

文章编号:1001-9081(2009)05-1230-03

## 基于中间节点效应的无标度网络演化模型研究

饶 浩<sup>1</sup>, 杨 春<sup>2</sup>, 陶少华<sup>3</sup>

(1. 韶关学院 计算机科学系, 广东 韶关 512005; 2. 华中师范大学 计算机科学系, 武汉 430079;

3. 许昌学院 计算机科学与技术系, 河南 许昌 461000)

(sail2009@yeah.net)

**摘 要:**原 BA 模型以网络中已存在的各个节点与新增节点的连接相互独立为前提。然而,在真实系统中,当网络中一个节点与新增节点连接后,该节点对其邻居节点与新增节点的连接会存在影响。针对该现象,提出了基于中间节点效应的无标度网络演化模型。首先描述与定义了中间节点效应,然后给出了中间节点效应模型的生成算法,并从理论上分析了该模型的度分布情况,最后利用仿真验证了理论分析的正确性,并就度分布、群聚系数、平均路径长度等复杂网络参数与原 BA 模型进行了对比,结果表明此模型能生成无标度网络并且更符合现实网络的演化过程。

**关键词:**BA 模型;中间节点效应;连接概率;度分布

**中图分类号:**TP301 **文献标志码:**A

## Research on new evolution model of scale-free network

RAO Hao<sup>1</sup>, YANG Chun<sup>2</sup>, TAO Shao-hua<sup>3</sup>

(1. Department of Computer Science, Shaoguan University, Shaoguan Guangdong 512005, China;

2. Department of Computer Science, Huazhong Normal University, Wuhan Hubei 430079, China;

3. Department of Computer Science and Technology, Xuchang University, Xuchang Henan 460000, China)

**Abstract:** A new evolution model of scale-free network for the recognition of interim node effect was proposed concerning the limitations of the usual B-A model and the phenomenon generally existing in the real scale-free network. Firstly, the description and definition of the interim node effect were given. Then the generation algorithm for the interim node effect model was presented, and the theoretical analysis of the nodes' degree distribution of the model was given and proved in simulation. Simulation results show that the new model still inherits the scale-free feature but matches the evolving process of scale-free networks much better than the usual B-A model in terms of the power-law exponent, the clustering coefficient and the average path length.

**Key words:** B-A model; interim node effect; connection probability; degree distribution

## 0 引言

自从发现真实网络的无标度性质<sup>[1-2]</sup>以来,有关真实网络中各种宏观性质的微观生成机制、网络的演化规律等一系列问题的研究成为目前研究人员广泛关注的热点。BA 无标度网络模型<sup>[1]</sup>首次从网络演化的角度来研究网络的一些宏观性质的起源,开辟了网络研究的新纪元。BA 模型将真实系统通过自组织生成无标度网络的现象归功于两个主要因素<sup>[1]</sup>:增长性和优先连接,并据此成功解释了复杂网络中集散节点的存在,即“富者愈富”现象。但是,同在真实网络中得到的实验结果相比较,我们不难发现 BA 模型存在许多局限性。首先,BA 模型只能生成度分布的幂律指数固定为 3 的网络,而所观测到的实际网络的幂律指数通常在 2~3 的范围内取值<sup>[3]</sup>。此外,对许多的真实网络而言,它们的度分布并不是幂指数型的,而是表现为其他的形式<sup>[4]</sup>,如指数中止和小变量饱和等<sup>[5-6]</sup>。这些实验结果说明真实网络在演化过程还受到许多其他因素的影响,而在 BA 模型中这些因素没有被考虑。

BA 模型预测了网络中节点的数量以及网络中的各个节点拥有的连接数随着时间的增长而动态增加,并以网络中老

节点已拥有的连接数作为新增节点选择连接对象的依据。但是,另一方面,许多实例又表明在真实的无标度系统中一个节点的连接与增长率并不仅仅只依赖于对方本身已拥有的连接数量。例如,在人际关系网中,我们不仅仅因为某人的“名气”大而认识或结交他,也可能是因某个中间人的介绍而认识对方或是因为双方都认识某人而相互结交;在文献引用网络中,有些文献是因为比较权威而被引用,但是当引用了一篇文章后,就更易查阅和研究这篇文章所引用的文献,从而引用那些文献等。所有这些例子都说明了一个简单的事实:真实的无标度系统的优先连接特性除了依据老节点的连接数量外,还受这些“中间人”、“中间文献”等的影响。我们把这一现象统称为中间节点效应。本文提出了中间节点效应对网络增长过程中节点优先连接的影响,并系统地研究了在中间节点效应作用下无标度网络的演化模型。

## 1 中间节点效应

由原 BA 模型的优先连接思想可以看出,网络中已存在的各个节点与新增节点的连接是相互独立的。然而,在真实系统中,当网络中一个节点与新增节点连接后,该节点对其邻居节点与新增节点的连接可能存在影响。如图 1,当  $v_i$  连接

收稿日期:2008-12-02;修回日期:2009-02-14。 基金项目:河南省基础与前沿技术研究计划项目(082102210085)。

作者简介:饶浩(1976-),男,广东兴宁人,讲师,硕士,主要研究方向:计算机网络、智能信息处理; 杨春(1984-),男,湖北宜都人,硕士研究生,主要研究方向:复杂网络、P2P 网络; 陶少华(1978-),女,河南商丘人,讲师,硕士,主要研究方向:复杂网络。

了  $v_r$  后(图(b)),再当  $v_s$  从  $v_p, v_q$  中选择连接节点时(图(c)),  $v_r$  可能会充当中间节点影响  $v_s$  的选择,从而导致  $v_s$  连接  $v_p$  (图(d))。

**定义1** 对于无向图  $G = (V, E)$ ,  $V$  是网络中节点的集合,  $E$  是网络中边的集合,  $v_p, v_q, v_r \in V, (v_r, v_p) \in E$ , 且  $k_p = k_q$  ( $k_i$  为节点  $i$  的度),  $v_s$  为新增节点, 若  $(v_s, v_r) \in E$ , 则连接概率  $P((v_s, v_p) \in E) \geq P((v_s, v_q) \in E)$ , 这种影响即为中间节点效应。此定义对于有向图亦然。

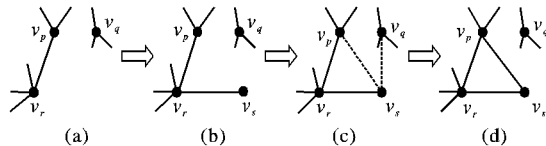


图1 新节点连接示意图

中间节点效应的思想类似于小世界效应,很多实际网络都具有这种现象。以搜索引擎的工作原理为例,系统维护一个超链队列,其中包含一些起始 URL,搜集器从这些 URL 出发,下载相应的页面,并从中抽取出新的超链加入到队列中,此过程不断重复直至队列为空。从中可看出,队列中几乎所有的 URL 都充当了中间节点。也正因如此,搜索引擎系统才能迅速扩大搜索范围。

## 2 基于中间节点效应的无标度网络演化模型

### 2.1 中间节点效应模型生成算法

将中间节点效应的影响值用参数  $\beta$  表示,并令  $\beta$  服从(0-1)分布,即若新增节点与网络中某节点  $v_i$  的连接受中间节点的影响,则令  $\beta_i = 1$ ; 否则,令  $\beta_i = 0$ 。

因此,给出考虑中间节点效应的无标度网络演化模型生成算法如下。

1) 增长性:假设网络最初有  $m_0$  个节点。在每个时间间隔加入一个新节点,每次加入的新节点通过  $m$  ( $m \leq m_0$ ) 条新加入的连接边与网络中已有的  $m$  个节点相连。

2) 优先连接性:新增加的节点与网络中已存在的节点  $v_i$  进行连接的概率  $\Pi_i$  被定义为:

$$\Pi_i = q_i / \left( \sum_j q_j \right) \quad (1)$$

其中,  $q_i$  被定义为  $q_i = \frac{k_i}{\sum_j k_j} (1 + \beta_i)$ ,  $k_i$  为节点  $v_i$  的度,定义

$$\beta_i = \begin{cases} 1, & \text{存在中间节点 } v_i \\ 0, & \text{不存在中间节点} \end{cases}$$

从式(1)可看出,本生成算法的连接概率已不完全依赖于老节点的连接度数,而要受  $\beta_i$  即中间节点效应的影响,当  $\beta$  取值全为 0,即不考虑中间节点效应时,此模型即化为原 BA 模型。

### 2.2 中间节点效应模型度分布研究

度分布<sup>[8]</sup>是复杂网络的一个重要统计特征,节点的度指的是与该节点连接的边数,度分布则表示为节点度的概率分布函数  $P(k)$ ,即网络中的节点有  $k$  条边的概率。在目前的研究中,两种度分布较为常见,一是指数度分布,即  $P(k)$  随着  $k$  的增大以指数形式衰减;另一种是幂律分布,即  $P(k) \sim k^{-\gamma}$ ,其中  $\gamma$  称为幂律指数,不同  $\gamma$  的网络,其动力学性质也不同。无标度网络的度分布满足第二种形式即幂律分布。下面从理论上分析中间节点效应模型的度分布形式。

假设  $k_i$  为连续实变量,由动态方程可得:

$$\frac{\partial k_i}{\partial t} = m \cdot \Pi_i = \frac{mk_i(1 + \beta_i)}{2mt + \sum_j k_j \beta_j} \quad (2)$$

式(2)的初始条件是:节点  $i$  在  $t_i$  时刻进入系统,其度数  $k_i(t_i) = m$ ,且  $t > 0$ , 于是,解微分方程可得:

$$P(k_i(t_i) < k) = \begin{cases} P\left(t < \frac{k\beta_i}{m(\beta_i - 1)} + \left[\left(\frac{k}{m}\right)^{\frac{2}{1+\beta_i}} - 1\right] \frac{\sum_{j \neq i} k_j \beta_j}{2m} + \left(t_i - \frac{\beta_i}{\beta_i - 1}\right) \left(\frac{k}{m}\right)^{\frac{2}{1+\beta_i}}\right), & \beta_i \neq 1 \\ P\left(t < \frac{k}{2m} \left(\ln k + 2t_i + \frac{\sum_{j \neq i} k_j \beta_j}{m} - \ln m\right) - \frac{\sum_{j \neq i} k_j \beta_j}{2m}\right), & \beta_i = 1 \end{cases} \quad (3)$$

由于时间  $t$  服从均匀分布,所以:

$$p(t_i) = \frac{1}{m_0 + t}$$

因此,可以计算得到节点的度分布:

$$p(k) = \frac{\partial P(k_i(t_i) < k)}{\partial k} = \begin{cases} \frac{1}{m_0 + t} \left[ \frac{2}{1 + \beta_i} m^{\frac{2}{1+\beta_i}} k^{\frac{-3-\beta_i}{1+\beta_i}} \left(t + \frac{\sum_{j \neq i} k_j \beta_j}{2m}\right) + \frac{\beta_i}{1 + \beta_i} m^{\frac{1-\beta_i}{1+\beta_i}} k^{\frac{-2}{1+\beta_i}} \right], & \beta_i \neq 1 \\ \frac{1}{m_0 + t} \left( \frac{mt}{k^2} + \frac{\sum_{j \neq i} k_j \beta_j}{2k^2} + \frac{1}{2k} \right), & \beta_i = 1 \end{cases} \quad (4)$$

由式(4)可得,当  $t \rightarrow +\infty$  时,节点的度分布趋于:

$$p(k) \approx \frac{2}{1 + \beta_i} m^{\frac{2}{1+\beta_i}} k^{\frac{-3-\beta_i}{1+\beta_i}} \quad (5)$$

可以看出,式(5)满足的  $P(k) \sim k^{-\gamma}$  形式,节点的度分布

服从幂律分布,且幂律指数  $\gamma = \frac{3 + \beta_i}{1 + \beta_i}$ , 并与  $\beta_i$  有关。当  $\beta_i = 0$  时,  $p(k) \approx 2m^2 k^{-3}$ ,  $\gamma = 3$ , 即为原 BA 模型的度分布;当  $\beta_i = 1$  时,  $p(k) \approx m k^{-2}$ ,  $\gamma = 2$ 。由于  $\beta_i$  服从(0-1)分布,其期望  $E(\beta_i) = P\{\beta_i = 1\} \in [0, 1]$ , 所以,在本模型中,  $\gamma$  的取值应在 2 到 3 之间。显然,此结果比原 BA 模型更符合实际观测得到的网络幂律分布指数的取值。

## 3 仿真与数据分析

按照本文提出的模型生成算法,在计算机上模拟了网络的演化过程,根据得出的数据验证了度分布理论分析的正确性,并就聚集系数、平均路径长度等复杂网络参数将中间效应模型与原 BA 模型进行了对比。

### 3.1 度分布验证

在用仿真模拟此模型生成过程中,取网络节点数  $N = 5010$ , 初始节点数  $m_0 = 10$ , 每次加入的新节点与网络中已有的  $m = 8$  个节点相连。对参数  $k$ ,  $P(k)$  分别取对数,得到节点度分布的仿真结果如图 2 所示。

从图 2 可看出,网络的节点度分布服从幂律分布,且通过

线性拟合,我们得到其幂律指数  $\gamma = 2.9267$ 。可见,仿真结果与理论分析比较吻合,从而验证了理论分析的正确性,也证明了中间效应模型生成的是无标度网络。

### 3.2 聚集系数对比

网络的聚集系数  $C$  描述了网络的聚集程度,是一个用来刻画网络局域性质的量。节点  $i$  的聚集系数定义为<sup>[9]</sup>:

$$C(i) = \frac{2E(i)}{k_i(k_i - 1)}$$

其中,  $E(i)$  为节点  $i$  的邻居间实际存在的边数。整个网络的聚集系数是每个节点聚集系数的平均值:

$$C = \frac{1}{N} \sum_{i=1}^N C(i)$$

其中,  $N$  为整个网络的节点总数。

根据聚集系数的定义,分别考察了  $N$  取 1 010、1 510、2 010、2 510、3 010、3 510、4 010、4 510、5 010 九种情况下中间效应模型与原 BA 模型的聚集系数值,如图 3。

如图 3 所示,在各同等网络规模下,中间效应模型的聚集系数均大于原 BA 模型。这是因为前者考虑了中间节点效应,如果新增节点  $v_i$  与某节点  $v_r$  连接后,则  $v_i$  更容易与  $v_r$  的邻居节点  $v_p$  相连接,即更容易聚类,所以中间效应模型的聚类效应比原 BA 模型更明显。

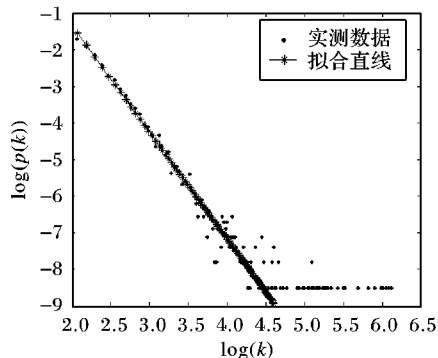


图2 中间效应模型的节点度分布

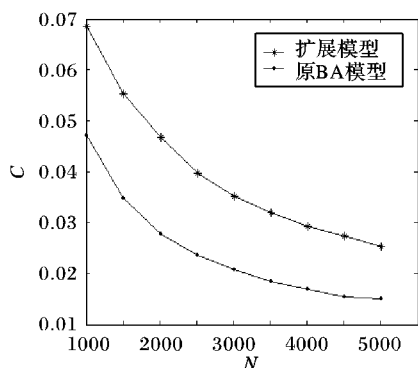


图3 聚类系数比较

### 3.3 平均路径长度对比

传输延迟是影响网络性能、信息传递的重要因素,而网络距离和直径是度量传输延迟的重要参数。但网络距离和直径都无法对网络的总体特征进行评价,因此需要引入网络的平均路径参数来度量<sup>[10]</sup>。平均路径长度定义为<sup>[9]</sup>:

$$L = \frac{1}{N(N-1)} \sum_{i,j \in V} d(i,j)$$

其中  $d(i,j)$  定义为从节点  $v_i$  到达另一个节点  $v_j$  的最短路径长度,  $V$  表示网络中节点集合。

按上述定义,采用类似于聚集系数对比的仿真方法,对中间效应模型与原 BA 模型的平均路径长度作了对比(如图

4)。结果显示,在各网络规模下,前者的平均路径长度均小于原 BA 模型。由于中间效应模型更容易聚类,所以其平均路径长度比原 BA 模型短,此结果比较合理。

## 4 结语

原 BA 模型将真实系统通过自组织生成无标度网络的现象归功于增长性和优先连接两个特性,并以网络中老节点已拥有的连接数作为新增节点优先连接的唯一依据。但是,许多实例又表明在真实的无标度系统中一个节点的连接与增长率并不仅仅只依赖于对方本身已拥有的连接数量,还要受中间节点效应的影响。本文根据真实系统中存在的中间节点现象提出了基于中间节点效应的无标度网络演化模型,从理论上分析了该模型的度分布情况,并利用仿真验证了理论分析的正确性,证明了该模型能生成无标度网络,最后还就聚集系数、平均路径长度两个属性与原 BA 模型分别进行了对比,所有的实验结果均表明中间效应模型比 BA 模型更符合真实网络的演化过程。

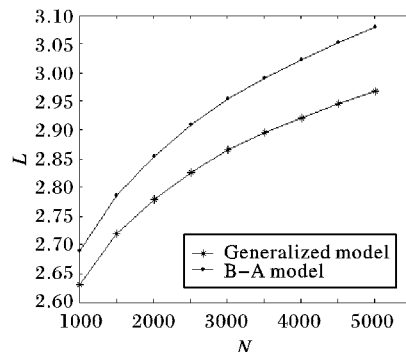


图4 平均路径长度比较

### 参考文献:

- [1] ALBERT R, BARABÁSI A L. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [2] ALBERT R, BARABÁSI A L, JEONG H. Power-law distribution of the world wide Web[J]. Science, 2000, 287(5461): 2115.
- [3] ALBERT R, BARABÁSI A L, JEONG H. Mean-field theory for scale-free random networks[J]. Physica A, 1999, 272(1/2): 173-187.
- [4] LIU YU-HUA, TAO SHAO-HUA, XU KAI-HUA, et al. On information dimension self-similarity of complex networks [C]// Second International Conference on Complex Systems and Applications—Modeling, Control and Simulations: ICCSA 2007. Jinan, China: Watam Press, 2007: 1265-1268.
- [5] KRAPIVSKY P L, REDNER S. Organization of growing random networks [J]. Physical Review E, 2001, 63(6): 66-123.
- [6] LIU YU-HUA, TAO SHAO-HUA, XU KAI-HUA, et al. A new evolving model of complex networks [C]// The 4th International Conference on Impulsive Dynamical Systems and Applications: ICID-SA2007. Nanchang, China: Watam Press, 2007: 1803-1805.
- [7] TouchGraph GoogleBrowser V1.01 [CP/OL]. [2008-08-12]. <http://www.touchgraph.com/TGGoogleBrowser.html>.
- [8] ALBERT R, BARABÁSI A L. Statistical mechanics of complex networks [J]. Reviews of Modern Physics, 2002, 74(1): 47-97.
- [9] BARRAT A, WEIGT M. On the properties of small-world network models [J]. The European Physical Journal B - Condensed Matter and Complex Systems, 2000, 13(3): 547-560.
- [10] WATTS D J, STROGATZ S H. Collective dynamics of small-world networks [J]. Nature, 1998, 393(6684): 440-442.