

与地理环境相关的中国铁路客运网拓扑结构^{*}

谭江峡¹⁾ 王杜鹃²⁾ 王鑫¹⁾ 王茹^{1)†} 蔡勖¹⁾

1) 华中师范大学 物理科学与技术学院, 武汉 430079)

2) 湖北教育学院, 物理与电子工程系, 武汉 430060)

(2008 年 3 月 5 日收到, 2008 年 4 月 2 日收到修改稿)

以中国铁路车站作为“节点”, 每辆列车经过的相邻两个停靠车站之间连接一条“边”, 构成有方向有权重的中国铁路客运网. 首先研究了该网络的拓扑结构, 包括连接度、聚集系数、最短路径和强度, 结果表明中国铁路客运网的连接度分布、强度分布都是介于指数分布和幂率分布之间, 是一个具有小世界性质的阶层网络. 修建铁路需考虑人口分布、行政区域等因素. 铁路固定设施成本高, 修建完成后很难做变动, 因此需考虑诸多空间地理环境对中国铁路客运网的影响, 如站点的连接度和站点的相连站点之间的平均行驶距离之间的关系、车站的分布密度与人口密度的关系、省市集团的强度等. 这些结果表明中国铁路客运网的修建也是受到地理环境约束的, 呈现出优化性.

关键词: 铁路客运网, 拓扑统计, 小世界, 地理环境

PACC: 0520, 0560, 0630C

1. 引言

近年来复杂网络的研究引起了各行各业的兴趣. 如用复杂网络的方法研究因特网^[1], 食物网^[2], 化学反应网^[3], 人类语言网^[4]等等. 交通网(铁路网、公交网、航空网)在人们日常生活和经济发展中扮演着至关重要的角色. 大量的人流和物流通过交通网在不同地区之间进行流通. 因此, 交通网络的研究既具有理论的重要性, 又包含实际应用的良好前景, 为许多科技工作者所重视. 印度铁路网^[5]、波士顿和维也纳地铁网^[6,7]、波兰^[8]、中国^[9,10]和世界^[11]若干城市内的公共交通网, 以及中国^[12,13]、美国^[14]和世界^[15]航空网都已经取得了一定研究成果.

本文从 [www : \ huoche.com](http://www.huoche.com) 网站上收集包含 32 个省、直辖市、自治区(不包括台湾省、澳门特区)的中国铁路客运网信息, 以火车车站作为“节点”, 每辆列车经过的相邻两个停靠车站之间连接一条“边”, 构成了 $N = 3110$ 个结点, $E = 27531$ 条边的有方向有权重的网络. 首先研究该网络的拓扑结构性质, 包括连接度、聚集系数、最短路径和强度, 然后研究在受到地理环境影响下的网络性质, 如站点的连接度和

站点的相邻站点之间的平均行驶距离之间的关系, 车站的分布密度和省市的集团强度等.

2. 中国铁路客运网网络拓扑结构

使用 $k_{in}(i)$, $k_{out}(i)$ 和 $k_{all}(i)$ 代表结点 i 的入、出、总的连接度, 可以得到入、出、总连接度分布. 为了减少由于有限数据量引起的统计误差, 我们使用在经济中广泛采用的累积分布. 累积度分布 $P_{cum}(k) = \int_k^\infty P(k) dk$ ^[16] 如图 1 所示, 在双对数坐标下, 累积度分布拟合后的曲线表现为 SPI(shifted power law)^[17] 具有如下形式:

$$P_{cum}(k) \propto (k + \alpha)^{-\eta}, \quad (1)$$

或者写为等价形式 $\log(P_{cum}(k)) \propto (-\eta) \times \log(k + \alpha)$. 由图可以看到, 三种分布都可以用相同的关键指数 $\alpha = 3.6$ 和 $\eta = 3.5$ 拟合, 该分布是典型的 SPL

分布. 中国铁路客运网的平均连接度 $k = \frac{\sum_i k_i}{N} = 3.41$ (见表 1), 即平均每个车站只有 3 个邻车站.

车站的连接度反应了该车站与其他车站联系的紧密程度, 车站的聚集系数则反应了该车站的邻车

^{*} 国家自然科学基金(批准号: 70501027, 70401020, 10647125, 10635020)和国家教育部基金(批准号: 306022)资助的课题.

[†] 通讯联系人. E-mail: wr0124@gmail.com

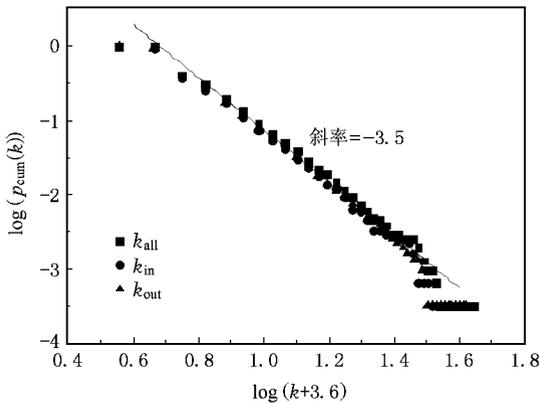


图1 中国铁路客运网的累计入、出、总连接度分布

站之间联系的紧密程度. 节点 i 的聚集系数 $c_i = \frac{2l_i}{k_i(k_i - 1)}$, k_i 是节点 i 的连接度, l_i 是这 k_i 个相邻节点之间相互连接的边数. 图 2 显示了聚集系数 c 和 k 的关系, 拟合得到

$$c(k) \propto k^{-\lambda}, \quad (2)$$

其中 $\lambda = 0.8$. 聚集系数的这种标度行为表明该网络具有层次结构^[18-20]. 平均聚集系数 $c = 0.28$ (见表 1) 表明中国铁路客运网中车站之间的相互联系并不紧密. 通过统计知道, 在 3310 个车站中有 1551 个车站的聚集系数为 0, 说明大约一半的车站只有一辆列车通过. 将中国铁路客运网的平均聚集系数与相同连接度和系统大小的随机图的平均聚集系数 $c_{rand} = \frac{k}{N} = 0.001$ 比较, 发现要远远大于随机图的平均聚集系数, 为小世界效应性质之一.

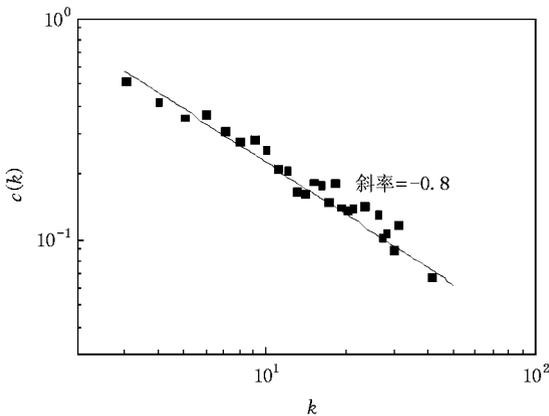


图2 中国铁路客运网的聚集系数 c 作为连接度 k 的函数

小世界效应另外一性质就是具有较小的平均最短路径^[21]. 用路径 d_{ij} 表示从节点 i 到节点 j 所

要经过的边数, i 和 j 之间的最短路径 $d_{\min}(ij)$ 是 d_{ij} 中的最小值. 图 3 反映的是中国铁路客运网的最短路径分布. 网络的平均最短路径 $L = \frac{\sum d_{\min}(ij)}{N(N-1)} = 11.3$ (见表 1), 即网络中任意两个结点之间平均要经过 11 条连线. 将中国铁路客运网的平均最短路径与随机图的 $L_{rand} = \frac{\ln N}{\ln k} = 6.6$ 进行比较可知, 中国铁路客运网的平均最短路径和随机网的相当, 由上可知中国铁路客运网具有较大的平均聚集系数和较小的平均最短路径, 即小世界效应.

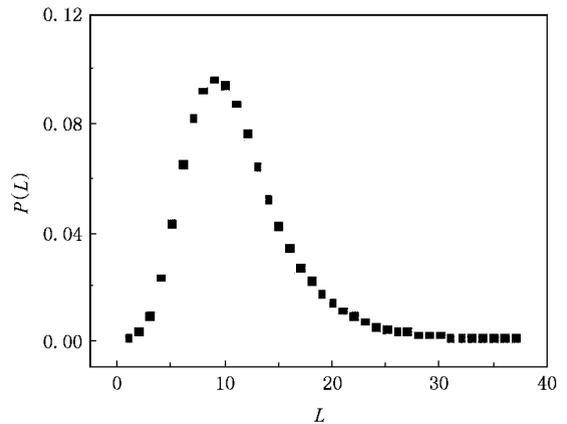


图3 中国铁路客运网的最短路径分布

对于中国铁路网, 每条铁轨不仅仅只有一辆列车通过, 也会有多辆列车通过, 所以使用权重矩阵 W_{ij} 来表示通过结点 i 到结点 j 的列车次数. 我们用强度 s_i 代表从结点 i 到结点 j 的流量 $s_i = \sum_j w_{ij}$ 和连接度类似, 在考虑到方向性下, 我们用 s_{in}, s_{out}, s_{all} 表示结点的入、出、总强度, 得到如图 4 所示的累积强度分布. 累积强度分布介于指数和幂率分布之间, 可以由 Laherrere 和 Sornette 在 1998 提出的 stretched exponential distribution (SED)^[22] 较好拟合:

$$P(x) dx = c \left(\frac{x^{c-1}}{x_0^c} \right) \exp \left[- \left(\frac{x}{x_0} \right)^c \right] dx, \quad (3)$$

其累积分布为

$$P_{cum}(x) = \exp \left[- \left(\frac{x}{x_0} \right)^c \right], \quad (4)$$

也可以写为 $\log[-\log P_{cum}(x)] \propto c \log(x)$. 当 c 趋于 1 时接近指数分布, 当 c 趋于 0 时趋于幂率分布, 这里的三种分布的 c 都等于 0.55.

累积强度分布是典型的 SED 分布而累积连接

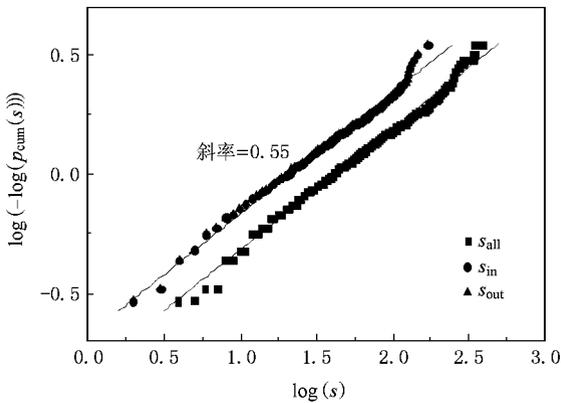


图4 中国铁路客运网入、出、总累积强度分布

度是典型的 SPL 分布,两者之间有较大差异,为了更清楚探究强度的内在含义,我们探讨了强度作为连接度的函数图,如图5所示.图5显示中国铁路客运网的强度 s 和 k 的关系拟合得到

$$s(k) \propto k^\alpha, \quad (5)$$

其中 $\alpha = 1.41$. 平均强度 $s = 17.70$ (见表1).表1列举了中国铁路客运网的一些物理量,如系统大小 N 、平均连接度、平均聚集系数、平均最短路径、平均强度.

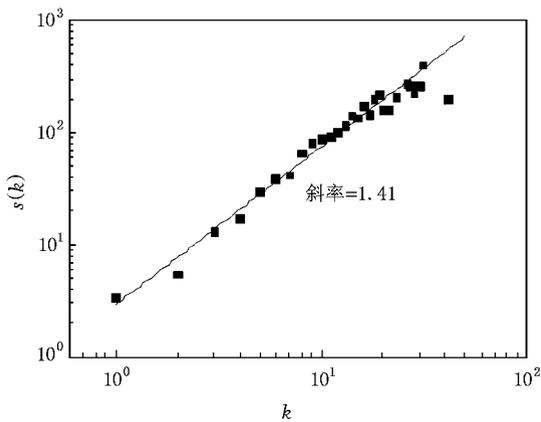


图5 中国铁路客运网的强度 s 作为连接度 k 的函数

表1 中国铁路客运网特征量

物理量	中国铁路客运网
N	3310
k	3.41
c	0.28
L	11.34
s	17.70

3. 地理环境对中国铁路客运网的影响

由于铁路固定设施成本高,修建车站时需考虑人口流动需要,而行政区的划分必定使人出行和物资流通有区域性,因此现实地理环境可能会造成网络中车站的连接有偏好.下面主要从这几个方面来研究空间地理环境对中国客运铁路网的影响:站点的连接度和站点的相邻站点之间的平均行驶距离之间的关系,车站的分布密度和省市的集团强度.

当节点的度增大时,网络的总地理长度也会随着变化^[23].车站的度是车站对应节点在拓扑空间的特征量,每个车站和邻车站通过车轨相连,车站和邻车站之间的行驶距离反应了该车站的真实空间的地理特征,故而我们考虑这两个量之间的联系. r_{ij} 是列车从车站 i 到行驶车站 j 的地理距离,定义 $R_i = \sum_j r_{ij}$ 是与经过 i 车站的所有列车的到相邻车站的行驶距离之和,则 $G_i = \frac{R_i}{k_i}$ 反映了车站 i 与邻车站之间的平均行驶距离.图6为每个车站平均行驶距离分布,该分布也是具有 SPL 行为,表明中国铁路客运网中大部分车站相距较近,只有少数车站距离其他车站较远.车站近似均匀地分布在地理版图中,这有利于人们出行时选择车站.

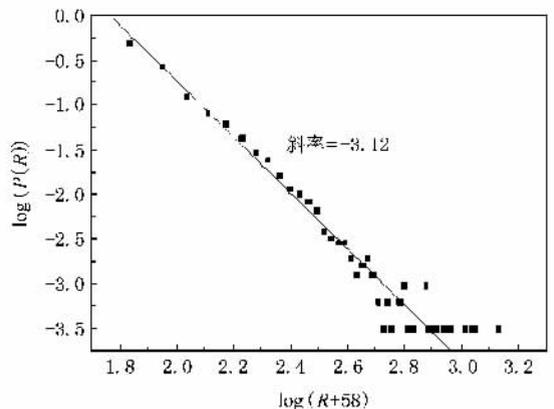


图6 中国铁路客运网车站平均行驶距离分布

图6是对每个车站的相邻站点的行驶距离和连接度比值的分布图,为了更深入探究每个车站的连接站点的行驶距离和连接度之间的关系,我们探讨连接度和行驶距离之间的函数关系.假设 $R_i(k)$ 为度为 k 的 i 车站所经过的所有列车到相邻车站的总

行驶长度,而连接度为 k 的车站有 n 个, $R(k) =$

$$\sum_{i=1}^n R_i(k). \text{ 如图 7 在双对数坐标拟合得到}$$

$$R(k) \propto k^\beta, \quad (6)$$

其中 $\beta = 1.84$. $R(k)$ 随着度的增大而迅速的增加,说明车站的连接度越大,到相邻车站的平均行驶距离越长,查看原始数据,发现连接度较大的车站是重要的大站,会有一些直达火车运行,所以导致较大的相邻车站平均行驶距离.

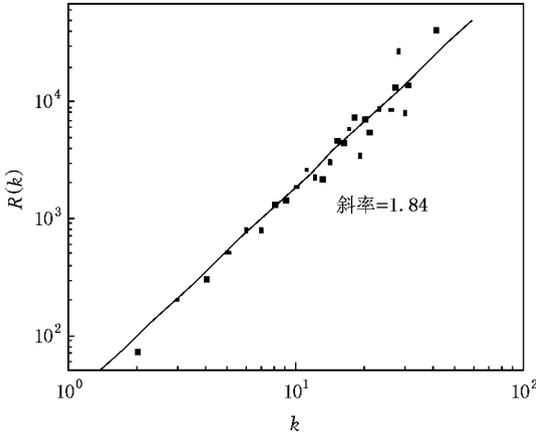


图 7 中国铁路网 $R(k)$ 作为 k 的函数图

车站是公共交通设施,因此修建车站时要考虑车站密度的分布是否能够最大化满足成本和需求之间的平衡. Newman 研究认为,如果将网络按一定的规则分成不同地区^[24],设施的最优分布是设施的分布密度 ρ_f 与人口分布密度 ρ_p 的关系满足关系式: $\rho_f \propto \rho_p^{2/3}$. 基于这个思想,统计中国铁路客运网数据涉及到的中国城市(299个)和省市(32个)的车站分布密度和人口分布密度,分析车站密度 ρ_f (个/ km^2) 和人口密度 ρ_p (人/ km^2) 之间的关系,如图 7 所示. 在双对数坐标下将图中数据点与最优分布的虚线 $\rho_f \propto \rho_p^{2/3}$ 做比较,发现中国火车站分布接近最优,数据散落在斜率为 0.95 的直线附近.

分析表明,中国铁路客运网受现实世界的影响也存在按照行政区划分的集团,即集团内部的连线密度要远远大于集团之间的连线密度^[25]. 每个省市(由于网站没有台湾和澳门的车站信息,故而没有考虑这两个区域)成为一个集团,这样中国铁路客运网可划分成 32 个集团. 每个集团内有多个车站,集团内的车站相互之间有连接,和其他集团内的车站也有连接. 用 Q_s 表示省份集团强度,对于集团 S ,集团

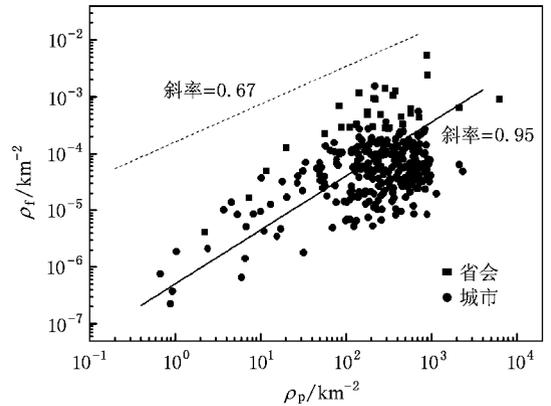


图 8 省市车站密度与人口密度关系

强度 $Q_s = \frac{L_s}{L} - \left(\frac{d_s}{L}\right)^2$, 其中 L_s 是集团 S 内节点之间边的总数, d_s 是集团 S 内节点的连接度之和, L 是网络的总边数. 集团强度反映了集团内部与外部之间的联系紧密程度,集团强度越大,集团相对与外界交流减少. 表 2 列出了集团强度最大的前 10 个省市,数据表明省份的集团强度普遍不大,而且强度较大的省市的车站数量也较大,如表 3 所示. 表明对于中国铁路客运网,因为省内出行距离较短,行人不偏向于选择火车出行,而是更喜欢乘坐灵活性较大的汽车,故而省市的集团强度不会大,另一方面,虽然如北京市、河南省由于拥有全国最大的车站站点北京、郑州和徐州,并且这些站点的连接度也是全国所有车站中最大的, $k_{\text{北京}} = 42$, $k_{\text{郑州}} = 22$, $k_{\text{徐州}} = 19$,但是由于这些地区的车站数量并不是最多,故而其集团强度排名不是很前. 总体来说,集团强度更多的是受到集团内车站数量的影响,然后会在此基础上受到中心枢纽车站的影响.

表 2 前 10 个最大集团强度的省市及其对应的车站数量列表

省市	Q_s	车站数量 N_s
黑龙江省	0.09499	313
内蒙古	0.06872	303
辽宁省	0.06028	204
吉林省	0.06013	218
四川省	0.05599	233
河北省	0.05476	200
山西省	0.03801	145
贵州省	0.03471	157
河南省	0.0334	104
陕西省	0.03123	111

4. 结 论

本文对有方向有权重的中国铁路客运网进行研究,分析了该网络拓扑结构和与地理环境相关的一些物理量.中国铁路客运网具有较大的平均聚集系数和较小的平均最短距离,是一个具有等级性的小世界网络,这些结果和中国铁轨网^[26]的结论一致,但是中国客运网的累积连接度分布和累积强度分布都是介于指数和幂率之间的分布,不同于文献^[26]中的幂率分布,这些差异是由于中国客运网是建立

在中国铁轨网上,列车行驶和铁轨分布有所不同,而且本文只是考虑了有停靠的车站,一些路过并没有停靠的小站并不在本文考虑范围之类.铁路是公共交通设施,与地理环境,人口分布,行政区域划分有密切联系,考虑这些因素对客运网的影响.研究表明,中国铁路客运网中的车站连接度越大,与之连接的路段的行驶距离就越长,大部分的车站相距较近,只有少数车站与其他车站需要较长的车轨来连接,通过研究车站密度和人口密度的关系,发现行政区域划分对于中国铁路的分布也有影响,使得车站的分布接近最优.

- [1] Barabasi A L ,Albert R 1999 *Nature* **401** 130
- [2] Williams R J ,Martinez N D ,Berlow E L ,Dunne J A ,Barabasi A L 2002 *Ecology Letters* **7** 558
Montoya J M 2002 *Journal of Theoretical Biology* **2** 405
- [3] Ito T ,Chiba T ,Ozawa R ,Yoshida M ,Hattori M ,Sakaki Y 2001 *National Acad Sciences* **3** 12
- [4] Markosova M ,Nather P 2006 *Sixth International Conference on Hybrid Intelligent Systems (HIS'06)* 37
- [5] Sen P ,Dasgupta S ,Chatterjee A 2003 *Phys. Rev. E* **67** 36106
- [6] Latora V ,Marchiori M 2002 *Physica A* **314** 109
- [7] Seaton K A ,Hackett L M 2004 *Physica A* **339** 635
- [8] Sienkiewicz J ,Holyst J A 2005 *Phys. Rev. E* **72** 046127
- [9] Zhang P ,Chen K ,He Y ,Zhou T ,Su B ,Jin Y ,Chang H ,Zhou Y P ,Sun L C ,Wang B H ,He D R 2006 *Physica A* **360** 599
- [10] Chen Y Z ,Li N ,He D R 2007 *Physica A* **376** 747
- [11] Ferber C V ,Holovatch T ,Holovatch Y ,Palchykov V 2006 eprint : physics/0608125
- [12] Li W ,Cai X 2004 *Phys. Rev. E* **69** 046106
Liu H K ,Zhou T 2007 *Acta. Phys. Sin.* **56** 106 (in Chinese) [刘宏鲲、周 涛 2007 物理学报 **56** 106]
- [13] He Y ,Zhu X ,He D R 2004 *Int. J. Mod. Phys. B* **18** 2595
- [14] Chi L P ,Wang R ,Su H *et al* 2003 *Chin. Phys. Lett.* **20** 1393
- [15] Guimera R ,Mossa S ,Turtschi A ,Amaral L A N 2005 *Proc. Natl. Acad. Sci.* **102** 7794
- [16] Newman M E J 2005 *Contemporary Physics ,Taylor and Francis Ltd* **46** 323
- [17] Fu C H ,Zhang Z P ,Chang H ,Tao J R ,Chen Z H ,Dai Y L ,Zhang W ,He D R 2008 *Physica A* **387** 1411
- [18] Ravasz E ,Somera A L ,Mongru D A *et al* 2002 *Science* **297** 1551
- [19] Yan G ,Fu Zg Q ,Ren J ,Wang W X 2007 *Phys. Rev. E* **75** 016108
- [20] Muff S ,Rao F ,Caffisch 2005 *Phys. Rev. E* **72** 056107
- [21] Zhao W ,He H S ,Lin Z C ,Yang K Q 2006 *Acta. Phys. Sin.* **55** 3906 (in Chinese) [赵 伟、何红生、林中才、杨孔庆 2006 物理学报 **55** 3906]
- [22] Laherrere J ,Sornette S 1998 *Eur. Phys. J. B* **2** 525
- [23] Xie Y B ,Zhou T ,Bai W J ,Chen G R ,Xiao W K 2007 *Phys. Rev. E* **75** 036106
- [24] Michael T G ,Newman M E J 2006 *Phys. Rev. E* **74** 016117
- [25] Newman M E J 2006 *Proc Natl Acad Sci.* **103** 8577
Du H F ,Li S Z ,Marcus W F ,Yue Z S ,Yang X S ,2007 *Acta. Phys. Sin.* **56** 6886 (in Chinese) [杜海峰、李树茁、Marcus W F、悦中山、杨绪松 2007 物理学报 **56** 6886]
Li K P ,Gao Z Y 2007 *Chin. Phys.* **16** 2304
Wang R ,Chi L P ,Cai X 2008 *Chin. Phys. Lett.* **25** 1502
- [26] Li W ,Cai X 2007 *Physica A* **382** 693

The topological railway network of China constrained by the geographic factors^{*}

Tan Jiang-Xia¹⁾ Wang Du-Juan²⁾ Wang Xin¹⁾ Wang Ru^{1)†} Cai Xu¹⁾

¹ *College of Physical Science and Technology, Huazhong (Centre China) Normal University, Wuhan 430079, China*

² *College of Physical Science and Electronic Engineering, Hubei University of Education, Wuhan 430060, China*

(Received 5 March 2008 ; revised manuscript received 2 April 2008)

Abstract

We made a detailed empirical analysis of the railway network of China (RNC), in which the stations are treated as nodes and a link exists between every two nodes if there is a train stop at each node. Firstly, we explore the statistical properties of the directed and weighted RNC with respect to the distribution of degree and strength, clustering coefficient and the shortest path length. The results exhibit that the distribution of degree and strength are between the exponential distribution and scale-free. Moreover, the RNC is a hierarchical small-world network. Secondly, the construction of the railway is the result of balance between the economical cost and the degree of need such as the population density and the number of administrative divisions. So the topological properties constrained by the geographic factors are taken into consideration, such as the relation between railroad length and degree, the density distribution of the stations and the population and the community strength of the province. These results proved that the RNC is constrained by the geography and approaches to the optimal construction.

Keywords : railway network of China, topology statistics, small world, geographic constraint

PACC : 0520, 0560, 0630C

^{*} Project supported by the Ministry of Education of China (Grant No. 306022) and the National Natural Science Foundation of China (Grant Nos. 70501027, 70401020, 10647125 and 10635020).

[†] Corresponding author. E-mail : wr0124@gmail.com