Physica A 451 (2016) 440–450

Contents lists available at ScienceDirect

Physica A

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

Measuring mixing patterns in complex networks by Spearman rank correlation coefficient

Wen-Yao Zhang [a], Zong-Wen Wei [a], Bing-Hong Wang [a,*], Xiao-Pu Han [b]

[a] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, People's Republic of China

[b] Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou 311121, People's Republic of China

HIGHLIGHTS

- A novel measure of assortative mixing based on the Spearman correlation coefficient is proposed.
- The modified Spearman rank satisfies a linearity condition.
- The linear relation is an important factor for inferring other parameters of networks.
- The simple exponent model can generate networks with certain coefficient directly.

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form 17 December 2015
Available online 29 January 2016

Keywords:
Spearman coefficient
Mixing patterns
Complex networks

ABSTRACT

In this paper, we utilize Spearman rank correlation coefficient to measure mixing patterns in complex networks. Compared with the widely used Pearson coefficient, Spearman coefficient is rank-based, nonparametric, and size-independent. Thus it is more effective to assess linking patterns of diverse networks, especially for large-size networks. We demonstrate this point by testing a variety of empirical and artificial networks. Moreover, we show that normalized Spearman ranks of stubs are subject to an interesting linear rule where the correlation coefficient is just the Spearman coefficient. This compelling linear relationship allows us to directly produce networks with any prescribed Spearman coefficient. Our method apparently has an edge over the well known uncorrelated configuration model.

1. Introduction

Complex network is an interdiscipline that described the complex relationship in nature, social and engineering systems, which recovers the rich diversity of real-world networked system, and found to have lots of common structural features [1–5]. One of the most important results is that the degree distribution of many large networks follows a power law, $P(k) \sim k^{-\gamma}$ (degree k of a node is the number of nodes connected to it), which significantly deviates from the Poisson distribution of classical Erdös–Rényi(ER) random networks [1,6–8]. These networks are called scale-free network.

In order to capture the inner structures and behaviors of real-world networked systems, along with the degree distribution, the mixing patterns in networks are also essential. Disassortative networks imply that the nodes with high degree tend to be connected to low-degree nodes, while assortative networks prefer connections between high-degree nodes. A wide variety of technological and biological networks are disassortative, while lots of social networks are assortative

* Corresponding author.
E-mail address: bhwang@ustc.edu.cn (B.-H. Wang).

http://dx.doi.org/10.1016/j.physa.2016.01.056
0378-4371/

[5,6,9,10]. It plays an important role in many fields, such as mean distance, robustness, stability, percolation thresholds, epidemic spreading and synchronization of oscillators [9–19]. To measure the mixing patterns in complex networks, Newman introduced a method based on Pearson correlation coefficient [9,10].

However, further studies indicate that Pearson coefficient has a serious drawback, its value critically depends on the size and degree distribution of networks. In particular, it would converge to zero for large scale-free networks [20–22]. This drawback strongly impedes the quantitative comparison of different networks. This point cannot be ignored, as the size of nowaday networks is getting larger and larger, for instance, scientific collaboration networks [23], the World-Wide-Web [24].

In this paper, to solve the above drawback of Pearson-coefficient-based measures, we propose a novel method based on Spearman correlation coefficient to measure the mixing patterns in complex networks. In mathematics, Spearman coefficient is usually used to quantify how well two columns of data monotonically depend on each other. It is rank-based, nonparametric, independent of the network size and degree distribution. Hence it could overcome the aforementioned drawback of Pearson coefficient and work better. Moreover, we discover that the rank orders of stubs are statistically in a linear correlation, and the correlation coefficient is just the Spearman coefficient. We argue that this kind of correlation universally exists in complex networks. It can be used as an essential factor to determine the joint degree probability distribution of networks. Based on this linear relationship and Marrows' results on ranking model [25–27], we give a simple exponent approximate form to describe the jointed degree probability distribution of networks, which could directly generate a network with any prescribed Spearman coefficient.

This paper is organized as follows: In Section 2, we introduce Spearman correlation coefficient for networks and indicate that it is scale-independent and effective. In Section 3, we show that the normalized Spearman ranks of links fit a linear correlation. Applying this to a simple exponential model, we show how to generate networks with a prescribed Spearman coefficient. In the last section, we give our conclusions.

2. Spearman coefficient of mixing patterns in networks

2.1. Normalized Spearman ranks of networks

Scale-independent measurement is important to the study of complex networks. As the commonly used measurement of networks' mixing patterns, Pearson coefficient, depends on the size of a network and will vanish for large scale-free networks [20–22]. In contrast, Spearman rank correlation coefficient is a nonparametric measure of statistical dependence and assesses the monotonic relation between two variables.

For Spearman rank correlation coefficient, the variable's rank is used instead of the value itself, which is the average of their positions in the ascending order of the values (see Appendix A). A perfect monotone function occurs a value of −1 or 1 for Spearman coefficient, and 0 occurred to no correlation. In complex networks, the variable of mixing patterns that we focus on is the degrees to the two stubs of a link. Let N_k be the number of nodes of degree k in the network, and there are kN_k stubs of degree k. Sort these stubs by degrees, we can obtain the Spearman rank X_k^0 of a stub with degree k,

$$X_k^0 = \frac{\left(\sum_{i=k_{\min}}^{k-1} iN_i + 1\right) + \sum_{i=k_{\min}}^{k} iN_i}{2} = \sum_{i=k_{\min}}^{k-1} iN_i + \frac{1}{2}kN_k + \frac{1}{2}$$

where $k_{\min}$ is the minimum degree of the network.

In order to make X_k^0 lie in the range $[0, 1]$ with average value 1/2, we make a shift to it and normalize it by the total stubs of the network (see (B.3)). Then the normalized Spearman rank (NSR) of a stub with degree k is defined as,

$$X_k = \frac{X_k^0 - \frac{1}{2}}{\sum_{i=k_{\min}}^{k_{\max}} iN_i} = \frac{\sum_{i=k_{\min}}^{k-1} iN_i + \frac{1}{2}kN_k}{\sum_{i=k_{\min}}^{k_{\max}} iN_i}, \tag{1}$$

where $k_{\max}$ is the maximum degree of the network.

In Fig. 1 and Table 1, we give an example of the Spearman rank and our normalized rank to the stubs in a simple network.