

基于社团结构的城市地铁网络模型研究*

丁益民^{1)†} 丁卓²⁾ 杨昌平¹⁾

1) (湖北大学物理学与电子技术学院, 武汉 430062)

2) (华南理工大学工商管理学院, 广州 510640)

(2012年7月30日收到; 2012年12月6日收到修改稿)

本文运用复杂网络理论, 对我国北京、上海、广州和深圳等城市的地铁网络进行了实证研究. 分别研究了地铁网络的度分布、聚类系数和平均路径长度. 研究表明, 该网络具有高的聚类系数和短的平均路径长度, 显示小世界网络的特征, 其度分布并不严格服从幂律分布或指数分布, 而是呈多段的分布, 显示层次网络的特征. 此外, 它还具有重叠的社团结构特征. 基于实证研究的结果, 提出一种基于社团结构的交通网络模型, 并对该模型进行了模拟分析. 模拟结果表明, 该模型的模拟结果与实证研究结果相符. 此外, 该模型还能解释其他类型的复杂网络(如城市公共汽车交通网络)的网络特性.

关键词: 复杂网络, 地铁网络, 小世界, 社团

PACS: 89.75.-Hc, 89.75.-k, 02.50.-r, 05.10.-a

DOI: 10.7498/aps.62.098901

1 引言

复杂网络是一种用来描述自然、社会以及工程技术中相互关联的理论. 近 10 年来, 复杂网络的研究受到了来自各个领域的研究人员的重视^[1-4]. 在复杂网络的研究中, 聚类系数、平均路径长度和度分布是描述复杂网络的特征物理量. 大量的实证研究表明, 许多复杂网络同时具有高的聚类系数和短的平均路径长度, 呈现小世界特征^[5]. 还有一些复杂网络的度分布满足幂律分布, 呈现无标度网络的特征^[6]. 人们还发现有些网络(如蛋白质网络)还具有模块化的结构特征^[7,8]. Song 等把盒计数法推广用于研究复杂网络, 指出对于许多复杂网络也存在类似于分数维的自相似指数, 从而也具有某种内在的自相似性^[9]. Palla 等指出自然界和人类社会中的许多复杂网络都可描述为具有重叠连接的社团结构^[10]. 最近, Clauset 等指出层次结构是建构复杂网络最基本的组织原则之一, 进一步揭示了复杂网络具有层次结构和社团结构的特征^[11].

在复杂网络的研究初期, 人们主要研究的是复

杂网络的拓扑特性, 对于空间网络 (spatial network) 则较少涉及^[4], 在这些网络中, 网络的拓扑结构要受到空间的限制, 例如, 在空间网络中, 长距离的连接要受到几何距离的限制. 我们知道, 不论是 WS 小世界网络还是 BA 无标度网络, 都没有考虑空间对网络演化的限制. 近年来, 人们对航空网络^[12-15]、铁路网络^[16,17]、城市公共汽车交通网络^[18-21] 和城市地铁网络^[22-25] 等空间网络进行了大量的实证研究, 研究发现, 这些网络显示出一些不同的特性. Li 等对中国航空网络进行了实证研究, 研究表明, 中国航空网的度服从所谓的“双段”幂律分布, 而该特征无法用先前提出的各类模型解释^[12]. 最近钱江海等的研究表明, 中国航空网与外部经济环境, 即与国民生产总值 (GDP) 具有相关性, 为此, 提出了一种耦合指数增长节点的演化模型, 解释了中国航空网的双段幂律度分布现象^[15].

大量的实证研究表明, 航空网络、铁路网络、城市公共汽车交通网络和城市地铁网络等空间网络都显示出小世界网络的特征, 即具有高的聚类系数和短的平均路径长度. Sen 等对印度铁路网络进行了研究^[16], 研究表明, 印度铁路网络的度分布服

* 国家自然科学基金(批准号: 11074067)资助的课题

† 通讯作者. E-mail: dymhubu@sina.com

从指数规律, 并具有高的聚类系数 ($C = 0.69$) 和短的平均路径长度 ($L = 2.16$), 是一种典型的小世界网络. Seaton 等对波士顿地铁网络的研究表明^[23], 该网络也显示出小世界网络的特征 ($C > 0.9, L < 2.0$). Sienkiewicz 等对波兰 22 个城市的公共汽车交通网络进行了全面的实证研究, 研究表明, 度分布既有服从幂律分布, 也有服从指数分布, 但都显示出小世界网络的特征. 近年来, 我国学者在公共汽车交通网络的领域进行了一系列的研究. Chen 等对北京、上海、南京和杭州等四大城市的公共汽车交通网络进行了研究, 研究表明, 度分布服从指数分布, 并且都显示出小世界网络的特征 ($C > 0.7, L < 3.0$). Yang 等研究表明, 公共汽车交通网络具有层次结构特征, 并提出了一个网络模型对北京、上海和杭州实证研究结果进行了解释^[20]. Ding 等对北京、广州、武汉和珠海等不同规模的城市公共汽车交通网络进行了实证研究, 研究表明, 这些网络的度分布并不是严格地服从幂律分布或指数分布. 为此, 提出了一个类似于适应度模型的网络模型对实证研究结果进行了解释^[21].

随着社会经济的发展, 城市规模越来越大, 城市的交通问题成为阻碍城市发展的的重要因素之一. 为了改变城市的交通状况, 作为城市最便捷的交通工具地铁在我国得到了快速地发展. 我国北京、上海等城市的地铁网络已成为世界上规模最大的地铁网络之一, 因此, 研究我国城市地铁系统的网络结构与动力学特性, 不论在理论上还在现实中, 都具有重要的意义. 本文首先对我国北京、上海、广州和深圳等四大城市的地铁网络的度分布、聚类系数和平均路径长度等网络特性进行实证研究, 并在此基础上, 结合近年来人们对复杂网络社团结构的认识, 提出了一个基于社团结构的网络模型, 并通过数值模拟对实证研究结果进行解释.

2 实证研究

2.1 建模方法

一个具体的复杂网络可抽象为一个由点集 V 和边集 E 组成的图 $G = (V, E)$. 节点数记为 $N = |V|$, 边数记为 $M = |E|$. E 中每条边都有 V 中一对点与之相对应. 一个地铁网络也可以抽象为一个图. 地铁站点可定义为节点, 但节点间边的定义却存在着不同的描述方法. 根据站点连接关系一般可将地铁网络的描述方法归结为两种^[17]: 一种是被称为 L

空间 (Space L) 方法, 即地铁站点视为节点, 若两个站点在某一条地铁线路上是相邻的, 那么它们就有边相连; 另一种是 P 空间 (Space P) 方法, 即地铁站点视为节点, 若两个站点有直达地铁线路, 那么它们就有边相连. 图 1 给出了用两种不同描述方法表示的一个简单的地铁网络. 很显然 Space L 方法构造的网络是 Space P 方法构造网络的子网络. Space L 方法构造的网络反映了地铁站点之间的地理联系, 保留了地铁网络基本的拓扑结构特性, 而 Space P 方法构造的网络则可以很好地反映地铁网络的换乘状况. 由于地铁是一个快捷的交通工具, 认为具有直达地铁线路的站点有边相连在地铁网络中具有明显的合理性. 因此, 在本文中, 我们将采用 Space P 方法对我国城市地铁网络进行研究.

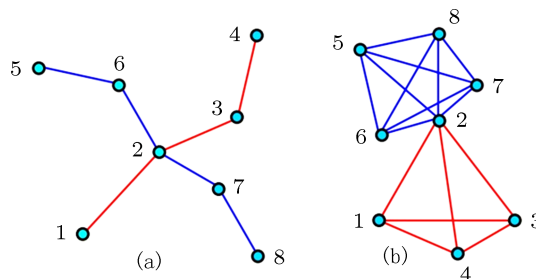


图 1 地铁网络拓扑结构图 (a) Space L : 地铁站点为节点, 若两个站点在某一条地铁线路上是相邻的, 则它们就有边相连, 节点的度即为地铁站点上地铁运行线路的方向数; (b) Space P : 地铁站点视为节点, 若两个站点有直达地铁线路, 那么它们就有边相连, 节点的度即为地铁站点上能直达的地铁站点数

2.2 统计参量

2.2.1 度分布

度是描述节点在网络中重要性的物理量. 节点 i 的度 k_i 定义为与此节点连接的边的数目, 网络中所有节点 i 的度 k_i 的平均值称为网络平均度, 定义为 $\langle k \rangle$. 网络中节点的度分布用分布函数 $p(k)$ 表示, 描述的是一个任意选择的节点的度恰好为 k 的概率. 为了减小统计起伏和涨落, 通常采用累积度分布 $P(k) = \sum_{k=k'}^{\infty} p(k')$ 来描述网络中站点的度分布状况, 本文将研究网络的累积度分布.

2.2.2 聚类系数

聚类系数是表征复杂网络的特征物理量. 假设网络中的一个节点 i 有 k_i 条边将它和其他节点相连, 这 k_i 个节点就称为节点 i 的邻居, 而聚类系数 c_i 是指它所有邻居节点之间实际存在的边数 e_i 和总

的可能的边数目 $k_i(k_i - 1)/2$ 之比, 即

$$c_i = \frac{2e_i}{k_i(k_i - 1)} = \frac{\sum_{j,m} a_{ij} a_{jm} a_{mi}}{k_i(k_i - 1)}, \quad (1)$$

式中, a_{ij} 为网络的连接矩阵. 网络中所有节点 (节点总数为 N) 的聚类系数的平均值称为网络的聚类系数, 定义为 C , 即

$$C = \langle c \rangle = \frac{1}{N} \sum_{i \in N} c_i. \quad (2)$$

2.2.3 平均路径长度

平均路径长度是表征复杂网络小世界特性的物理量. 网络中两个节点 i 和 j 之间的距离 d_{ij} 定义为连接这两个节点的最短路径上的边数. 网络中任意两个节点之间距离的最大值称为网络的直径 D , 网络平均路径长度 L 为所有节点对之间距离的平均值

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij}, \quad (3)$$

式中, N 为网络节点总数.

2.3 数据分析

根据上一节的定义和描述, 我们对 2010 年北京、上海、广州和深圳等四城市的地铁网络 [26,27]

在 p 空间中的统计参量进行研究. 分别计算其网络节点数、边数、平均度、特征路径长度、网络直径和聚类系数以及和所研究的网络具有相同数目的节点和边的随机网络的平均路径长度和聚类系数. 计算结果如表 1 所示.

根据表 1 的计算结果可知: 2010 年北京市的地铁系统由 10 条地铁线路、124 个地铁站点组成. 在 Space P 中, 最大的节点度 k_i 是 55, 平均度为 20.64, 即无需换乘平均可直接到达其他 20.64 个站点. 聚类系数 $C = 0.9434$, 具有很高的聚类系数 (接近 1), 这是由于地铁网络是由若干个聚类系数为 1 的高聚类社团 (每一条地铁线路可看成一个全局耦合网络子图) 组合而成. 平均路径长度 $L = 2.364$, 网络直径 $D = 4$, 即任意两个站点之间平均转车次数不到两次, 但也存在着从一个站点到另外一个站点至少需要转车 3 次才能到达的情形. 与同种规模的 ER 随机网络 ($C_{ER} = 0.1249$, $L_{ER} = 2.096$) 相比, 北京市地铁网络同时具有高的聚类系数和短的平均路径长度, 显示出小世界网络的特征. 对我国上海、广州和深圳等其他几个城市地铁网络的实证研究, 也得出了相同的结论: 地铁网络都具有高的聚类系数和短的平均路径长度, 即具有小世界效应.

表 1 四个城市地铁网络统计参数表 (2010)

网络参数及特征量	北京	上海	广州	深圳
城市人口 I	12	15	7	5.5
站点数 N	133	234	123	121
线路数 M	10	12	8	5
网络直径 D	4	4	4	2
平均度 $\langle k \rangle$	20.64	29.20	19.98	28.63
平均路径长度 L	2.449	2.604	2.420	1.631
聚类系数 C	0.9434	0.9213	0.9507	0.9687
随机网络 L_{ER}	2.096	1.9798	2.116	2.140
随机网络 C_{ER}	0.1249	0.1108	0.1317	0.1292

四大城市在 Space P 中的累积度分布情况如图 2 所示 (星号表示实证研究数据).

从图中可以看出, 初看起来, 地铁网络的累积度分布在半对数坐标上可近似看成一条直线, 表明地铁网络节点度分布服从指数分布, 其中上海地铁网络的累积度分布规律较为明显, 这可能是因为上海地铁网络规模较大. 我们知道, 如果网络增长时的连边原则是随机连接时, 节点的度分布是指数型

的分布, 这意味着在城市地铁网络中新增站点与已有站点之间的连接可视为随机连接. 这是因为, 受空间及工程技术的限制, 不宜建造高密度线路的地铁站点. 站点度呈随机分布更具合理性. 然而, 仔细观察, 不难发现地铁网络的累积度分布具有分段分布的规律. 这一方面说明地铁网络具有层次结构的特征, 另一方面是由于地铁网络规模太小统计出现明显涨落的缘故, 从而导致地铁网络与 ER 随机网

络、WS 小世界网络以及 BA 无标度网络有所不同. 我国四大城市地铁网络的实证研究结果, 对我国其他城市地铁网络的建构具有良好的指导意义, 首先, 网络的随机特性决定了不宜建立地铁线路太多的地铁总站点, 地铁站点应相对均匀地分布, 以避免

地铁站点发生交通拥堵; 其次, 网络构建可采用层次化的结构方式, 这样有利于保证地铁网络中各站点的连通性. 为了理解地铁网络演化过程, 下一章我们将引入一个网络模型对地铁网络的特征进行模拟研究.

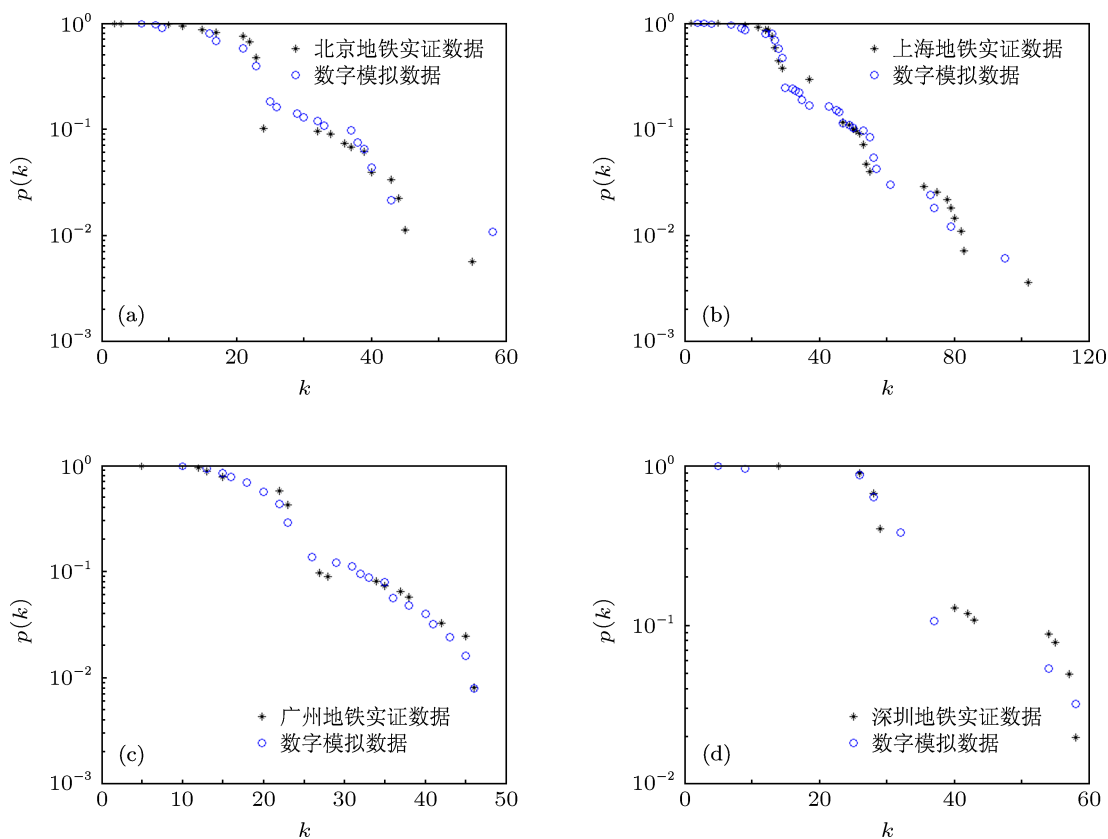


图2 四大城市地铁网络在半对数坐标上的累积度分布图 (a) 北京地铁网络; (b) 上海地铁网络; (c) 广州地铁网络; (d) 深圳地铁网络

3 网络模型

3.1 模型

基于对以上四大城市地铁网络实证研究的分析, 不难发现, 地铁网络的演化过程与 BA 无标度网络有所不同, 网络的增长不再是在原网络中, 通过引入一个新的节点并增加连边产生, 而是依次以原网络中的某一个 (或几个) 节点为起点构建一个新的全局耦合网络 (高聚类社团) 而形成. 下面我们引入一个基于社团结构的小世界网络模型, 该网络模型的构造算法如下 (如图 3):

1) 开始: 初始时刻 $t = 0$, 从一个具有 n_1 个节点, 每个节点具有 $n_1 - 1$ 条边的全局耦合网络开始.

2) 选择: 每个时间步, 以概率 p 随机地在网络中选择 m_1 个节点作为新加入的全局耦合网络的起点.

3) 生长: 以选择的 m_1 个节点为起点构建一个新的具有 n_2 个节点, 每个节点具有 $n_2 - 1$ 条边的全局耦合网络, 以此类推, 网络生长演化.

下面我们考虑网络模型的一种特殊情况. 当 $m_i = 1, n_1 = n_2 = \dots = n_i = 4, p = 1$ 时, 网络的演化过程如图 4 所示. 在这种情况下, 网络结构显示出与新陈代谢网络 (参见文献 [7, 8]) 相类似的由模块以某种迭代的方式生成的等级网络 (hierarchical organization of modularity). 可见这种网络模型与近年人们对复杂的网络具有模体与层次结构的认识相符 [11].

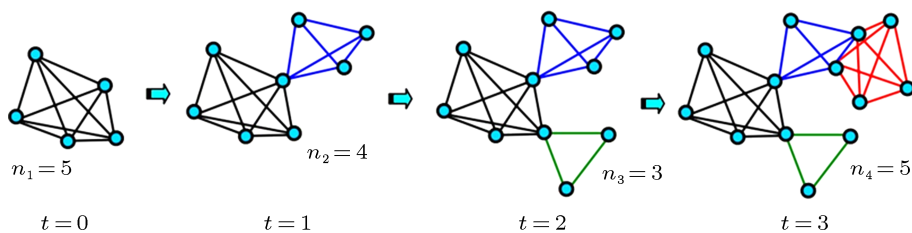


图3 网络模型演化图(这里 $n_1 = 5, n_2 = 4, n_3 = 3, n_4 = 5$. $m_1 = m_2 = 1, m_3 = 2$. n_i 是全局耦合网络 S_i 的节点数, m_i 是每次选择的节点数, t 是时间步)

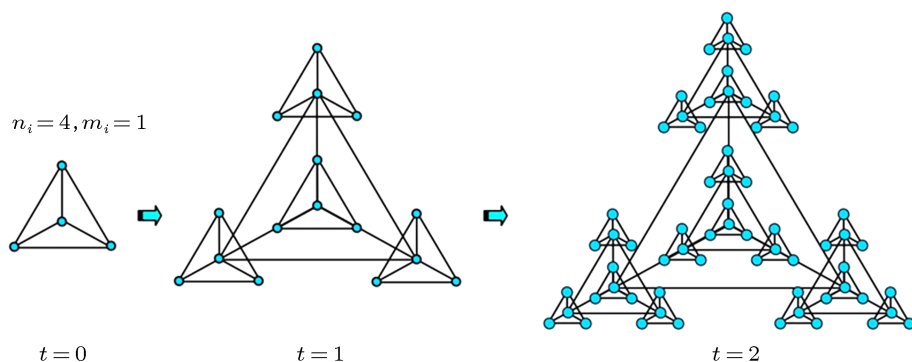


图4 网络模型演化图(特殊情况)(这里 $n_1 = n_2 = \dots = n_i = 4, m_1 = m_2 = \dots = m_i = 1, p = 1$. n_i 是全局耦合网络 S_i 的节点数, m_i 是每次选择的节点数, t 是时间步)

3.2 数字模拟

现在我们根据网络模型对我国四大城市的地铁网络进行数字模拟. 图2中圆形数据曲线表示与四城市地铁网络规模相近网络的累积度分布模拟

图. 由此可见, 模拟图与实证研究的结果相一致, 均呈现分段的度分布规律. 表2为其他统计参量的模拟计算值. 与表1中相近规模的地铁网络相比, 模拟数据与实证研究数据也基本相符. 该网络模型可以较好地解释四大城市地铁网络的特性.

表2 网络模型特征参量模拟数据表

网络参数及特征量 (模拟数据)	北京	上海	广州	深圳
站点数 N	160	246	121	95
线路数 M	10	12	8	5
网络直径 D	4	4	3	3
平均度 $\langle k \rangle$	21.61	25.48	22.59	25.58
平均路径长度 L	2.379	2.522	2.082	1.796
聚类系数 C	0.9079	0.9023	0.9102	0.9335

在数字模拟过程中, 输入较小网络参量(站点数 N , 线路数 M). 可以模拟地铁网络的网络特性. 若输入较大网络参量, 还可以模拟更大规模的城市公共汽车网络.

根据该模型我们模拟与杭州公共汽车交通网络规模相当的网络(线路为150, 站点数为843

时), 模拟结果为: $C = 0.7295, L = 2.642, D = 5, \langle k \rangle = 43.44$, 而网络的累积度分布如图5所示, 可见, 该网络的累积度分布在对数坐标中近似为一条直线, 服从指数分布. 这些结果与Chen等对杭州公共汽车交通网络的实证研究结果相符($C = 0.78, L = 2.60, D = 5, \langle k \rangle = 44.46$)^[19].

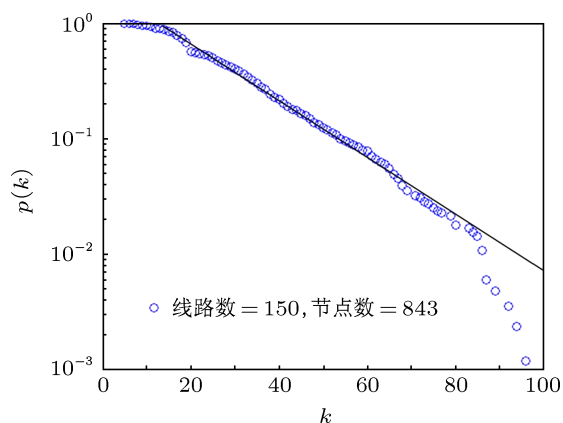


图5 网络模型累积度分布模拟图(其中,线路数 $L=150$,节点数 $N=843$,圆形表示模拟数据,直线表示拟合直线)

4 结论

本文从复杂网络的角度出发,对我国北京、上

海、广州和深圳等四大城市的地铁网络进行实证研究.对地铁网络的度分布、聚类系数和平均路径长度等特征参量进行统计计算.结果表明:四大城市的地铁网络都具有高的聚类系数和短的平均路径长度,即都具有小世界效应.而度分布服从分段分布的规律,显示出该网络具有层次结构的特征.在此基础上,本文提出了一个基于社团结构的小世界网络模型,并对该模型进行数字模拟,模拟结果与实证研究相符.对该网络模型的特殊情况进行分析,发现该网络模型演化为具有自相似行为的、并具有层次结构的模体网络,这一结果与近年人们对复杂网络具有层次结构与社团结构的结论相符.该网络模型不仅能解释地铁网络的网络特性,还可以解释其它城市公共交通网络的网络特性,对其他类型复杂网络的研究也会有一定的参考价值.

- [1] Albert R, Barabasi A L 2002 *Rev. Mod. Phys.* **74** 47
- [2] Dorogvtsev S N, Mendes J F F 2002 *Adv. Phys.* **51** 1079
- [3] Newman M E J 2003 *SIAM Rev.* **45** 167
- [4] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D U 2006 *Physics Reports* **424** 175
- [5] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [6] Barabási A L, Albert R 1999 *Science* **286** 509
- [7] Milo R, Shen-Orr S, Itzkovitz S, Kashan N, Chklovskii D, Alon U 2002 *Science* **298** 824
- [8] Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U 2004 *Science* **303** 1538
- [9] Song C, Havlim S, Makse H A 2005 *Nature* **433** 392
- [10] Palla G, Derenyi I, Farkas I, Vicsek T 2005 *Nature* **435** 814
- [11] Clauset A, Moore C, Newman M E J 2008 *Nature* **453** 98
- [12] Li W, Cai X 2004 *Phys. Rev. E* **69** 046106
- [13] Gastner M T, Newman M E J 2006 *Eur. Phys. J. B* **49** 247
- [14] Liu Hong-Kun, Zhou Tao 2007 *Acta Phys. Sin.* **56** 0106 (in Chinese)
- [15] Qian J H, Han D D, Ma Y G 2011 *Acta Phys. Sin.* **60** 098901 (in Chinese) [钱江海, 韩定定, 马余刚 2011 物理学报 **60** 098901]
- [16] Sen P, Dasgupta S, Chatterjee A 2003 *Phys. Rev. E* **67** 036106
- [17] Kurant M, Thiran P 2006 *Phys. Rev. Lett.* **96** 138701
- [18] Sienkiewicz J, Holyst J A 2005 *Phys. Rev. E* **72** 046127
- [19] Chen Y Z, Li N, He D R 2007 *Physica A* **376** 747
- [20] Yang X H, Chen G, Sun B, Chen S Y, Wang W L 2011 *Physica A* **390** 4660
- [21] Ding Y M, Ding Z 2012 *Int. J. Mod. Phys. B* **26** 1250090
- [22] Latora V, Marchiori M 2002 *Physica A* **314** 109
- [23] Seaton K A, Hackett L M 2004 *Physica A* **339** 635
- [24] Domenech A 2009 *Physica A* **388** 4658
- [25] Zhang J H, Xu X M, Hong L, Wang S L, Fei Q 2011 *Physica A* **290** 4562
- [26] <http://urbanrail.net/index.html>
- [27] [http:// ditie.bus.58.com](http://ditie.bus.58.com)

The network model of urban subway networks with community structure*

Ding Yi-Min^{1)†} Ding Zhuo²⁾ Yang Chang-Ping¹⁾

1) (*Faculty of Physics and Electronic, Hubei University, Wuhan 430062, China*)

2) (*School of Business Administration, South China University of Technology, Guangzhou 510640, China*)

(Received 30 July 2012; revised manuscript received 6 December 2012)

Abstract

In this paper, we present the empirical investigation results for the urban subway networks in China. The results show that all the urban subway networks have high clustering coefficient and small character path length, which exhibit a small-world behavior, the degree distributions take multiplicative exponential function forms. Otherwise, these networks are hierarchically organized by overlapping cliques, which are all the globally coupled networks. To explain these results, we introduce a network model, which is in good agreement with the empirical results; in addition, this model can explain the evolutionary procedure of other networks, such as the urban bus transport networks or the film actor networks.

Keywords: complex networks, urban subway network, small-world, community

PACS: 89.75.-Hc, 89.75.-k, 02.50.-r, 05.10.-a

DOI: 10.7498/aps.62.098901

* Project supported by the National Natural Science Foundation of China (Grant No. 11074067).

† Corresponding author. E-mail: dymhubu@sina.com