Physica A 387 (2008) 5287–5294

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Estimating the clustering coefficient in scale-free networks on lattices with local spatial correlation structure

Anastasios A. Tsonis [a,*], Kyle L. Swanson [a], Geli Wang [b]

[a] Department of Mathematical Sciences, Atmospheric Sciences Group, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, United States

[b] Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 13 December 2007
Received in revised form 26 February 2008
Available online 5 June 2008

PACS:
89.75.-k
89.75.Fb
89.75.Hc
92.60.Bh
92.70.Gt

Keywords:
Networks
Scale-free networks
Clustering coefficient
Climate
Complex systems

ABSTRACT

When we study the architecture of networks of spatially extended systems the nodes in the network are subject to local correlation structures. In this case, we show that for scale-free networks the traditional way to estimate the clustering coefficient may not be meaningful. Here we explain why and propose an approach that corrects this problem.

1. Introduction

In classical networks the relative position of nodes and links is arbitrary and there are no spatial correlations between nearby nodes (for example, a network of scientists where a link indicates that the two corresponding scientists have co-authored a paper). In such networks two popular measures are the characteristic path length (or diameter) and the clustering coefficient [1]. The characteristic path length is the average number of links in the shortest path between two nodes, and it indicates the degree of information transfer in the network (the smaller the diameter the easier the transfer). The definition of the clustering coefficient in classical networks is illustrated in Fig. 1. The top panel shows a hypothetical network and the links of a specific node (the central one, in this case denoted as i). According to this information the central point is connected to eight other nodes (i.e. its degree is eight). These eight nodes are called the neighbors of node i and define the closest neighborhood of this node. To estimate the clustering coefficient for this node we then find the number of distinct links between these eight neighbors, Δ_i. For any number of k_i neighbors there are at most $k_i(k_i-1)/2$ possible connections. This happens when each node in the neighborhood is connected to every other node in the neighborhood (the central point is not part of the neighborhood). In our example there are five links between the neighbors. Thus, $\Delta_i = 5$. Then, the clustering coefficient for node i is $C_i = 2\Delta_i/(k_i-1)k_i$ [1]. Based on this definition the clustering coefficient varies between [0, 1]. The average C_i over all nodes provides the clustering coefficient of the network, C. For a fully connected network $C = 1$ and for

* Corresponding author.
E-mail address: aatsonis@uwm.edu (A.A. Tsonis).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.05.048

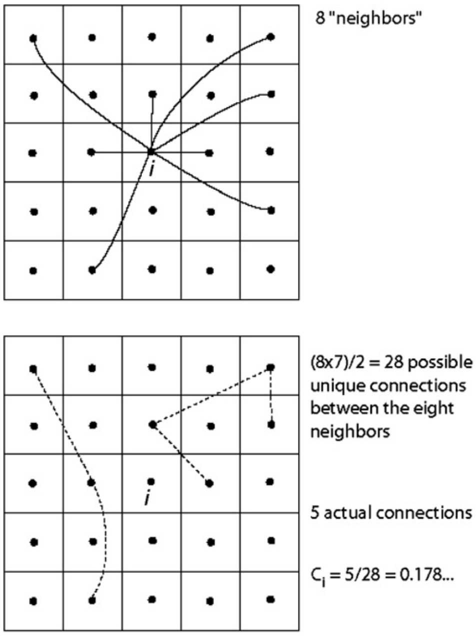

Fig. 1. Illustration of the method to estimate the clustering coefficient (see text for details).

a random network $C = \langle k \rangle / N$ where $\langle k \rangle$ is the average number of links per node and N is the total number of nodes in the network. The clustering coefficient relates to the local "cliqueness" and the higher it is the better the network can withstand the effect of link removal, which tends to fragment the network thereby making it less stable (more random).

2. The problem

In the specific case of scale-free networks [2], which are characterized by the presence of supernodes, the clustering coefficient as defined in [1] predicts that removal of the supernodes results in a decrease of the clustering coefficient. This, however, may not be the case for a scale-free network constructed from fields with spatial correlations. Such scale-free networks are more realistic in the physical world and may exhibit interesting features and properties [3]. For example, consider the network constructed from the extratropical (30 N–90 N) 500 hPa field (Fig. 2(top)). A 500 hPa value indicates the height of the 500 hPa pressure level and provides a good representation of the general circulation (wind flow) of the atmosphere. Here the data in the period December–March at a resolution of $2.5^\circ \times 2.5^\circ$ are used. For each grid point a time series of monthly anomaly values in the period December–February from 1950 to 2004 is available. Thus, it is assumed that the general atmospheric circulation is represented by a grid of oscillators each one of them representing a dynamical system varying in some complex way. This field is characterized by local spatial correlations extending up to some characteristic scale as well as long range spatial correlations. In constructing the network each grid point is a node and two nodes are considered as connected if the absolute value of the correlation coefficient of their respective time series is greater or equal to 0.5. The architecture of the resulted network is presented in Fig. 2(top), which shows the area weighted number of total links (connections) at each geographic location. More accurately, it shows the fraction of the total area that a point is connected to. This is a more appropriate way to show the architecture of the network because the network is a continuous network defined on a sphere [4]. Thus, if a node i is connected to N other nodes at λ_N latitudes then its area weighted connectivity, $\tilde{C}_i$ (which is analogous to the degree of node i), is defined as

$$\tilde{C}_i = \sum_{j=1}^{N} \cos \lambda_j \Big/ \sum_{over\ all\ \lambda\ and\ \varphi} \cos \lambda \tag{1}$$

where ϕ is the longitude. In the above expression the denominator is the area of the northern hemisphere surface north of 30 N and the numerator is the area of that surface a node is connected to. We observe that the resulted network is characterized by the presence of dominant supernodes; a property of scale-free networks [2]. As has been extensively discussed [4–6] this network has indeed properties of scale-free networks and these supernodes correspond to two major atmospheric teleconnection patterns: the Pacific North America (PNA) [7] pattern and the North Atlantic Oscillation (NAO) [8–10].

As mentioned above, in aspatial scale-free networks removal of supernodes results in a smaller clustering coefficient; a direct consequence of the network becoming more fragmented and more random [11]. But in our case, we find that this