

基于二部图的公共交通网络模型

卢鹏丽, 贾春旭, 沈万里

(兰州理工大学计算机与通信学院, 兰州 730050)

摘 要: 利用现有方法对兰州市公共交通网络进行建模时, 不能得到真实网络参数, 或复杂度过高。为此, 以二部图作为基本模型, 将车次作为“上集”, 站点作为“下集”, 建立兰州市公共交通网络模型。计算并分析数据, 验证其符合小世界特性和无标度网络, 并利用 Laplacian 特征值和最大连通子图相对值来分析网络的性能和连通情况。分析结果证明, 该模型在减少网络存储空间的同时能保证计算结果的准确, 且对其优化也较简单直观。

关键词: 复杂网络; 二部图; 公共交通网络; 无标度网络; 小世界; 网络故障

Public Transport Network Model Based on Bipartite Graph

LU Peng-li, JIA Chun-xu, SHEN Wan-li

(School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China)

【Abstract】 Some of the modeling methods are not very suitable to establish the model of public transport network of Lanzhou for the reason that the real parameters of the network cannot be got or the method is with high self-complexity. A model based on bipartite graph is proposed, in which, the bus number as the “top set”, the station as the “bottom set”, public transport network model of Lanzhou is built. A lot of data are computed and analyzed, which indicates that this network fits with small-world and scale-free network, and analyzes the performance and connectivity with the maximum value of connected subgraph and Laplacian spectrum. Result proves that the use of bipartite graph to express, analyze and calculate the complex network not only guarantees the calculation result when the storage space drops relatively, but also seems direct-viewing and simple when optimizing the complex network on the bipartite graph.

【Key words】 complex network; bipartite graph; public transport network; scale-free network; small-world; network fault

DOI: 10.3969/j.issn.1000-3428.2012.03.087

1 概述

要理解网络结构与网络行为之间的关系, 就需要对实际网络的结构特征有很好的了解, 并在此基础上建立合适的网络结构模型。在文献[1]关于小世界网络, 以及文献[2]关于无标度网络的开创性工作之后, 人们从不同的角度出发, 提出了各种各样的网络拓扑结构模型。

文献[3]认为所有的复杂网络都可以利用二部图进行表示, 本文在文献[3]的研究基础上对兰州市公共交通网络进行建模, 并试图通过二部图对网络进行简单的改进。

2 相关研究

在对兰州市公共交通网络建模时, 对一些模型进行分析:

复杂网络的基础建模方法是 Bollobás 提出的 ER 随机图, 但是它的度分布是 Poisson 分布, 这说明了, 大部分的节点度都相似, 而这在实际的网络中基本上是不可能的。

可以通过给定度分布来利用随机图模型(配置模型或 MR 模型^[4])。但是, 在对度分布比较合理的假设下, 随着网络节点 n 的不断增加, 其聚类系数就有可能趋向于 0 了。这种方法适用于给定节点数目、给定度分布和聚类系数。

但是, 毕竟大部分实际网络既不是完全随机的, 也不是完全规则的。Dorogovstev 和 Mendes(DM)方法^[5]可以产生一个符合幂律分布和高聚类系数的网络。但是这种方法同时也会有一些相对较差的特性(如没有度为 1 的节点)。

此外, 还有一些其他的建模方法, 如在文献[6]提出的 Space L(SL)和 Space P(SP)方法中, SL 的建模方法会得到一些相对不真实的参数, 如平均路径长度较大、聚类系数较小; 而 SP 方法得到的参数较好, 但复杂度较高。

3 相关统计参数

3.1 平均路径长度

网络中 2 个节点 i 和 j 之间的距离 d_{ij} 定义为连接这 2 个节点的最短路径上的边数。网络的平均路径长度 L 定义为任意 2 个节点之间的距离的平均值, 即:

$$L = \frac{1}{2N(N-1)\sum_{i,j \in G, i \neq j} d_{ij}} \quad (1)$$

其中, N 为网络节点数。网络的平均路径长度也称为网络的特征路径长度(characteristic path length)^[7]。

3.2 聚类系数

一般地, 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。而这 k_i 个节点之间实际存在的边数 E_i 和总的可能的边数 $k_i(k_i-1)/2$ 之比就定义为节点 i 的聚类系数 C_i , 即:

$$C_i = 2E_i / (k_i(k_i-1)) \quad (2)$$

整个网络的聚类系数是指所有节点聚类系数的平均值, 它是考察网络的集团化程度^[7]。

3.3 度与度分布

网络中所有节点 i 的度 k_i 的平均值称为网络的(节点)平均度, 记为 $\langle k \rangle$ 。完全规则网络的度分布为 Delta 分布。而完全

基金项目: 甘肃省自然科学基金资助项目(0809RJZA017); 兰州理工大学校基金资助项目(0914ZX136)

作者简介: 卢鹏丽(1973—), 女, 副教授、博士, 主研方向: 算法设计与分析, 图论, 复杂网络; 贾春旭、沈万里, 硕士研究生

收稿日期: 2011-08-19 **E-mail:** jiachunxu.leo@gmail.com

随机网络的度分布近似为 Poisson 分布。这意味着当 k 远大于 $\langle k \rangle$ 时, 度为 k 的节点实际上是不存在的。

近几年的大量研究表明, 许多实际网络的度分布明显不同于 Poisson 分布。特别地, 许多网络的度分布可以用幂律形式来更好地描述, 称为无标度网络^[7]。

4 模型建立

4.1 二部图建模理论

根据图论, 一个二部图可表示为: $G=(\downarrow, \uparrow, E)$, 其中, $\downarrow$ 和 $\uparrow$ 是 2 个不邻接的点集; $E=\downarrow \times \uparrow$ 是边的集合。

给出一个二部图 $G=(\downarrow, \uparrow, E)$, 就可以很容易得到其原图的视图, 被称做“下集投影” ($\downarrow$ -projection), 记 $G'=(\downarrow, E')$ 为二部图 G 的下投影集, 若 (u, v) 属于 E' , 当且仅当在 G 中 u, v 都与同一个节点相连(图 1)。

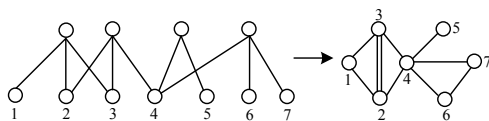


图 1 二部图及其“下集投影”

在一个二部图的“下集投影”中, 每一个上集节点与其对应的下集节点, 都会产生一个完全子图^[3]。

4.2 公共交通网络模型的建立

结合二部图理论分析兰州市公共交通网络, 认为在一条线路上的所有站点都应归到一类, 所以将站点抽象为“下集节点”, 不同线路的相同站点不重复记录, 而将线路抽象为“上集节点”, 一个线路的站点都与同一个“上集节点”相连。根据兰州市交通图所提供的信息, 最后可得到含 491 个节点和 1 031 条边的二部图模型。

如果只考虑站点之间的关系, 则可以生成包含 397 个节点和 4 466 条边的“下集投影”。与二部图比较, 虽然节点数目下降, 但边的数目却大幅上升, 这说明了二部图可以用较少的信息来表达复杂网络。

由于被分析和研究的目标是公共交通网络, 因此若在“下集投影”中出现多重边, 那么就将其看做一条边。以兰州市公共交通网络为例: 1 路车与 16 路车的线路都经过“火车站”与“盘旋路”, 那么其建模过程见图 2。

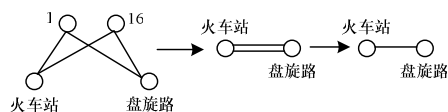


图 2 有多重边的建模过程

5 网络分析

5.1 公交网络的小世界特性

小世界网络就是指既具有较短的平均路径长度, 又具有较高的聚类系数的网络。利用二部图将兰州市公共交通网络建模后, 计算其平均路径长度 $L \approx 2.400$, 聚类系数 $C \approx 0.797$ 。表 1 为与北京和杭州的对比, 其中北京与杭州已被证实其网络具有小世界特性^[8]。

表 1 各城市的平均路径长度和聚类系数比较

城市	网络大小 N	网络连边 E	平均路径长度 L	聚类系数 C
北京	6 235	431 864	2.928	0.736
杭州	1 404	59 163	2.651	0.720
兰州	397	4 466	2.400	0.797

从表 1 可以看出, 三地的平均路径长度在 2.40~2.93 之间, 也就是说平均经过 2 次换乘就可以到达目的地; 三地的聚类

系数都在 0.7 以上, 这说明其网络的连通性较好, 集团化程度较高, 所以其网络符合小世界特性。

5.2 网络的无标度特性

网络中节点度的分布情况可用分布函数 $P(k)$ 来描述。 $P(k)$ 表示的是一个随机选定的节点的度恰好为 k 的概率。图 3 为兰州市公交网络“下集投影”的度分布图。

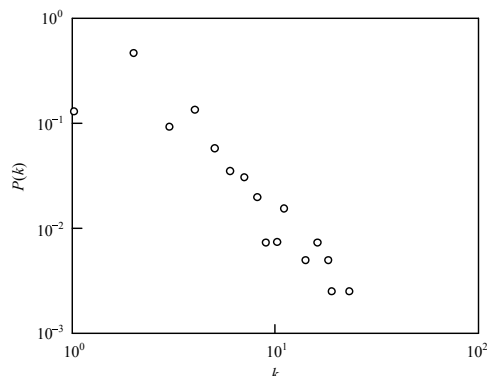


图 3 节点度分布

用一个幂律函数 $P(k) \propto k^{-r}$ 表示度分布, 通过曲线拟合能进一步求得幂指数 r 为 2.34 ± 0.204 , 则可得出兰州市的公交网络节点度分布符合幂律分布。

5.3 网络故障分析

在很多实际网络中, 一个或少数几个节点或边发生的故障会通过节点之间的耦合关系引起其他节点发生故障, 这样就会产生连锁效应, 最终导致相当一部分节点甚至整个网络的崩溃。

利用最大连通子图相对值 MCR ^[9] 来表示网络的性能, 这里 MCR 定义为:

$$M = N' / N \quad (3)$$

其中, N' 表示相继故障结束后网络的最大连通子图包含的节点个数。假设公共交通网络中某一节点出现比较严重的堵车, 则认为这个节点发生故障, 从而导致网络流量负荷的重分布。在网络受到攻击后, 可以利用 Laplacian 谱来分析网络的连通情况。

设网络为图 $G=(V, E)$, V 表示网络的节点集合; E 表示连接的集合, 其度矩阵和邻接矩阵分别记为 $D(G)$ 和 $A(G)$ 。则图 G 的 Laplacian 矩阵定义为 $L(G) = D(G) - A(G)$ 。设 $0 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_n$ 是图 G 的 Laplacian 特征值, 其中, 0 的个数即为图 G 的连通分量个数。表 2 显示了蓄意攻击和随机攻击对网络的影响对比(限于篇幅, 选取部分数据罗列)。

表 2 蓄意攻击和随机攻击对网络的影响对比

被攻击 节点个数	蓄意攻击			随机攻击		
	平均路径	M	为 0 的 Laplacian 特征根	平均路径	M	为 0 的 Laplacian 特征根
1	3.734	1.000	1	2.400	1	1
5	4.403	0.972	6	2.406	1	1
10	4.513	0.897	15	2.407	1	1

通过表 2 的数据对比可以看到, 如果最初移去的节点是随机选择的, 则网络的性能几乎没有变化($M \approx 1$); 然而, 如果移去的是度最大或者负荷最大的节点, MCR 会大大下降; 随着出现故障节点的增多, 为 0 的 Laplacian 谱个数也在不断增多, 最终网络将溃散成碎片。由此可以得出无标度网络对随机的故障有非常高的鲁棒性, 而对蓄意攻击具有高度的脆弱性。

(下转第 269 页)