

基于复杂网络理论的人际关系网络模型的构造和特性分析

Construction and Characteristics of Relationships Network Model Based on Complex Network Theory

周 卫¹, 刘会会¹, 王 念²

ZHOU Wei¹, LIU Hui-hui¹, WANG Nian²

(1. 桂林电子科技大学计算机科学与工程学院, 广西桂林 541004; 2. 桂林电子科技大学电子工程与自动化学院, 广西桂林 541004)

(1. School of Computer Science and Engineering of the Guilin University of Electronic Technology University, Guilin, Guangxi, 541004, China; 2. School of Electronic Engineering and Automation of the Guilin University of Electronic Technology University, Guilin, Guangxi, 541004, China)

摘要: 在小世界、无标度网络模型的基础上构造出人际关系网络模型, 并对人际关系网络模型进行了特性分析和仿真实验。结果表明, 该网络模型具有平均路径长度与网络规模呈对数形式增长, 度分布服从幂律增长, 具有等级结构, 聚类系数较大的特征, 能很好地符合实际的人际关系网络。

关键词: 网络 网络模型 人际关系模型

中图分类号: TP393, O233 **文献标识码:** A **文章编号:** 1002-7378(2012)01-0019-04

Abstract: This paper analyzes the characteristics of the relationships network, proposes the relationships network model based upon small world and scale free networks model, and also conducts a theoretical analysis and simulation. The results show that the characteristics of this network model is the average length path, increasing logarithmically with the number of nodes, power law degree distribution, hierarchical organization, and big clustering coefficient, which can efficiently describe the real network.

Key words: network, network model, relationships network model

人际关系网或者社会交往圈, 其实就是一群人或团体按照某种关系连接在一起而构成的一个系统。这个系统包含的关系可能是多种多样的, 比如朋友关系、同事关系、家庭关系、在线交友关系等等。这些关系纷繁复杂, 构成一个典型的复杂网络。近年来已有大量关于人际关系网络的研究, 如著名的六度分离实验^[1]、科学家合作网^[2]、演员合作网、论文引用网^[3]、在线网络社区^[4]等。研究表明, 人际关系网络具有 4 个性质: 小世界特性, 即两个认识的人

通过很少的几个人就可以联系起来; 无标度特性, 即有些人交际能力很强, 拥有很多朋友, 而有些人却拥有较少的朋友; 高聚类性, 即一个人的两个朋友很可能彼此也是朋友, 体现了“物以类聚, 人与群分”的特点; 层次性, 人际关系网络的层次性体现在很多方面, 如辈分、上下级、关系的亲疏等。

Watts 等^[5]在 1998 年提出了小世界网络模型。该模型很好地描述了复杂网络的小世界特性, 其度分布可近似用 Poisson 分布来表示, 但是实际网络的连接度分布函数都具有幂律形式。Barabási 等^[6]在 1999 年提出 BA 模型。该模型考虑了实际网络的增长特性和优先连接特性, 具有小世界特性, 度分布符合幂律分布, 但是当网络规模较大时, BA 无标度网络不具有明显的聚类特征。

收稿日期: 2011-09-20

修回日期: 2011-09-25

作者简介: 周 卫(1987-), 男, 硕士研究生, 主要从事网络协议设计、形式化技术和嵌入式开发研究。

为克服以上模型的缺陷,本文提出了一个基于BA无标网络模型的人际关系网络模型,通过理论分析和仿真实验得出,该模型具有小世界特性;度分布符合幂律分度;较大的聚类系数,而且还具有等级结构,重要的统计特性符合真实的人际关系网络,可以很好地描述人际关系网络。

1 人际关系网络模型的构造

人际关系网络模型可形式化定义为一个五元组,即 $N = (m_0, e_0, m, t)$ 。其中, $m_0 (m_0 \geq 3)$ 表示初始网络的节点数, e_0 表示初始网络的边数, m 表示一次增加的节点数, t 表示网络迭代演化的时间步。

人际关系网络模型构造算法如下:初始网络中有 $m_0 (m_0 \geq 3)$ 个节点相互连接,初始边数为 e_0 。在每经过一个时间步 t ,增加 m 个节点,并与网络中随机选择的一条边的两个端点相连。经过 t 个时间步的迭代,形成一个具有 $N(t) = m_0 + mt$ 个节点、 $E(t) = e_0 + 2mt$ 条边的网络。在此构造算法中 m 是可调参数,通过改变它的值,来改变网络的统计特性值。

2 人际关系网络模型的特性分析

2.1 网络平均路径长度

在人际关系网络模型中,网络的平均路径长度为任意两个节点之间的最短距离的平均值,也就是网络内两个人之间最短关系链中的朋友的平均个数,定义为

$$L = \frac{\delta(N)}{N(N-1)/2} = \frac{1}{N(N-1)/2} \sum_{1 \leq i < j \leq N} d(i, j), \quad (1)$$

其中, N 为网络节点数, $d(i, j)$ 表示连接节点 i 和 j 的最短路径上的边数。新加入的节点只与一条边的两个端点相连,因此,它的加入并不影响网络中已经存在的节点对之间的距离。根据文献[6],有

$$\sigma(N+1) = \sigma(N) + \sum_{i=1}^N d(i, N+1). \quad (2)$$

假设节点 $N+1$ 被添加到三角形 $\Delta y_1 y_2 y_3$ 的一条边上,则(2)式可写成

$$\sigma(N+1) = \sigma(N) + \sum_{i=1}^N (D(i, y) + 1) = \sigma(N) + N + \sum_{i=1}^N D(i, y), \quad (3)$$

其中, $D(i, j) = \min \{d(i, y_1), d(i, y_2), d(i, y_3)\}$ 。把 $\Delta y_1 y_2 y_3$ 看成是一个点 y , 则有 $D(i, y) = d(i, y)$ 。由于 $d(y_1, y) = d(y_2, y) = d(y_3, y) = 0$, 则(3)式可

写成

$$\sigma(N+1) = \sigma(N) + N + \sum_{i \in \Gamma} d(i, y), \quad (4)$$

其中, $\Gamma = \{1, 2, \dots, N\} - \{y_1, y_2, y_3\}$ 。 $\sum_{i \in \Gamma} d(i, y)$ 表示 y 节点到网络中所有其他 $N-2$ 个节点的总距离。则

$$\sum_{i \in \Gamma} d(i, y) \approx (N-3)L(N-2). \quad (5)$$

由于平均距离 $L(N)$ 按照 N 单调递增, 则 $(N-3)L(N-2)$ 可写成

$$(N-3)L(N-2) = \frac{2\delta(N-2)}{n-2} < \frac{2\delta N}{N}. \quad (6)$$

结合上面(4)式、(5)式和(6)式,可以得到

$$\delta(N+1) < \delta(N) + N + \frac{2\delta(N)}{N}. \quad (7)$$

再根据(6)式,可以得到关于 $\delta(N)$ 的方程

$$\frac{d\delta(N)}{N} = N + \frac{\delta(N)}{N}. \quad (8)$$

解这个方程,可得

$$\delta(N) = N^2 \ln N + a, \quad (9)$$

其中, a 是常数。由于 $\delta(N) \sim N^2 L(N)$, 那么 $L(N) \sim \ln N$ 。这说明网络的平均路径长度 L 的增加速度与网络规模 N 的对数成正比。

2.2 网络节点的度分布

度是单独节点的属性中的一个重要概念,在本模型中指一个人朋友的数量。求无标度网络的度分布的理论研究主要有三种方法:连续场理论、主方程法和速率方程法。我们采用主方程法进行分析。

设 $p(k, t_i, t)$ 表示 t_i 时间步加入到网络中的节点在 t 时刻度为 k 的概率。根据人际关系网络模型的构造算法可知,从 t 时刻到 $t+1$ 时刻有 m 个节点和 $2m$ 条边加入到网络中。度为 k 的节点所拥有的边占网络中总边数的比例为 $\frac{k}{e_0 + 2mt}$ 。因此,当一个新节点与随机选择的一条边的两个端点进行连接时,恰好选中度为 k 的概率为度为 k 的节点所拥有的边占网络中总边数的比例。也就是,度为 k 的节点以概率 $\frac{k}{e_0 + 2mt}$ 变为,度为 $k+2m$ 的节点以 $1 - \frac{k}{e_0 + 2mt}$ 的概率保持度不变。由此得到递推关系式^[7]

$$p(k, t_i, t+1) = \frac{k-m}{e_0 + 2t} p(k-m, t_i, t) + (1 - \frac{k-m}{e_0 + 2t}) p(k, t_i, t), \quad (10)$$

网络的度分布为 $p(k) = \lim_{t \rightarrow \infty} \left(\frac{1}{t} \sum_{t_i} p(k, t_i, t) \right)$ 。由此可得到 $p(k)$ 所满足的方程

$$p(k) = \frac{(1+m)^t \cdot (k/2)^{-1-\ln(2+m)/\ln(1+m)} + 1}{(1+2m)^t + 1} \quad (11)$$

因此,当 $t \rightarrow \infty$ 时,

$$p(k) \approx \left(\frac{k}{2}\right)^{-1-\frac{\ln(2+2m)}{\ln(1+m)}}, \quad (12)$$

即幂指数 $\gamma = 1 + \frac{\ln(1+2m)}{\ln(1+m)}$, 幂指数与参数 m 相关。当 $m=1$, $\gamma \approx 3$; 当 m 无限大于 1 时, $\gamma \approx 2$ 。我们可以根据实际的网络和仿真结果来调节 m , 以构建合适的网络。

2.3 网络的等级结构

研究表明,一些网络(如细胞网络^[8])中的拓扑模块是按等级组织起来的。等级结构或等级层次都是研究复杂网络拓扑结构的全新的视角。复杂网络具有等级结构的一个重要的量化标志是节点的聚类系数服从幂律分布 $C(k) \sim k^{-1}$ 。

根据人际关系网络模型的构造算法可知,当任一节点 x 开始进入系统并与边的两端点建立连接时,节点 x 的度 $k_x=2$, 节点 x 的 k_x 个节点之间实际存在的边数 $E(x) = 1$, 则聚类系数 $C(x) = \frac{2E(x)}{k(x)(k(x)-1)} = 1$ 。在以后的演化中,当新节点与节点 i 连接时,必定同时与 i 的一个邻居进行连接。在迭代过程中发现:节点度 k 与 $C(k)$ 存在一一对应的关系。对度为 k 的节点,其聚类系数为

$$C(k) = \frac{1+(k-2)}{k(k-1)/2} = \frac{2}{k}, \quad (13)$$

即 $C(k) \sim k^{-1}$, 因此从理论上证明了本文建立的人际关系网络具有等级结构。节点度很小的节点具有高的集聚系数,而度很高的 Hub 节点具有低的集聚系数。

2.4 网络的聚类系数

由前面的分析可得 $E(x)$ 和 $k(x)$ 的关系如下

$$E(x) = k(x) - 1. \quad (14)$$

根据聚类系数的定义可得

$$C(x) = \frac{2(k(x)-1)}{k(x)(k(x)-1)}. \quad (15)$$

因此 $C = \frac{2}{N} \sum_{i=1}^N \frac{1}{k(i)}$, 其中 $k(i)$ 是节点 i 的度。

将 $\sum_{i=1}^N f(k_i)$ 写成积分的形式^[6]

$$\sum_{i=1}^N f(k_i) = \int_{k_{\min}}^{k_{\max}} N p(k) f(k) dk, \quad (16)$$

其中, k_{\max} 和 k_{\min} 是网络中度的最大值和最小值。则(16)式可写成

$$C = 2 \int_{k_{\min}}^{k_{\max}} \frac{p(k)}{k} dk. \quad (17)$$

由前面分析可知

$$p(k) = \frac{(1+m)^t \cdot (k/2)^{-1-\ln(2+m)/\ln(1+m)} + 1}{(1+2m)^t + 1} =$$

$$\alpha k^{-1-\frac{\ln(1+2m)}{\ln(1+m)}},$$

由人际关系网络模型的构造算法可知 $k_{\min} = 2$, 且 k_{\max} 远大于 k_{\min} 。 α 又满足如下等式

$$\int_{k_{\min}}^{k_{\max}} p(k) dk = 1. \quad (18)$$

结合(11)式,(17)式和(18)式,可求得

$$C = \frac{m+3}{2m+3} \left[\ln\left(\frac{m+1}{m}\right) - \frac{m+1}{m} \right] \cdot \frac{(1+2m)^t}{[(1+2m)^t + 1]}. \quad (19)$$

当 $t \rightarrow \infty$ 时, $C \approx \frac{m+3}{2m+3}$ 。因此,聚类系数与 m 有关,且随着迭代步的增加,将逐渐趋近于一个非零的常数。当 $m=1$ 时, $C \approx 0.8$; 当 m 较大时, $C \approx 0.5$ 。可根据实际的网络特性来调节参数 m 。从以上分析可知,我们建立的人际关系网络模型具有小世界特性。

3 仿真实验

为了检测理论分析的正确性以及验证参数 m 对人际关系网络的统计特性的影响,根据人际关系网络模型的构造算法,用 Matlab 仿真了网络的演化过程(图 1~3)。

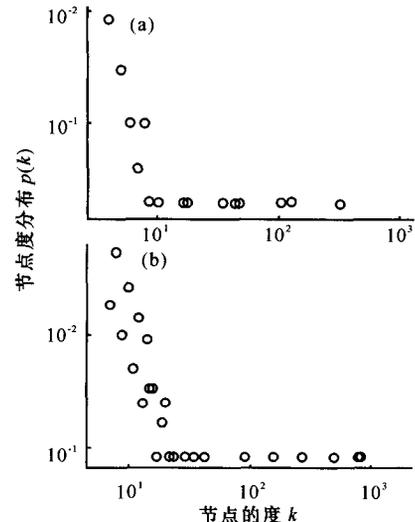


图 1 网络节点度的概率分布
(a) $m=1$; (b) $m=2$ 。

从图1可以看出,度分布大致服从幂律分布,仿真结果与理论分析的结果基本一致。

由图2可知,网络聚类系数与参数 m 有关,对比两图1(a)和图1(b)可知: m 值越大,网络的聚类系数越小,但是随着网络规模的增大,网络的聚类系数增大,并趋于一个常数。

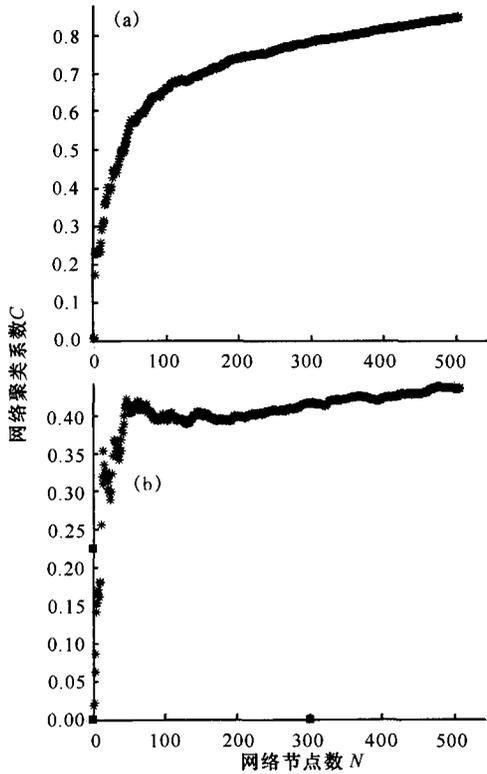


图2 网络聚类系数分布
(a) $m=1$; (b) $m=5$ 。

从图3可以看出,参数 m 对变化趋势只起到微观的作用,随着网络规模的不断增加,网络的平均路径呈对数形式增长。尽管网络的节点数巨大,但网络的平均路径长度却小的惊人,说明模型很好的表现出小世界特性。

本文建立的人际关系网络模型具有较大的聚类系数、非常小的平均路径长度,不仅是一个无标度网络,也是一个小世界网络。该模型比单纯的 WS 小世界模型和 BA 无标度模型更接近于真实的人际关系网,对虚拟社会系统的建模仿真,具有一定的理论指导意义。

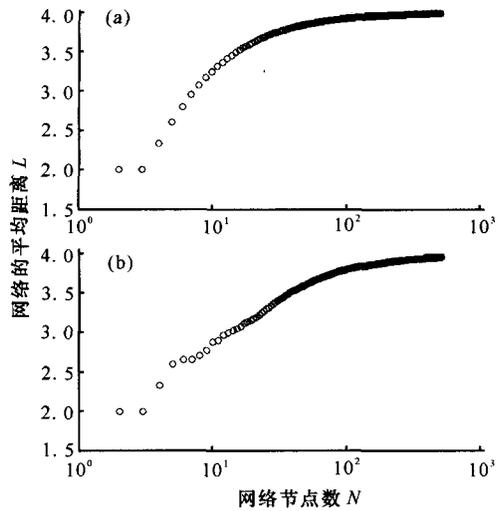


图3 网络平均距离变化趋势
(a) $m=1$; (b) $m=5$ 。

参考文献:

- [1] Ebel H, Mielsch L, Bornholdt S. Scale-free topology of e-mail networks[J]. Phys Rev, 2002, 66: 60-64.
- [2] Barabasi A L. Linked: the new science of networks [M]. American Journal of Physics, 2003, 71(4): 409.
- [3] Newman M. The structure and function of complex networks[J]. SIAM Review, 2003, 45: 167-256.
- [4] 胡海波, 徐玲, 王科, 等. 大型在线社会网络结构分析[J]. 上海交通大学学报, 2009(4): 587-591.
- [5] Watts D J, Strogatz S H. Collective dynamites of 'small - world' networks [J]. Nature, 1998, 6684 (393): 440-442.
- [6] Zhou Tao, Yan Gang, Wang Bing hong. Maximal planar networks with large clustering coefficient and power-law degree distribution[J]. Phys Rev E, 2005, 71: 275-287.
- [7] 裴伟东, 夏玮, 王全来, 等. 一类三角形结构动态复杂网络演化模型分析[J]. 中国科学技术大学学报, 2010 (11): 1186-1190.
- [8] Stelling J, Sauer U, Szallasi Z, et al. Robustness of cellular functions[J]. Cell, 2004, 118: 675-685.

(责任编辑:尹 闯)