Vol 18 No 11, November 2009
1674-1056/2009/18(11)/5091-06

Chinese Physics B

© 2009 Chin. Phys. Soc.
and IOP Publishing Ltd

A preliminary investigation on the topology of Chinese climate networks*

Wang Ge-Li(王革丽)[a)†] and Anastasios A Tsonis[b)]

[a)] *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029, *China*

[b)] *Department of Mathematical Sciences, Atmospheric Sciences Group, University of Wisconsin-Milwaukee, Milwaukee, WI* 53201, *USA*

(Received 5 December 2008; revised manuscript received 20 April 2009)

Complex networks have been studied across many fields of science in recent years. In this paper, we give a brief introduction of networks, then follow the original works by Tsonis *et al* (2004, 2006) starting with data of the surface temperature from 160 Chinese weather observations to investigate the topology of Chinese climate networks. Results show that the Chinese climate network exhibits a characteristic of regular, almost fully connected networks, which means that most nodes in this case have the same number of links, and so-called super nodes with a very large number of links do not exist there. In other words, though former results show that nodes in the extratropical region provide a property of scale-free networks, they still have other different local fine structures inside. We also detect the community of the Chinese climate network by using a Bayesian technique; the effective number of communities of the Chinese climate network is about four in this network. More importantly, this technique approaches results in divisions which have connections with physics and dynamics; the division into communities may highlight the aspects of the dynamics of climate variability.

Keywords: Chinese climate network, complex systems, small world, community

PACC: 9260X

1. Introduction

Recent years have witnessed a growing interest in the study of complex networks. Complex networks describe a wide range of systems, including societies, ecosystems, economies, scientific collaborations, the Internet, and other systems.[1−11]

A network is a set of interconnected nodes, in which a node is a fundamental unit usually with specific dynamical or information content. The nodes in a network can be anything. For example, in the network of scientists, the nodes are scientists who are connected to other scientists if they have collaborated. Generally speaking, there are four basic types of networks: regular networks, classical random networks, small-world networks, and scale-free networks.

Regular networks are networks with a fixed number of nodes, each node having the same number of links connecting it in a specific way to a number of neighbour nodes. They have a high (local) clustering coefficient and a high number of degrees of separation between any two distant points.

Classical random networks are networks where the nodes are connected at random, and do not exhibit local clustering, the number of degrees of separation between any two faraway points is very small. The problem with these networks is that they have a very small clustering coefficient and thus are not very stable. Besides, the degree distribution $P(k)$ of this network is often found to be a Poisson one (the degree distribution $P(k)$ gives the probability that a node in the network is connected to k other nodes).

A small-world network is a type of graph in which most nodes are not neighbours of one another, but most nodes can be reached from every other one by a small number of hops or steps, which is a superposition of regular and classical random graphs. Such networks exhibit a high degree of local clustering but a small number of long-range connections which makes the number of degrees of separation between faraway points as small as if the networks were random. The degree distribution of small-world networks is also a Poisson distribution.

In recent years the term 'small-world effect' has thus taken on a more precise meaning: networks are said to show the small-world effect if the value of average path length scales logarithmically or slower with network size for a fixed mean degree. In general, re-

*Project supported by the National Natural Science Foundation of China (Grant Nos 40890052 and 40505018).

†Corresponding author. E-mail: wgl@mail.iap.ac.cn

gardless of which definition of the clustering coefficient is used, the values tend to be considerably higher than for a random graph with a similar number of vertices and edges.[4]

A network that is scale-free will have the same properties no matter what the number of its nodes is. It is more like a map showing an airline's route, which has a few hubs connecting with many other points (super nodes) and many points connected to only a few other points. Their defining characteristic is that their degree distribution follows a power law relationship defined by $P(k) \sim k^{-\gamma}$. The coefficient γ may vary approximately from 2 to 3 for most real networks; however, in some cases it can also take a value between 1 and 2. An indicator of scale-free characteristics for a network is whether it obeys the power-law.

Scale-free networks are not only efficient in transferring information, but due to the high degree of local clustering they are also very stable.[12] Because there are only a few super nodes, chances are that accidental removal of some nodes will not include the super nodes. It was also shown that scale-free networks have the properties of small-world networks, but the small-world networks from Watts and Strogatz (1998)[9] are not scale free.

2. Climate networks

2.1. Connectivity of the networks

Approaches to construct networks were applied to the climate system by assuming that climate is represented by a grid of oscillators, each of them representing a dynamical system varying in some complex way. Ideas on applying networks to the climate system have been presented by Tsonis *et al.*[13,14] Considering the global grid geopotential height reanalysis data by the National Centers for Environmental Prediction (NCEP) and available through the National Oceanic and Atmospheric Administration (NOAA) online archive at http://www.cdc.noaa.gov/cdc/reanalysis, we follow their studies and give some other analyses here.

The data used here are arranged on a grid with a resolution of 5° latitude ×5° longitude. In this network each grid point is a node, for each grid point winter (December, January, and February) anomaly geopotential height values from year 1949 to 2006 are used. This results in 72 points in the east–west direction and 37 points in the north– south direction for a total of $n = 2,664$ points. These 2,664 points will be assumed to be the nodes of the network.

In order to define the 'connections' between the nodes, the correlation coefficient r at lag zero between the time series of all possible pairs of nodes ($n(n-1)/2 = 3547116$ pairs) is estimated. Note that even though r is calculated at zero lag, a connection should not be thought of as being 'instantaneous'. The fact that the values are seasonal introduces a time scale of at least 3 months to each connection. In this network two nodes are considered as connected if the absolute value of the correlation coefficient of their respective time series is greater than or equal to 0.5; this criterion is based on a statistical significance test (Student's t-test).[14]

Once links are constituted, the architecture of this network is established. This is a more appropriate way to show the architecture of the network because it is a continuous network defined on a sphere, rather than a discrete network on a two dimensional grid.

Thus, if a node i is connected to N other nodes at λ_N latitudes, then its area-weighted connectivity $\tilde{C}_i$ is defined as

$$\tilde{C}_i = \sum_{j=1}^{N} \cos\lambda_j \Big/ \sum_{\text{over all } \lambda \text{ and } \varphi} \cos\lambda, \qquad (1)$$

where φ is the longitude.

The architecture of the resulted network is presented in Fig.1(a), which shows the area-weighted number of total links (connections) at each geographic location with data of the 500 hPa winter anomaly geopotential height. It shows that in the extratropics it appears that certain nodes possess more connections than the rest, which is a characteristic of scale-free networks. In the Northern Hemisphere the presence of regions where such super nodes exist (in China, North America, and the northeast Pacific Ocean) was noted. Similarly, several super nodes are visible in the Southern Hemisphere. Here in this paper, we also give the connectivity distribution of the climate network with data of the 300 hPa and 700 hPa winter anomaly geopotential heights which are shown in Figs.1(b) and 1(c) respectively. We still find that similar super nodes exist, and other plentiful structures are also provided in the tropics zone in the lower layer.

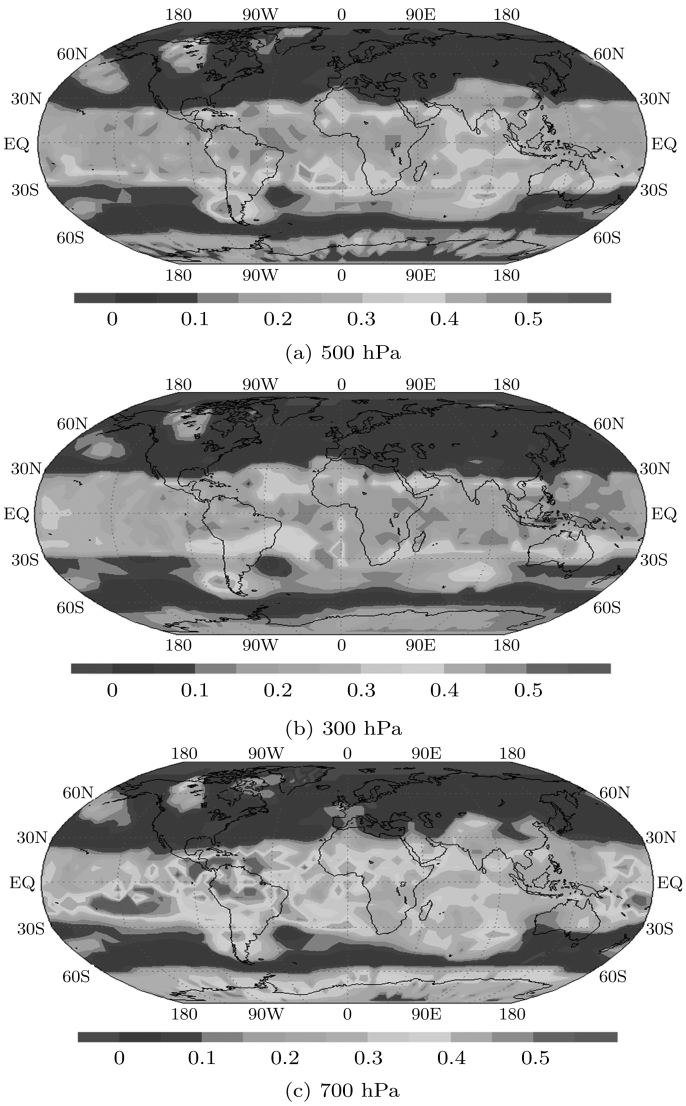


Fig.1. Global distribution of connectivity: (a) with data of 500 hPa; (b) 300 hPa and (c) 700 hPa.

Tsonis *et al*[14] demonstrate that those super nodes may be associated with major teleconnection patterns in climate. For example, the super nodes in North America and the northeast Pacific Ocean coincide with the well-known Pacific North America (PNA) pattern. In the Southern Hemisphere super nodes are also consistent with some of the features of the Pacific South America (PSA) pattern. Furthermore, results show that the average path length L is approximately equal to 9600 km by using the distance on the sphere and the clustering coefficient C equal to 0.56, while for a random network with the same number of nodes and the same average links as this climate network, its average path length is $L_{\text{random}} \cong 7500$ km and its corresponding clustering coefficient $C_{\text{random}} = 0.19$. These values indicate that indeed $L \geq L_{\text{random}}$ and $C > C_{\text{random}}$, which will make this global network close to a small-world network.

On the other hand, for a degree distribution of nodes in the extratropical region of 30°–65°N and 30°–65°S, it is noted that, on a double-logarithmic plot, this degree distribution of nodes does obey a power law, which appears to exhibit a scaling regime similar to those observed in scale-free networks. Also, the slope of it is around –2.0, in agreement with other scale-free networks;[7] but to the tropical zone (20°N–20°S), its corresponding degree distribution does not exhibit the power law regime, but indicates a characteristic of regular almost fully connected networks, where most nodes have almost the same large number of connections, and no super nodes exist. It is

noted that the whole network is a fusion of a fully connected tropical network and a scale-free extratropical network.

Consider that super nodes exist in China (approximately located in the Qinghai–Tibet domain), as shown in Fig.1, and also that China has very extensive meteorological observations.[15−17] Recent studies have paid attention to Chinese climate research through analysis of networks.[18−21] Based on these works, we investigate the topology structure and features of the Chinese climate network. The data we used was the 160 annual average surface temperature derived from the China Meteorological Administration during the period of year 1950 to 1999. These $m = 160$ points are assumed to be the nodes of the network. In order to define the 'connections' between the nodes, the correlation coefficient r at lag zero between the time series of all possible pairs of nodes ($m(m-1)/2 = 12720$ pairs) is estimated. According to formula (1), the area-weighted connectivity of this case is shown in Fig.2, which exhibits the structure of the Chinese climate network. It provides information on links (connections) at each station in China. Unfortunately we have rare data information on the Qinghai–Tibet plateau, so we can get less information there. It is noted that super nodes exist in North China and Southeast China, which are the two important climate zones related to the monsoon effect. However, these values do not make a big difference (say from 0.01 to 0.09), thus to some extent it appears that in the Chinese climate network (in this case) all nodes possess more or less the same number of links, which is a feature of a sort of fully connected network.

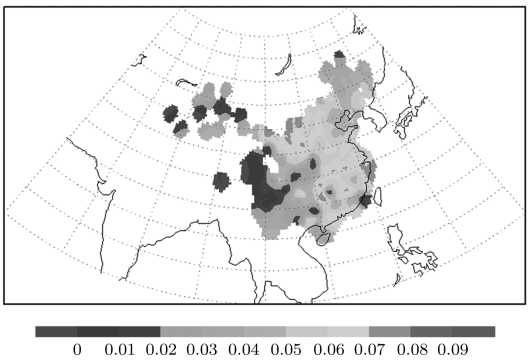


Fig.2. Distribution of connectivity in China.

Moreover, we present the degree distribution of nodes in double-logarithmic coordinates in Fig.3, where no power law exists. This also demonstrates the similar idea that this network shows a characteristic of regular, almost fully connected networks with many nodes with the same number of links and that no highly connected super nodes exist there. In other words, though nodes in the extratropical region provide a property of scale-free networks (Fig.1), they still have other different local fine structures inside.

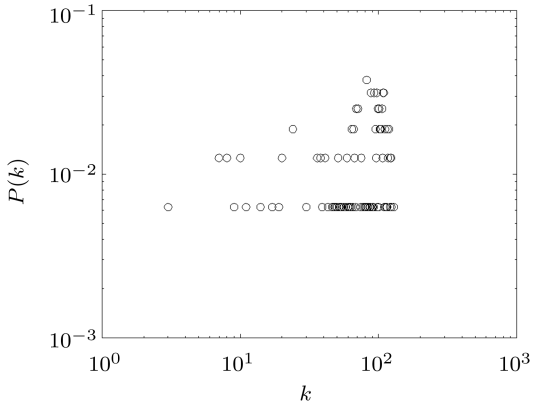


Fig.3. Degree distribution in a double-logarithmic plot with 160 observations.

2.2. Community of the Chinese climate network

One application in studying the networks on describing interactions between a multitude of networks is a community structure in which nodes are grouped into communities, with a large number of interactions within communities and fewer interactions between communities. In recent years, more and more attention has been paid to detecting and characterizing the community structure with various degrees of success.[22−24] For instance, Girvan and Newman[25] gave a method of detecting the community structure based on the betweenness of the edges in the network, and Newman[23] used a leading eigenvector method for detecting community structure.

In this paper, we use a Bayesian approach to network community provided by Hofman and Wiggins (2008).[24] A brief introduction to this method follows.

We use the data (an adjacency matrix from an N-node network), to evaluate the Bayesian evidence $p(D|\boldsymbol{K}) = \int \mathrm{d}\theta\ p(D|\ \theta, K)p(\theta|\boldsymbol{K})$, where D is the data, θ are the parameters and latent community assignments, and K is the number of communities, by integrating over all parameter and latent variable settings.

In doing so, we infer the posterior distribution $p(\theta|D, K)$ over parameters and community assignments and use the evidence to perform complexity control, and determine the optimal number of communities.

To calculate the exact evidence, we integrate over parameters analytically and exploit implicit symmetries to compute the remaining sum over community assignment orders of magnitude faster than complete enumeration, which requires K^N integrand evaluations for N nodes and K communities. While this method can, in principle, be used to evaluate the evidence for networks of arbitrary size, runtimes scale too quickly with N to be practically applicable for large-scale networks.

Figure 4 shows the division into communities of the Chinese climate network using this Bayesian technique. The total number of communities is 11; however, they consist of a very few points between a small number of dominant communities. Thus, it is noted

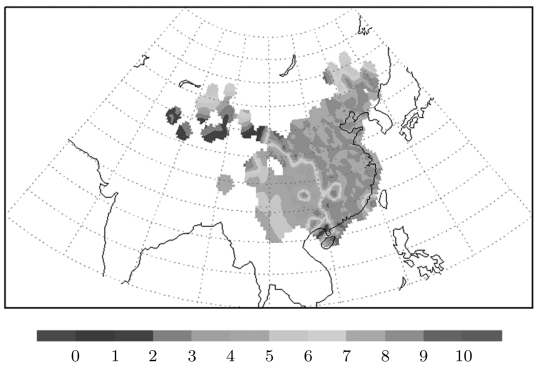


Fig.4. Application of the Bayesian technique to detecting the community of the Chinese network.

that the effective number of communities is about four in this network. More importantly, this technique approaches results in divisions which have connections with physics and dynamics. The network is obviously divided into 4 parts, northeast of China, northwest of China, southwest of China and east of China. This division corresponds to the most dominant climate factor, monsoon: specifically the transition from a moist continental climate in northeast China to an arid continental climate in northwest China, and a southwest monsoon from the Indian Ocean and a southeast monsoon from the Pacific Ocean. The division into communities may highlight the aspects of those dynamics that dominate inter-annual climate variability.

3. Summary

Studying networks provides a new and innovative way to investigate nonlinear systems and is becoming an advancing field in climate research. In this paper, following Tsonis *et al* (2004, 2006, 2008),[13,14,26] we present a brief introduction to networks and some results of the application of networks to climatic data, especially focusing on the Chinese climate data, which indicate that networks delineate some key features of the climate system. According to calculations on the connectivity and community of Chinese climate networks, it is shown that the climate system is an unquestionably complex system, but it is not complicated, where some simple principles exist inside it. This will be a good suggestion for investigating climate processes. In addition, this paper presents preliminary research on the Chinese climate network. Other information with which to understand networks, such as the minimal spanning tree (MST) and synchronization, exist and are worth continuing in this field.

Acknowledgement

We would like to thank Chris Wiggins for sharing his codes of the Bayesian approach for community.

References

[1] Redner S 1998 *Eur. Phys. J.* **4** 131
[2] Camacho J, Guimera R and Amaral L 2002 *Phys. Rev. Lett.* **88** 228102
[3] Newman M E J 2001 *Proc. Natl. Acad. Sci.* **98** 404
[4] Newman M E J 2003 *SIAM Rev.* **45** 167
[5] Newman M E J 2006 *PNAS* **23** 8577
[6] Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, Tomkins A and Wiener J 2000 *Computer Networks* **33** 309
[7] Albert R, Jeong H and Barabasi A L 1999 *Nature* **401** 130
[8] Albert R and Barabasi A L 2002 *Rev. Mod. Phys.* **74** 47
[9] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[10] Strogatz S H 2001 *Nature* **410** 268
[11] Chen H B, Fan Y, Fang J Q and Di Z H 2009 *Acta Phys. Sin.* **58** 1383 (in Chinese)
[12] Barabasi A L and Bonabeau E 2003 *Sci. Amer.* **288** 60
[13] Tsonis A A and Roebber P J 2004 *Physica* A **333** 497
[14] Tsonis A A, Swanson K L and Roebber P J 2006 *Bull. Amer. Meteor. Soc.* **87** 585
[15] Feng G L, Gong Z Q, Zhi R and Zhang D Q 2008 *Chin. Phys.* B **17** 2745
[16] Mo J Q and Lin W T 2004 *Acta Phys. Sin.* **53** 996 (in Chinese)

[17] Feng G L, Dong W J and Jia X J 2004 *Chin. Phys.* **13** 413
[18] Wang X J, Gong Z Q and Zhi R 2009 *Acta Phys. Sin.* **58** 6651 (in Chinese)
[19] Zhi R, Gong Z Q, Zheng H M and Zhou L 2009 *Acta Phys. Sin.* **58** 2113 (in Chinese)
[20] Gong Z Q, Zhou L, Zhi R and Feng G L 2008 *Acta Phys. Sin.* **57** 5351 (in Chinese)
[21] Zhou L, Gong Z Q, Zhi R and Feng G L 2008 *Acta Phys. Sin.* **57** 7380 (in Chinese)
[22] Rosvall M and Bergstrom C T 2007 *Proc. Nal. Acad. Sci.* **18** 7327
[23] Newman M E J 2006 *Phys. Rev.* E **74** 036104
[24] Hofman J M and Wiggins C H 2008 *Phys. Rev. Lett.* **100** 258701
[25] Girvan and Newman M E J 2002 *Proc. Nat. Acad. Sci.* **99** 7821
[26] Tsonis A A, Swanson K L and Wang G L 2008 *Journal of Climate* **21** 2990