Physica A 390 (2011) 4027–4033

Contents lists available at SciVerse ScienceDirect

Physica A

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

The effect of hub nodes on the community structure in scale-free networks

Yang Wang, Yanqing Hu, Zengru Di, Ying Fan *

Department of Systems Science, School of Management and Center for Complexity Research, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 20 October 2010
Received in revised form 14 April 2011
Available online 5 July 2011

Keywords:
Complex network
Power law exponent
Hub nodes
Community structure

ABSTRACT

Many networks have two important features in common (1) the scale-free degree distribution $P(k) \propto k^{-\alpha}$ and (2) the community structure. In this paper, we focus on the relationship between these two features in complex networks. We first investigate the effect of the power law exponent α on the community structure in artificial networks and some real-world networks. Generally speaking, we find out that the networks with significant community structure, often have a large α. Second, hub nodes removal from scale-free networks affects the community structure more considerably than random removal. Our observation indicates that hubs may be the explanation for that scale-free networks often have fuzzy community structure.

1. Introduction

Networks, despite its simplicity, reveals a set of crucial features of real complex systems and has been an important tool in the past decades [1–5]. Interactions between proteins in living organisms, relations between people in society and connections between web can all be described as complex networks. Many networks are characterized by a broad distribution of node degree, called "scale-free networks", such as Internet [3] and some biological networks [6]. In recent years, scale-free networks have attracted particular attention for their mathematical properties, which sometimes lead to surprising physical consequences. The degree distribution of scale-free networks follows the power law distribution $P(k) \propto k^{-\alpha}$, where α is a constant parameter of the distribution known as the power law exponent parameter, and k represents node degree. In scale-free networks, there are quite a lot of nodes whose degrees are large, called "hubs". The power law exponent typically lies in the range $2 < \alpha < 3$ in real-world networks, although there are occasional exceptions. The scale-free networks often have some special properties, such as high robustness against random failures [7] and the absence of a threshold for the spreading of epidemics [8].

Another important feature of complex networks is the "community structure", the tendency for vertices to be divided into groups, with dense connections within groups and only sparse connections between them [9]. Many real-world networks have been shown to possess community structure, such as social networks [9,10], biochemical networks [11–13], and information networks such as the web [14]. Exploring network communities is important for three main reasons [15]: (1) to reveal network at a coarse level; (2) to better understand dynamic processes taking place on the network such as epidemic spreading [16,17] and synchronization [18,19]; (3) uncover relationships between the nodes which are not apparent by inspecting the graph as a whole while can typically be attributed to the function of the system. As a result, community structure has attracted much attention in recent years and there are two related questions. One is how communities can be detected in networks and a huge variety of community detection techniques have been developed in recent years, based on centrality measures [9], the modularity-based method [20–24], random walks [25,26], eigenvectors of matrices [27] and

* Corresponding author. Tel.: +86 10 58807876; fax: +86 10 58800141.
E-mail addresses: yanqing.hu.sc@gmail.com (Y. Hu), yfan@bnu.edu.cn (Y. Fan).

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.06.031

many other approaches (see Ref. [28] as a review). The other question is the concept of significance which is related to the robustness or stability of a partition against perturbations of the graph. If the community structure is significant, it will be uncovered even if the structure of the graph is modified. We know that some networks possess clear communities, while others do not. However, almost all algorithms could uncover community structures in networks in their own ways without considering whether these communities actually exist. Thus, the question on community identification regards measuring the "significance of communities". There are also some approaches explored in the literature. [29–33]. The scale-free degree distribution and the community structure are both important to complex networks. However, little work has been done so far to consider a fundamental question: How does scale-free degree distribution affect the significance of the communities? What is the relationship between scale-free degree distribution and the community structure? This is what we will try to access in this paper.

The previous investigations had a detailed discussion about the connection between hub nodes and modularity in biological networks [6,34,35]. They showed that how the removal of hubs affects the topology of the whole network, such as the size of the giant component and the path length, and divided hubs into two categories such as "date hubs" and "party hubs". Guimera and Amaral have also proposed a classification of the nodes based on their roles within communities [12]. However, what is the relationship between hub nodes and the significance of communities are still mostly unclear. Uncovering the relationship is the main goal of this paper. Knowing the effect of hub nodes on communities can also provide valuable insights into how communities are organized. First, we investigate how the power law exponent α affect the significance of the community structure in scale-free networks. We investigate both the artificial and the real-world networks, and find out that scale-free networks with significant community structure often have large power law exponent α. Second, we consider how the hub nodes in scale-free networks affect the community structure. We remove some hub nodes from artificial networks as well as some real-world networks and find out that hub nodes removal affect the community structure more considerably than random removal. Our observation indicates that in scale-free networks clusters are often not well separated due to the presence of the hubs.

The organization of this paper is as follows. In Section 2, we investigate the relationship between power law exponent α and the significance of the communities in artificial and real-world networks. How the hub nodes removal affect the community structure is proposed in Section 3. In Section 4, we discuss our results and the conclusion is given.

2. The effect of α on community structure

In this section, we investigate both artificial as well as some real-world networks, and focus on the relationship between the power law exponent α and the community structure.

Recently, Lancichinetti et al. have proposed a realistic artificial network called the LFR benchmark [36]. This benchmark is suitable for our investigation due to its two basic properties, namely the scale-free degree distribution and the community structure. In the benchmark each node is given a degree taken from a power law distribution $P(k) \propto k^{-\alpha}$ where k is the node degree, and the size of the communities is also taken from a power law distribution with exponent β. Furthermore, each node shares a fraction $1-\mu$ of its links with the other nodes of the same community, where μ is the mixing parameter. We can change the power law exponent α and keep other parameters constant in this benchmark. In this paper, the parameters chosen were $n = 1000$, $\beta = 1.0$, $\mu = 0.25$, average degree $\langle k\rangle = 20$, maximum degree is 50, maximum and minimum community sizes are 50 and 20, respectively.

There have been some indexes to evaluate the significance of community structure [23,29–31]. Moreover, the presence of near-zero eigenvalues of Laplacian matrix generally indicates the existence of strong communities, or nearly disconnected components [37]. The multiplicity of the (exactly) zero eigenvalue is equal to the number of disconnected components. This is a description of the significance of community structure. Then Hu et al. proposed an index H to evaluate the significance of the community structure using the spectra theory [31]. The method could quantify the significance of the community structure, and it is independent of partition algorithms.

We begin by defining the adjacency matrix A to be a matrix with elements: $A_{ij} = 1$, if there is a connection between vertices i and j, $A_{ij} = 0$, otherwise. The corresponding graph Laplacian L is defined as: $L_{ij} = -A_{ij}$ if $i \neq j$, and $L_{ii} = k_i$, where k_i is the degree of node i. λ_i is the eigenvalue and v_i is the corresponding eigenvector. We let $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$. The significance of community structure H is obtained [31]:

$$H = \frac{n}{k \sum_{j=c+1}^{n} \frac{1}{|\overline{\lambda}-\lambda_j|}} \tag{1}$$

where $\overline{\lambda}$ is the average value of λ_2 to λ_c, k is the average degree of the network, c is the number of communities in a network and n is the number of the vertices in the network.

Besides H, there are other indexes. In Ref. [29], Gfeller et al. proposed a method to evaluate the "robustness" of communities. They added a random noise over the edges, and they also computed "in-cluster probability", denoted by p_{ij}. This probability is defined as the fraction of times nodes i and j have been classified in the same cluster during several occurrences of the clustering algorithm. Then they defined an index denoted by ES, which represents the entropy of the