Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 366 (2006) 578–586
PHYSICA A
www.elsevier.com/locate/physa

Complex network properties of Chinese natural science basic research

Jianguo Liu*, Yanzhong Dang, Zhongtuo Wang

Institute of System Engineering, Dalian University of Technology, 2 Ling Gong Rd., Dalian 116023, Liaoning, PR China

Received 13 September 2005; received in revised form 30 September 2005
Available online 28 November 2005

Abstract

In this paper, we studied the research areas of Chinese natural science basic research from a point view of complex network. Two research areas are considered to be connected if they appear in one fund proposal. The explicit network of such connections using data from 1999 to 2004 is constructed. The analysis of the real data shows that the degree distribution of the *research areas network* (RAN) may be better fitted by the exponential distribution. It displays small world effect in which randomly chosen pairs of research areas are typically separated by only a short path of intermediate research areas. The average distance of RAN decreases with time, while the average clustering coefficient increases with time, which indicates that the scientific study would like to be integrated together in terms of the studied areas. The relationship between the clustering coefficient $C(k)$ and the degree k indicates that there is no hierarchical organization in RAN.

Keywords: Complex networks; Power-law distribution; Clustering coefficient; Evolution network; Chinese natural science basic research

1. Introduction

In the past few years there has been a growing interest in the study of complex networks. The boom has two reasons: the existence of interesting applications in several biological, sociological, technological and communications systems and the availability of a large amount of real data [1–11]. Recent works on the mathematics of networks have been driven largely by the observed properties of actual networks and the studies on network dynamics [12–22], optimization [23–28], and evolutionary [29–48]. It also makes sense to examine simultaneously data from different kinds of networks. Recent approaches with methodology rooted in statistical physics focus on large networks, searching for universality both in the topology of the real networks and in the dynamics governing their evolution [49]. These combined theoretical and empirical results have opened unsuspected directions for researches and a wealth of applications in many fields ranging from computer science to biology and sociology [3,4,6,50,51]. In this respect, three important results have been crystallized: First, it has been found that the degree distribution contains important information about the

*Corresponding author. Tel.: +86 13050537943.
E-mail address: liujg004@yahoo.com.cn (J. Liu).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2005.10.036

nature of the network, for many large networks following the exponential distribution and the power-law distributions. Second, most networks have the so-called small world property [2], which means that the average distance between different pairs of nodes is rather small. Third, real networks display a degree of clustering coefficient higher than expected for random networks [2,4]. Finally, the assortative mixing is studied to answer why social networks are different from other types of networks [53].

The scientific studies can be considered as being organized within a network structure, which has a significant influence on the observed study collective behaviors. The viewpoints of complex networks are of interest in studying scientific study networks, to uncover the structural characteristics of the networks built on RAN. In this paper, the research areas of natural science basic research is studied from the point view of complex network [54]. In the fund management department, such as National Natural Science Foundation of China (NSFC), the research areas are denoted by the code system, which have the tree structure to demonstrate the inclusion relation between the research areas, such as Physics $\rightarrow$ statistical physics $\rightarrow$ complex network. The leave codes of the code system always represent the research areas more specially. To make the network reflect the reality more accurately, in this paper, the nodes are defined as the leave nodes of the code system. Because the scientists can fill in the fund proposal two codes: the first application code and the second one, then if one requisition paper filled in two different codes one can consider that the research work is cross the two research areas. The proposals filled in only one code are not considered in RAN. By this definition, there are 371, 349, 367, 400, 456, 544 nodes in RAN from 1999 to 2004. Three complementary approaches allow us to obtain a detailed characterization. First, empirical measurements allow us to uncover the topological measures that characterize the network at a given moment, as well as the time evolution of these quantities. Second, the average distance of RAN decreases with time, which means that the distance between any pairs of research areas is getting short. Third, the average clustering coefficient increases with time, which means that the neighbors of one research area would like to be connected with each other.

This paper is organized as follows: In Section 2, the topological characteristics of RAN, such as the degree distribution, clustering coefficient, average path length, assortative coefficient and the relationship between the clustering coefficient and the degree k are investigated and visualized. In Section 3, the conclusion and discussion are given.

2. Data analysis of RAN

In this section, the topology and dynamics of the empirical network are investigated. The parameters that are crucial to the understanding of the topology of RAN are extracted. The analysis of the data could provide the development trend of Chinese natural scientific basic research system.

2.1. Degree distribution follows the exponential form

The degree distribution $P(k)$ presents the probability that a randomly selected node has k links, which has been studied extensively for various networks. Networks for which $P(k)$ has a power-law tail, are known as scale-free networks [51,55]. On the other hand, classical network models, including the Erdős–Rényi model [56,57] and the Watts and Strogatz model [1] have an exponentially decaying $P(k)$ and are collectively known as exponential networks. Another observed degree distribution form of real-life networks [6,58], named stretched exponential distribution (SED) [59] is of the form $P(x)\,\mathrm{d}x = \mu(x^{(\mu-1)}/x_0^{\mu})\exp(-(x/x_0)^{\mu})\,\mathrm{d}x$ and its accumulative distribution is $P(x) = \exp(-(x/x_0)^{\mu})$, which can be stated as $\ln P(x) \sim x^{\mu}$. Obviously, SED degenerates to an exponential distribution when μ is close to 1 and to a power law when μ is close to 0. When μ is between 0 and 1, the degree distribution is between a power law and an exponential function. The data on a single-logarithmic plane show that the degree distribution of RAN can well fitted of an exponential form, see Figs. 1 and 2.

It seems difficult to find a common function to fit the empirical network because of the fluctuations. However, the accumulative distribution of RAN has less fluctuations and is more stable. The distribution of all the data from 1999 to 2004 can be fitted by an exponential distribution, given by $P(k) = \exp(0.13k)$ (Fig. 3). The hub nodes of RAN and their degrees from 1999 to 2004 are demonstrated in Table 1.