Physica A 387 (2008) 5639–5646

Contents lists available at ScienceDirect

Physica A

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

Geographic coarse graining analysis of the railway network of China

Wang Ru [a,*], Tan Jiang-Xia [a], Wang Xin [a], Wang Du-Juan [b], Cai Xu [a]

[a] *Institute of Particle Physics, Hua-Zhong (Central China) Normal University, Wuhan 430079, PR China*

[b] *College of Physical Science and Electronic Engineering, HuBei University of Education, Wuhan 430060, PR China*

ARTICLE INFO

Article history:
Received 2 February 2008
Received in revised form 28 March 2008
Available online 7 June 2008

PACS:
98.75.HC
98.40.-a
05.10.Cc

Keywords:
Railway network
Statistics physics
Geographic coarse graining

ABSTRACT

We investigate the detailed, empirical analysis of the statistical properties of the Railway Network of China (RNC) in space L and space G, constructed by geographic coarse graining process. The RNC exhibits similar properties in the cumulative distributions of degree and strength in two spaces, and it presents the hierarchical structure, small-world behavior and assortativity, areciprocal connection both in space L and space G. We also investigate the path length that every train runs, the distribution of the railroad length per degree and the optimal distribution of stations.

1. Introduction

In recent years the physics community has paid a lot of attention to the field of complex network science after the works of Watts and Strogatz [1] as well as Barabasi and Albert [2]. A considerable amount of research has been done on different real-world networks, ranging from technological networks [1–4] (electric power grids, internet routers), biological networks [1–4] (chemical reaction networks,neural networks), social networks [1–4] (co-authorship networks, sex network). The transportation networks, such as air [5–7], bus [8], river [9], street [10], subway [11], railway networks [12–14], play crucial roles in people's daily lives and in world economies. Massive numbers of people and quantities of goods move through these networks. These networks were usually visualized by nodes representing stations or spots and by links between them representing their connections. Most previous studies focusing on topology properties of these transportation networks, i.e. non-geographical properties of the networks, such as small-world behavior and scale-free structure [1–4]. In this sense, every edge is of length 1, and the topological distance between two nodes is simply defined as the number of edges along the shortest path connecting them. To ignore the geographical effect is reasonable for some networked systems e.g., food webs [15], citation networks [16], metabolic networks [17], where the Euclidean coordinates of nodes and the lengths of edges have no physical meanings. Yet, many real-life networks, such as transportation networks, the Internet [18], and power grids [19], have well defined node positions and edge lengths and are called spatial or geographical networks [3]. Seldom network topology analysis considers the geographical constraint while it is unreasonable to ignore for these transportation networks. We will fill this gap on the investigation of the geographical constraint in the topology of the railway network.

In this paper, we will present an empirical analysis of railway network of China (RNC). After the construction of the first railway lines in 1876, the rapid development of China railway has now experienced five times acceleration. We collect the RNC data from Internet [20] which comprises the information of 3110 stations and 2252 number of trains in China

* Corresponding author.
E-mail address: wr0124@gmail.com (R. Wang).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.05.052

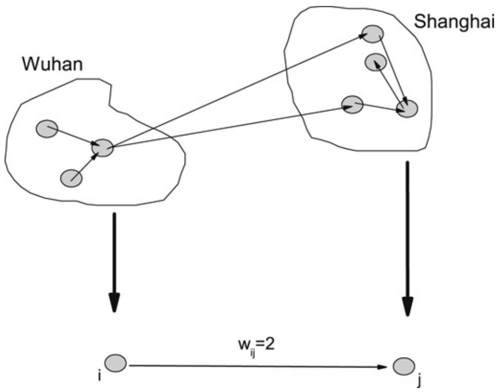

Fig. 1. Illustration of the coarse graining procedure form space L to space G. The three stations in city of Wuhan are merged into one node and four stations in city of Shanghai merge into one node in space G. After merging, the weight between the two nodes in space G is $W_{ij} = 2$.

excluding those in Macao and Taiwan due to the absence of the railway information. Different from other networks, the railway network has the complex properties of its own: (i) The railway network is a directed and weighted network. All the trains running through the two stations are directed and sorted as outgoing and incoming. Since there may have more than one trains passing through two stations, the number of the trains passing through can provide import information about how busy the station or the railway line is. (ii) The topological construction is embedded in a real physical space with nodes having well-defined positions. The balance between the economical cost and the amount of need such as the population density, fluidity of people or goods are all needed to take into consideration when planning to build a new station. In general, the structure of railway network is both a function of what is geographically feasible and what is technologically desirable. So, the topology structure of the RNC is strongly constrained by the information about the geography. This leads us to the idea of investigating two aspects of the railway network in China.

This paper is organized as follows. Section 2 gives the two different definitions of the RNC and the related matrix expression. Section 3 will explore its topological properties. Section 4 will take the weighted network into consideration. Section 5 analyzes the geographical ingredient in the structure of the RNC. The last section is a brief conclusion.

2. The network representation

A lot of work has been done in the definition of space L, which take the station as node and the link exists if there is a train passing through. When measuring the average distance between an arbitrary pair of stations, Sen et al. [21,22] proposed a definition of space P which describes the stations as nodes and two stations are considered to be connected by a link when there is at least one train which stops at both the stations. These two definitions result in slightly different topological properties [21–23].

By the end of 2006, The China administration are composed of 34 province which can be divided further into 333 cities. Inspired by the geographical embedding, we introduce the space G which is constructed by the geographic coarse graining process [24,25]. In contrast to the Kadanoff block spin renormalization group procedure in standard statistical mechanical systems that build on regular square box, we merge stations that are in the same city into a single node; there is a link between two nodes if one train has consecutive stops. In the same way, we can also merge the cities that are in the same province into a single node for the repeat geographic coarse graining, but due to the limited system size, we don't take this further step into consideration. According to the railway information we collected, the 3110 stations can merger into 298 cities, that is, the system size in space G is 298 while the system size is 3110 in space L. The illustration of the coarse graining procedure form space L to space G is presented in Fig. 1. In space L, the degree is the number of train routs that one can take from the chosen node, while in space G, the node degree is the number of train routs that travel to the other cities from the chosen city. For the different administrative regions of the Chinese government, the number of stations is different in distinct cities. So, from space L to space G, the number of nodes and links are decreased by ignoring parts of the local information.

The RNC can be represented as a directed weighted network and we can use connectivity matrix C^L and weight matrix W^L symbolized in space L, and that of C^G and W^G in space G. The entry of connectivity matrix C^L_{ij} is 1 if there is a link pointing from node i to node j and 0 otherwise in space L, and that of the weight matrix W^L_{ij} is the number of links pointing from node i to node j in space L.

We employ $k^L_{in}(i)$ and $k^L_{out}(i)$ to denote the in-degree and out-degree of a given node i in the directed RNC in space L, and $k^L_{all}(i)$ to represent the undirected degree of the undirected RNC. Using the connectivity matrix, we have:

$$k^L_{in}(i) = \sum_{i \neq j} \eta(C^L_{ij} - 1) \tag{1}$$