The Thirteenth International Symposium on Artificial Life and Robotics 2008(AROB 13th '08),
B-Con Plaza, Beppu, Oita, Japan, January 31-February 2, 2008

Method to extract stable links out of the noisy links in complex networks made from data by two body functions

Kazuko Yamasaki
Kenneth J. Mackin & Masanori Ohshiro
Kotaro Matusita & Eiji Nunohiro
Tokyo University of Information Sciences, Chiba, Japan
yamasaki@rsch.tuis.ac.jp

Abstract

We study the method to extract a reliable complex network from a set of time series which correspond to nodes. Normally whether a pair of nodes is connected (a link) or disconnected (no link) is decided by the two body functions calculated from the pair time series. But sometimes noise makes it difficult to decide clear links. We separate every time series into two parts by random 0/1 number to get two sample sets of time series, and we decide links for the two sample networks respectively. After repeating this process, we finally get reliable network by choosing common links among several sample networks. We apply the method to climate networks of temperature on the earth.

1 Introduction

Complex networks especially which show scale free have been studied earnestly these days. [3] There are two types of networks. One type of networks is direct expression of networks in real world, for example computer networks made by wired computers. And other type of networks has links defined by the similarity of the components, for example company networks made by correlation function of the stock prices. In latter case, the function which describes the similarity (for example correlation function) includes noise and sometimes it is difficult to extract stable links out of the noise. So we propose the empirical method to distinguish stable links and noise using data. First we explain the method to extract a network from a set of N time series using the two body functions, which describe the similarities between two time series, in simple case. We consider N time series of the length L, $x_i(t)\{i = 1, 2, .., N, t = 1, 2, .., L\}$ (or sources of these time series) as nodes of a network. To determine links of the network, the two body functions $C_{i,j}\{i, j = 1, .., N\}$ which describe the similarity of two time series are computed between every combination of $x_i(t)$. With given threshold σ, we consider a pair of nodes $\{i, j|w(C_{i,j}) > \sigma\}$ has a link. (Here function w expresses the strength of the similarity.) If the two body function $C_{i,j}$ holds the permutation symmetry on subscripts, the network is undirected. A lot of studies have been done in this category. Sometimes no meaningful criterion for inclusion/exclusion of links can be found. So we studied a method to extract the stable links from noisy data and applied climate networks.

2 Simple explanation of the method to extract reliable links

As we mentioned in introduction, sometimes the series $x_i(t)$ include a lot of noise and there is no criterion to decide meaningful links. And the value of threshold σ is not obvious. In these cases, we can test the reliability of the links by dividing a time series into some parts and make one network from one part respectively and check the consistencies of the networks. A straight forward separation is to separate the time axis into some divisions $0 < t < T_1, T_1 < t < T_2, \cdots$. For this separation, we have to consider stationarity of the time series. Another separation is to separate the time series by random numbers. When we separate the time series into two parts, we generate a time series $R(t)$ of the same length as the original time series by random number 0/1. (Fig.1) Next we rearrange the time series $x_i(t)\{i = 1, 2..N\}$ into two time series $x_{0,i}(s)\{i = 1, 2..N\}$ and $x_{1,i}(s)\{i = 1, 2..N\}$.

$$\{x_{0,i}(s)|s = 1, 2, ..L/2\} = \{x_i(t)|R(t) = 0, t = 1, 2, ..L\}$$

$$\{x_{1,i}(s)|s = 1, 2, ..L/2\} = \{x_i(t)|R(t) = 1, t = 1, 2, ..L\}$$

After repeating this process, we can get several sets of the time series from which networks can be extracted,

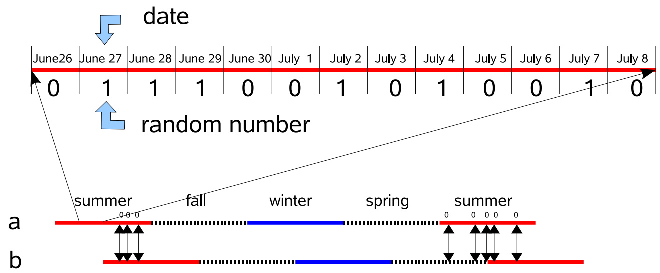

Figure 1: Separation of the time series by random numbers

and we choose common links in these networks for a stable network. In this method, we can not use the information of the memory of each time series but it is possible to extract the similarity of the non stationary time series. Similar method was used in [8]. In their work, they selected a replica time series from original one by random number allowing duplication.

3 Climate networks

Nowadays a lot of attention is paid to various networks , but only a few studies have been done about climate networks on the earth. [6],[7] But there are some phenomena which indicate the study of the climate networks may be useful and fruitful. In Japan it is said that the climate changes in the cycle of three cold days and four warm days, that is "3 kan 4 on" in Japanese. And sometimes the temperature, precipitation or some other climate observations have connection with each other beyond far distances, which is called teleconnection. In this work, we consider the temperature at a location changes cyclically, even if we subtract the seasonal effect from original temperature. And dynamics of the climate produces some correlation between temperatures in different locations with time delay. Imagine limit cycles weakly connected. Some of them are synchronized depending on connected parameters. See [2] The global National Centers for Environmental Prediction - National Center for Atmospheric Research(NCEP-NCAR) provides reanalysis dataset of various climate data in every 100-hPa atmospheric pressure levels and every grid with a resolution of 2° latitude × 2° longitude. We choose 47 locations from the grids so that density of the locations is almost equal on the earth and use 500-hPa and study the daily temperature data from 1979 to 2005.[1]

4 Similarity of the nodes in climate networks

We examined the above idea in the two methods, "the correlation method with shifts" and "the phase synchronization method with shifts", to extract the climate network from temperature time series. First we removed seasonal trends by subtracting the yearly mean temperature of the day from each day's temperature. If we symbolize the temperature of one of 47 locations i as $T_i(t,d)$, where t is the index of the day from 1/1/1979 to 12/31/2005 and d is the index of the day in a year from 1 to 365, the temperature anomaly can be expressed as follows, $x_i(t) = T_i(t,d) - \frac{1}{Y}\sum_{d=d'} T_i(t',d')$, where $Y = 27$ is the number of year in the record.

4.1 Correlation method with shifts

A simple method to get the similarity of the dynamics between locations is to use correlation function. We compute the cross correlation functions $C_{i,j}(\tau) \equiv Cor_t(x_i(t)x_j(t+\tau))$ with time delay daysτ ($-K < \tau < K$) for each pair of locations i and j . The strength of the similarity of the link is chosen to be $W_{i,j} = max_\tau(C_{i,j}(\tau) - mean_\tau(C_{i,j}(\tau)))/std_\tau(C_{i,j}(\tau))$, where max_τ, $mean_\tau$ and std_τ are the maximal value, mean value and standard deviation in the range of τ, respectively. So every ith-jth location pair has the set values of ($W_{i,j}, \tau_{i,j}^{max}$), where $max_\tau(C_{i,j}(\tau)) = C_{i,j}(\tau_{i,j}^{max})$.

4.2 Phase synchronization method with shifts

Next we examine more sophisticated method for dynamical system which can be modeled by weakly connected nonlinear oscillations. Recently phase synchronization method has been applied to a lot of complex dynamical systems, for example chaotic system , the binocular fixation eye movements [4] and temperature and precipitation in different regions etc [5]. In this method, we use nonlinear phase $\phi_i(t)$ which is defined by Hilbert transform. The generalized phase-difference is $\varphi_{n,m}(t) = n\phi_1(t) - m\phi_2(t)$ between two oscillators. After taking mod (π), if the generalized phase-difference is constant, we can say two oscillators are synchronized. But generally it has time dependence, so we have to see the histogram of $\varphi_{n,m}$ to judge whether two oscillators are systematically synchronized or not. Then we use the Shannon entropy. If the entropy equals 0, they are shifted constantly. If the entropy is large, the cycles of two oscillators are independent so they are not synchronized. We constructed the complex signals

$$z_j(t) = x_j(t) + iy_j(t) = A_j e^{i\phi_j(t)}$$

by Hilbert transform of $x_j(t)$. Now we have the nonlinear phase $\phi_j(t)$ at the jth location. Next we define the generalized phase-difference $\varphi_{n,m,i,j}(t)$ for various m and n values and every pair of the locations i and j.

$$\varphi_{n,m,i,j}(t) = (n\phi_i(t) - m\phi_j(t)) \bmod 2\pi$$

We create a histogram of $\varphi_{n,m,i,j}$ with M bins of size $2\pi/M$ and we get the distribution, p_k, of phase-difference of each bin k, that shows how the phase-difference of the pair of two locations occurs. To quantify the systematic occurrence of the phase-difference, we used the Shannon entropy S and define an index $\rho_{n,m,i,j}$.

$$\rho_{n,m,i,j} = \frac{S_{\max} - S}{S_{\max}}, \quad S = -\sum_{k=1}^{M} p_k \ln p_k$$

By definition, the maximum entropy and the range of the index are $S_{\max} = lnM$ and $0 \leq \rho_{n,m,i,j} \leq 1$. $\rho_{n,m,i,j} = 1$ means complete phase synchronization, in this case the distribution of the frequencies p_k shows sharp peak at a value of k. Next we shift the time $t \rightarrow t + \tau$ in one of the two time series $x_i(t)$, $x_j(t)$ and calculate the phase-difference $\rho_{n,m,i,j}(\tau)$ between one time series with shift τ and the other time series without shift. Finally we have max value of the index $\rho_{n,m,i,j}(\tau)$ in deferent values of τ. Here we consider only phase-difference $\rho_{1,1,i,j}$ ($n = 1$, $m = 1$). The strength of the similarity of the link is chosen to be the standardized max value of the index $\rho_{1,1,i,j}$. $W_{i,j} = max_\tau(\rho_{1,1,i,j}(\tau))/mean_\tau(\rho_{1,1,i,j}(\tau))$,where max_τ and $mean_\tau$ are the maximal and mean value of $\rho_{1,1,i,j}(\tau)$ in the range of τ, respectively. Then every ith-jth location pair has the set values of ($W_{i,j}, \tau_{i,j}^{max}$), where $max_\tau(\rho_{1,1,i,j}(\tau)) = \rho_{1,1,i,j}(\tau_{i,j}^{max})$.

5 Extraction of reliable links

We already explained the method simply in section 2. Now to make four networks for four seasons separately, we explain the method in both correlation networks and phase synchronization networks.

5.1 Correlation method with shift

The method for the correlation networks is:

(1) to divide time series a into 4 seasons.

(2) to put 0 or 1 on each day randomly.

(3) to shift time series b by τ days from time series a.

(4) to calculate correlation between daily data of time series a which are on summer and labeled 0 and corresponding daily data of shifted time series b.

(5) to repeat process (4) for 4 seasons and label 0 and 1.

(6) to repeat process (3) $\sim$ (5) for different shift τ.

(7) to put the different random number 0/1 and to repeat process (2) $\sim$ (6) S times.

(8) to repeat process (1) $\sim$ (7) for every combination of time series.

(9) to determine links which have strong positive and negative correlations with shift in 2S data sets respectively.

(10) to regard common links in 2S data sets as stable links.

Fig.2 shows the schematic explanation of above process (1) $\sim$ (10).

5.2 Phase synchronization

The method for the phase synchronization networks is:

(0) to transform time series a and b into phase time series by Hilbert transformation.

(1) Using these phase time series $\widetilde{a}$ and $\widetilde{b}$, do the same process as the correlation networks above (1) $\sim$ (10).

6 Result

The climate network extracted by phase synchronization in winter season with threshold $W_{i,j} > 3.5$ is shown in Fig. 2. The links of black lines show common links which are common in 21 data sets among 24 data sets (87.5%). Fig.3 shows the distribution of number of common links. In the case of the random time series, every pair of nodes in the network has

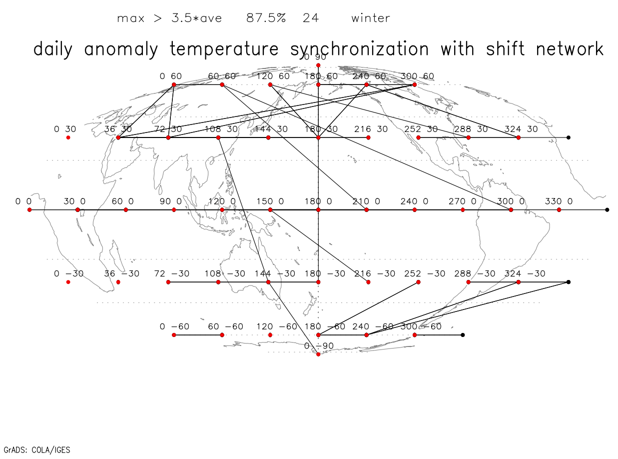

Figure 2: The climate network by phase synchronization in winter season with threshold $W_{i,j} > 3.5$, using 24 sample data sets. The links of black lines show common links which are common in 21 data sets among 24 data sets (87.5%).

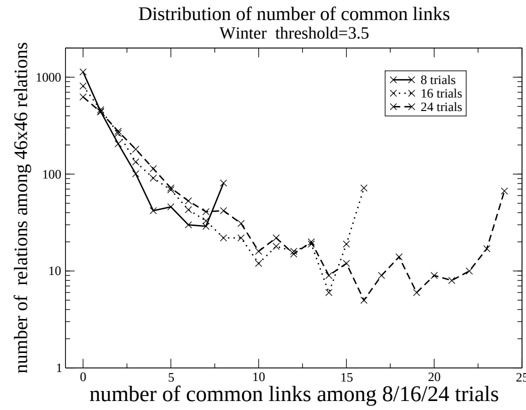

Figure 3: The distribution of the number of common links. y axis is the number of pairs of the nodes and x axis is the number of common links (s) among 8 samples(black),16 samples(dot light) and 24 samples(dot black)(S). In the case of the random time series, the probability is expected to be $C_s^S(1-\delta)^{S-s}\delta^s$, here δ is the probability that each pair has a link.

a link by the same probability of δ, so the probability that the fixed pair i, j has a link in s networks and does not have a link in $S - s$ networks among S networks is $C_s^S(1-\delta)^{S-s}\delta^s$. It decreases rapidly and monotonously according to the increase of s. In Fig. 3, the distributions are quite different from rapid and monotone decrease and the probability goes up when s approaches to S. This shows the common links in high percentage are meaningful.

7 Summary

We studied the method to increase the reliability of a link which is defined by the similarity of the two time series. We separated every time series into two parts by random number and got two sample sets of half length. Repeating this process gave us several sample sets of time series and enables us to choose common links among these sample sets. The distribution of the links shows our method is useful to increase reliability of the links. This work was supported by the Academic Frontier Joint Research Promoting Center and Joint Research Fee of Tokyo University of Information Sciences.

References

[1] "http://www.cdc.noaa.gov"

[2] S. H. Strogatz, (2001), "Exploring complex networks", *Nature*, Vol. 410, pp.268-276

[3] A. L. Barabsi, M. Newman, D. J. Watts, (2006), "The Structure and Dynamics of Networks"

[4] A. Gozolchiania, S. Moshela, et all, (2006), "Decaying of phase synchronization in parkinsonian tremor",*Physica A*, Vol. 366, pp. 552-560

[5] D. Rybsky, S. Havlin, A. Bunde, (2003), "Phase Synchronization in Temperature and Precipitation Records", *Physica A* Vol. 320, pp. 601

[6] A.A. Tsonis, P.J. Roebber, (2004), "The architecture of the climate network",*Physica A*, Vol. 333, pp. 497-504

[7] A.A. Tsonis, K.L. Swanson, and P. J. Roebber," What do networks have to do with climate?",*American meteorological society*, pp.585-595

[8] M. Tumminello, R. N. Mantegna, et all, (2007), "Spanning Trees and bootstrap reliability estimation in correlation based networks", *IJBC*, Vol. 17, No. 7, pp. 2319-2329