set are obtained by solving Equation (8) with constant step length equal to 0.1. The data is mapped into a phase space of dimension $m=50$ that is sufficiently large to study the topological structure. The time series or the orbit to be tracked consists of the first 500 data points. In addition, the network density[21] is used to choose the threshold τ, which is defined as the number of edges divided by the largest number of edges possible.

Figure 1(a) shows the attractor which known as “The Double Scroll” because of its shape in the (x,y,z) space generated by original data (to visualize details of the attractor, the step length for generating the data is set to be 0.0125). We choose the x component of system as the source data which is shown in Figure 1(b) to map to a network.



(a)



(b)

Figure 1. The time series pattern of Chua’s circuit system. (a) Attractor of system and the color indicates the temporal order of observations (from blue to orange); (b) x component of system

Based on the Pearson correlation coefficient and MIC, Figure 2 show the network and community structures and the placement algorithm is used to layout [22]. Figure 2(a) displays a network for the Chua’s system, which corresponds to the adjacency matrix of a Pearson correlation coefficient. We also note the similarity between Figure 2(a) and Figure 1(a), which recover the double-scroll pattern of the original attractor in the reconstructed phase space of the three-dimensional embedding vectors. Figure 2(b) shows the community structure noted as label G which are separated with the different box. From Figure 2(b), it can be seen that the vertices which locate closely in temporal and spatial relationship are grouped in a same community particularly in G1 and G4. In summary, Figure 2(a) and Figure 2(b) can just reveal the apparent and direct structure of time series.

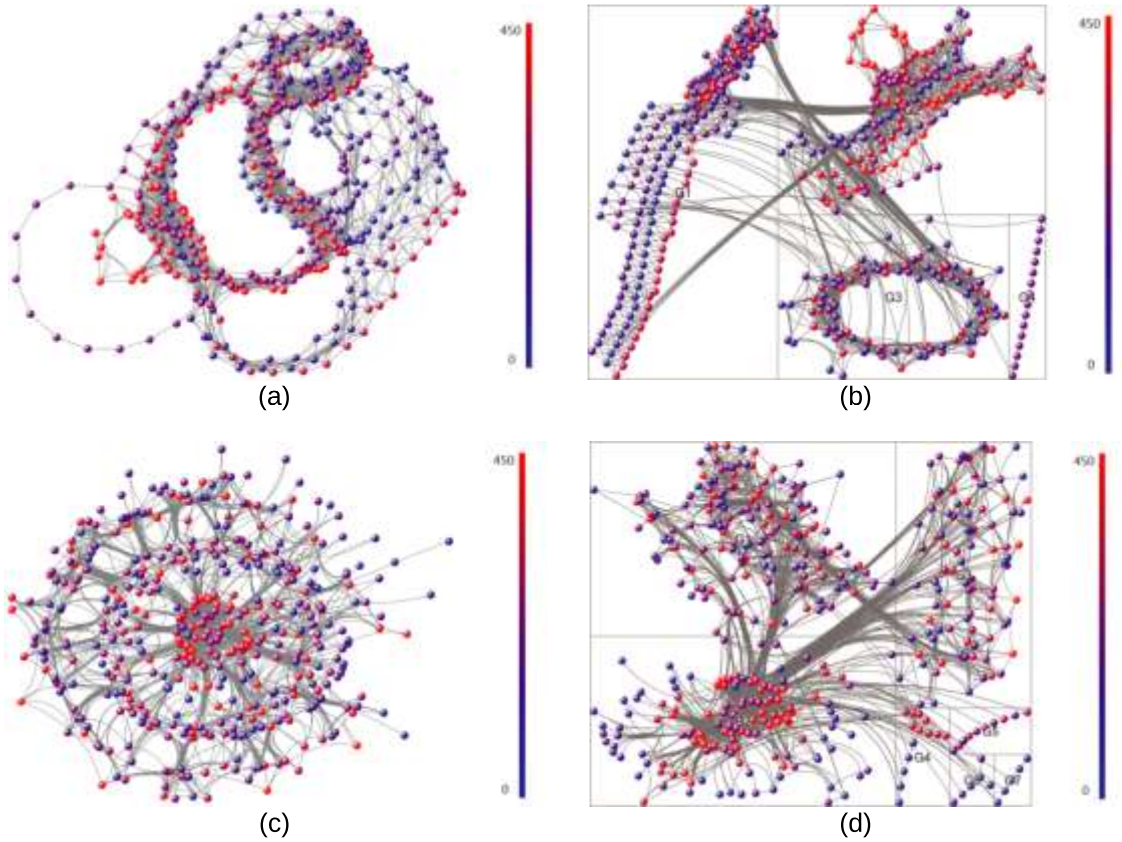


Figure 2. Graphical representation of the different complex networks and community structures constructed from the x coordinate of the Chua's system shown in Figure 1 based on the different relationship measure. The vertex color indicates the temporal order of observations (from blue to bright red noted as the color bar at the bottom of figure). (a) Networks structure based on Pearson correlation coefficient; (b) Community structure of Figure 2a; (c) Networks structure based on MIC; (d) Community structure of Figure 2c

To illustrate the capability of MIC, we show the network structure generated based on MIC in Figure 2(c) and Figure 2(d). Comparing with Pearson correlation coefficient, MIC can capture the hidden structure of a time series. Figure 2(c) reveals a pronounced structure with three major groups that are characterized by a ring-like topology. However, these groups do not correspond to the two scrolls of the attractor in Figure 1(a). To explain the topology structure in details, we group the vertices into seven communities which are shown in Figure 2(d). It can be seen that the vertices belonging to different period of time are grouped into a same community. These communities are the hidden structures of the network since the topology structure are not similar to the geometry property of the attractor.

To explain the difference between two networks with a numerical measure, $R_{G_{MIC},G_{PCC}} = 0.2379$ and $R_{G_{RAN},G_{PCC}} = 0.5428$ described in Equation (6) are calculated, where G_{MIC}, G_{PCC} and G_{RAN} represent the networks generated by MIC, Pearson correlation coefficient and the random graph, respectively. In general cases, the random network is very different from the special network in topology structure. Though $R_{G_{MIC},G_{PCC}}$ is smaller than $R_{G_{RAN},G_{PCC}}$, it still shows that the G_{MIC} is very distinct in topology structure from G_{PCC}.

4. Conclusion

In summary, we proposed a method to map a time series to a complex network with MIC. The goal of this work is to find the hidden community structure by exploring the new relationship measure between two vertices of a network. Each vector point in the reconstructed phase space is considered as a variable and MIC is used for measuring the relationship between different variables. We find that the MIC has a powerful ability to detect the nonlinear or hidden correlation among the vertices and thus the hidden community can be found effectively in a network.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under grant No.: 61362024 and the National Natural Science Foundation of Jiangxi Province, China under grant No.: 20132BAB201022.

References

[1] Hart JK, Martinez K. Environmental Sensor Networks: A revolution in the earth system science? *Earth-Science Reviews.* 2006; 78: 177-191.
[2] Zhang J, Jia L, Niu S, et al. A Space-Time Network-Based Modeling Framework for Dynamic Unmanned Aerial Vehicle Routing in Traffic Incident Monitoring Applications. *Sensors.* 2015; 15: 13874–13898.
[3] Liu B, Zhang F, Qu X. A Method for Improving the Pose Accuracy of a Robot Manipulator Based on Multi-Sensor Combined Measurement and Data Fusion. *Sensors.* 2015; 15: 7933-7952.
[4] Montoya JM, Pimm SL, Solé RV. Ecological networks and their fragility. *Nature.* 2006; 442: 259-264.
[5] Mislove A, Marcon M, Gummadi KP, et al. *Measurement and analysis of online social networks*. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement - IMC '07. New York, USA. 2007: 29.
[6] Zhao X, Liu Y, Wang X, et al. Disrupted small-world brain networks in moderate Alzheimer's disease: a resting-state FMRI study. *PloS one* 2012; 7.
[7] Steinhaeuser K, Ganguly AR, Chawla NV. Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Climate Dynamics.* 2011; 39: 889-895.
[8] Donges JF, Heitzig J, Donner RV., et al. Analytical framework for recurrence network analysis of time series. *Physical Review E.* 2012; 85: 46-105.
[9] Aldecoa R, Marín I. Surprise maximization reveals the community structure of complex networks. *Scientific Reports.* 2013; 3: 1060.
[10] Traag VA, Van Dooren P, Nesterov Y. Narrow scope for resolution-limit-free community detection. *Physical Review E.* 2011; 84: 16-114.
[11] Zhang J, Small M. Complex Network from Pseudoperiodic Time Series: Topology versus Dynamics. *Physical Review Letters*. 2006; 96.
[12] Xiang R, Zhang J, Xu XK, et al. Multiscale characterization of recurrence-based phase space networks constructed from time series. *Chaos.* 2012; 22: 13-107.
[13] Takens F. Detecting strange attractors in turbulence. *Dynamical systems and turbulence, Warwick 1980.* 1981; 898: 366-381.
[14] Yang Y, Yang H. Complex network-based time series analysis. *Physica A: Statistical Mechanics and its Applications.* 2008; 387: 1381-1386.
[15] Xu X, Zhang J, Small M. *Superfamily phenomena and motifs of networks induced from time series.* Proceedings of the National Academy of Sciences of the United States of America. 2008; 105: 19601–19605.
[16] Donner RV, Zou Y, Donges JF, et al. Recurrence networks—a novel paradigm for nonlinear time series analysis. *New Journal of Physics.* 2010; 12: 33-25.
[17] Reshef DN, Reshef YA, Finucane HK, et al. Detecting novel associations in large data sets. *Science.* 2011; 334: 1518-1524.
[18] Speed T. A correlation for the 21st century. *Science.* 2011; 334: 1502-1503.
[19] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature.* 1998; 393: 440-442.
[20] Clauset A, Newman M, Moore C. Finding community structure in very large networks. *Physical Review E.* 2004; 70: 66-111.
[21] Zhang J, Sun J, Luo X, et al. Characterizing pseudoperiodic time series through the complex network approach. *Physica D: Nonlinear Phenomena.* 2008; 237: 2856-2865.