

一种改进的演化型 Internet 拓扑模型

姚园园, 张志鸿

(郑州大学 信息工程学院, 河南 郑州 450052)

摘 要:通过分析传统的基于偏好依附的演化型 Internet 拓扑模型 BA 模型, 针对其考虑网络演化过程中事件的不完善性, 提出一种 BA 模型的改进 VBA 模型。该模型更全面地考虑了网络演化过程中加边、加点和去边三种事件。对 VBA 模型的拓扑度量值分析结果表明, 该模型能够生成比传统模型更优的具有幂律度分布和小世界特性的网络实例。

关键词: Internet 拓扑建模; 网络演化; 偏好依附; VBA 模型

中图分类号: TP393

文献标识码: A

文章编号: 1673-629X(2006)08-0088-03

A Novel Internet Topology Model Based on Evolution

YAO Yuan-yuan, ZHANG Zhi-hong

(School of Information Engineering, Zhengzhou University, Zhengzhou 450052, China)

Abstract: Analyzes a classical growing network model based on preferential attachment, the BA model. An improved BA model is presented which takes into account of adding either nodes or edges and deleting edges in the evolving networks. The analysis of its topology metrics shows that VBA model can generate better graphs with a power-law degree distribution and small-world properties.

Key words: Internet topology modeling; network's evolution; preferential attachment; VBA model

0 引言

Internet 可以看作是一个由计算机组成的不断演化的生态系统。对这个庞大的系统建模是在更高层次上开发利用 Internet 的基础, 具有十分重要的意义。对 Internet 拓扑建模涉及网络测量、图论、算法设计、统计学、数据挖掘、可视化以及数学建模等多个研究领域^[1]。

一般的网络演化包含 5 种现象: 加边、加边、重连、去边、去点。传统的网络演化模型 BA 模型和其扩展 GBA 模型, 最多只考虑了加边、加边和重连。这种涵盖事件的不完备性使得传统模型必然不能生成较完整的表现 Internet 大规模拓扑特征的实例。文中提出一种 VBA 模型来改进 BA 模型, 它根据 Internet 的实际情况, 考虑了网络演化过程中加边、加点和去边三种事件, 较好地解决了上述问题。

1 基本概念

定义 1 顶点的度数: $k_i = \sum_{l \in E} \delta_{il}^l$

其中 E 为边集, 若顶点 i 包含在边 l 的顶点中, 则 $\delta_{il}^l = 1$; 否则, $\delta_{il}^l = 0$ 。

定义 2 平均最短路径 (l) 是矩阵 $(d_{ij})_{N \times N}$ 的分布特

征值的平均值^[2]。其中, $d_{ij} = |l_{ij}|$, l_{ij} 表示两点间的最短距离, N 表示点集。

定义 3 聚集程度:

$$c_i = \left(\sum_{l \in E, x, y \in N_i} \delta_{il}^x \delta_{il}^y \right) / C_n^2$$

其中 N_i 表示顶点的近邻集合, $n = |N_i|$ 。实际上 c_i 是一个比率, 是 i 的最近邻之间边的数量与最近邻之间可能的边的总数之比。

定义 4 增长概率:

$$\prod(k_i) = \delta k_i / \delta k$$

其中 δk 为在一个时间片 ΔT 内增加的边的总数。这样比率 $\delta k_i / \delta k$ 就被定义为一条新边连接到点 i 的概率^[3]。

$k_s(t)$ 表示在 t 时刻, 第 s 点的度数。

$P(k, t)$ 表示在时间 t , 一个节点的度数为 k 的概率。

2 演化型网络拓扑模型

Internet 拓扑模型可分为两类: 一类是静态型, 描述 Internet 的拓扑特征; 另一类是演化型, 描述拓扑特征形成的机理。演化型拓扑模型在建模过程中把网络的动态进化规则作为主要的元素, 它有两个基本方面: 1) Internet 或者其他复杂网络是一个增长的网络, 其节点和连接数都随时间不断增加; 2) 连接都是由随机过程来确定, 这个随机过程具有节点局部属性的偏好性。

偏好依附 (preferential attachment) 的概念首先是由 Barabási 和 Albert 提出的^[4], 其核心思想是认为在很多真实网络中, 新增加的边并不是随机连接的, 而是倾向于和

收稿日期: 2005-11-08

作者简介: 姚园园 (1980-), 女, 河南南阳人, 硕士研究生, 研究方向为网络拓扑建模; 张志鸿, 博士, 副教授, 研究方向为语义 Web 过程、复杂系统建模。

具有较大度数的点连接。利用偏好依附这种简单的动力学机制的演化型网络拓扑模型,可以自然的生成具有幂律度分布和小世界特性的网络实例。

Barabási 和 Albert 提出的 BA 模型^[5]是基于偏好依附的最简单的模型,其建模方法基于如下两条规则:

规则 1 Growth:网络开始于有 m_0 个互连顶点的小核心,每个时间步增加一个新顶点,由新点引出 m ($m \leq m_0$) 条边连接到系统原有的点上。

规则 2 Preferential attachment:新边连接到原有点 s 上的概率正比于该点的度数 k_s ,为:

$$\prod[k_s(t)] = \frac{k_s(t)}{\sum_j k_j(t)}$$

根据相关文献,将 BA 模型的各种拓扑度量值总结如下:

BA 模型具有固定的平均度数, $\langle k \rangle = 2m$ 。度的分布为:

$$P(k, t) = 2m^2 \frac{t + \frac{m_0}{m} \langle k \rangle_0}{t + m_0} k^{-3}$$

当 $t \rightarrow \infty$ 时,不依赖于时间的幂律度分布为:

$$P(k) = 2m^2 k^{-3}$$

BA 模型的平均最短路径为:

$$\langle l \rangle \sim \frac{\log(N)}{\log \log(N)}$$

平均聚集程度为:

$$\langle c \rangle_N = \frac{m}{8N} (\ln N)^2$$

BA 模型的平均最短路径表现了很好的小世界性质;其平均聚集程度随着网络规模的增加而减小,最后当网络规模无限大时趋向于 0。

Barabási - Albert 的 BA 算法虽然只根据简单的动力学规则就生成了具有幂律度分布和小世界性质的生成图,然而其聚集程度,还有对较早结点总是具有最大连接数的假设却不符人们所观察到的 Internet。

3 VBA 模型

一般网络演化包含 5 种情况:加边、加边、重连、去边、去点,其中边的重连是指:一条边的一个端点 j 不变,而另一个端点 i 断开,然后连接到端点 i' 上,且 $i' \neq i$,当这两个事件在很短的间隔内发生时,就叫做重连。目前绝大多数网络演化模型都在这 5 种事件的范围内讨论。而 BA 模型只考虑了加边和加边两种事件。对 Internet 的研究表明,边重连的概率非常小,而同时边消失的概率是不能忽略的,而 BA 模型却没有考虑这些因素。

由此文中提出 BA 模型的一个改进,考虑加边、加点和去边三种事件。因为从重连的定义可以看出,边的重连实际上可以分为去边和加边两个事件,所以不再对重连事件进行单独考虑。

假设初始有 m_0 个互连的点,VBA 模型定义如下:

规则 1 以概率 q 断开 m 条边,每一条边都随机选择。

规则 2 以概率 p 增加 m 条新边,每条边的一个端点随机连接,另一个端点以概率 $\prod[k_i(t)]$ 连接。

规则 3 以概率 $1 - p - q$ 增加一个有 m 条边的新点,每条边的另一端以概率 $\prod[k_i(t)]$ 连接到已存在的点。其中 $k_s(t)$ 表示在 t 时刻第 s 个点的平均度数。

$$\prod[k_i(t)] = \frac{k_i(t)}{\sum_j k_j(t)} \quad (1)$$

这里用连续性方法^[2]来处理 VBA 模型。在 VBA 模型中,第 s 点的平均度数速率方程要同时考虑边的删除和增加。

当 t 非常大时,可以忽略最初 m_0 个顶点的核的贡献,这样根据规则 1、规则 2、规则 3 中边的删除与增加分别对增长速率的贡献, $k_s(t)$ 的完整的进化速率方程就可以表示为:

$$\begin{aligned} \frac{\partial k_s(t)}{\partial t} &= \frac{\partial}{\partial t} (k_s^{(1)}(t) + k_s^{(2)}(t) + k_s^{(3)}(t)) \\ &= \frac{m(p-2q)}{(1-p-q)t} + \frac{(1-q)k_s(t)}{2(1-2q)t} \end{aligned} \quad (2)$$

因为每一个时间步都以概率 $(1-p-q)$ 增加一个新点,那么通过时间 t 系统中点的总数是 $N(t) = (1-p-q)t$,总的度数增长可以表示为:

$$\sum_j k_j(t) = 2(1-2q)mt$$

通过边界状态 $k_s(s) = m$ 解上面的方程得到:

$$k_s(t) = [A(p, q, m) + m] \left(\frac{t}{S}\right)^{B(p, q)} - A(p, q, m) \quad (3)$$

其中,

$$A(p, q, m) = 2m \frac{(p-2q)(1-2q)}{(1-p-q)(1-q)}$$

$$B(p, q) = \frac{1-q}{2(1-2q)}$$

这样可以计算这个模型的度的分布,用 $p(k, s, t)$ 表示第 s 点在 t 时刻的度值为 k 的概率,令

$$p(k, s, t) = \delta(k - k_s(t))$$

其中 $\delta(x)$ 是狄拉克 δ 函数,使用 master 方程方法得公式(4):

$$\begin{aligned} P(k, t) &= \frac{1}{t + m_0} \sum_{s=0}^t p(k, s, t) \\ &= \frac{1}{t + m_0} \int_0^t \delta(k - k_s(t)) ds \\ &\equiv - \frac{1}{t + m_0} \left(\frac{\partial k_s(t)}{\partial s} \right)^{-1} \Big|_{s=s(k, t)} \end{aligned} \quad (4)$$

其中 $s = s(k, t)$ 是方程 $k = k_s(t)$ 的解。

根据公式(4)解本模型的度分布得到式(5):

$$\begin{aligned} P(k, t) &\simeq \frac{t}{t + m_0} \frac{[A(p, q, m) + m]^{1/B(p, q)}}{B(p, q)} \\ &\quad [A(p, q, m) + k]^{-1 - \frac{1}{B(p, q)}} \end{aligned} \quad (5)$$

当 $t \rightarrow \infty$ 时,可以近似得到:

$$p(k) \sim (k + k_0)^{-\gamma} \quad (6)$$

其中, $k_0 = A(p, q, m)$, $\gamma = 1 + 1/B(p, q)$ 。

要使该式有意义, 只要满足 $A(p, q, m) + m > 0$, 也就是 $p > 2q$ 。

VBA 模型的度分布指数可以根据参数 p, q 的变化来进行调节, 幂律分布与指数分布都可以在该模型的网络中出现, 但是 p, q 也存在着限制条件。 q 表示去边事件, 如果 q 大于某个域值, 去边事件占优势, 网络就呈现衰退的现象。通过对网络的经验观察, Internet 度分布指数的一般范围是 $2 < \gamma < 3$, 令

$$2 < 1 + 1/B(p, q) < 3$$

解得: 在 $0 < q < 1/3$ 的范围内都能够得到比较接近实际网络幂律度分布指数的实例。

这里, 根据偏好依附的性质, 大度数的点起到连接网络上大量节点的 hub 的作用, 为网络提供连接的捷径, 因而这个模型很自然地 and BA 图一样表现出良好的 small-world 性质。

下面计算聚集程度, 设 $P(s, s', t)$ 为 t 时刻两个点相互连接的概率, 并假设 $s > s'$, 得到

$$\begin{aligned} P(s, s', t) = & (1 - p - q)m \frac{k_s(s')}{2ms'} \\ & + \frac{p}{2} \left(\frac{m}{N(t)} \frac{mk_s(t)}{\sum_j k_j(t)} + \frac{m}{N(t)} \frac{mk_{s'}(t)}{\sum_j k_j(t)} \right) \\ & - q \frac{m}{N(t)} \frac{m}{N(t)} \end{aligned} \quad (7)$$

其中, 右边第一项表示点 s 增加时, 所引出的边对 s, s' 两点互连的贡献; 第二项表示在 t 时刻, 在已存在点之间新增的 m 条边对作为已存在的点 s 和 s' 互相连接的贡献; 第三项表示断开的边对两点互连的影响。某个点 l 在时刻 t 的聚集程度是一个比率, 即点 l 的邻居间相互连接的边数与所有可能的边的总数 $k_l(t)[k_l(t) - 1]/2$ 之比, 如下式(8):

$$c_l(t) = \frac{1}{k_l(t)[k_l(t) - 1]} \cdot \int_1^t ds' \int_1^t ds P(l, s, t) P(s, s', t) P(s', l, t) \quad (8)$$

结合上面的概率公式, 设

$$k_l(t)[k_l(t) - 1] \simeq k_l(t)^2$$

可解得平均聚集程度的一个近似值:

$$\langle c \rangle_N \approx \frac{(\ln N)^2}{\alpha N^\beta} \quad (9)$$

其中的参数 α, β 是以 p, q, m 为变量的表达式。

从这个结果可以看到, 聚集程度随着网络的规模而变化, 并且变化的程度决定于 p, q, m 的取值。如果取值适当, 可令聚集程度随网络规模发生极其缓慢的变化, 这样可以更接近现实网络中的聚集程度几乎恒定的特征。而在 BA 模型中, $\langle c \rangle_N$ 值随着网络规模的增加迅速衰减。相比之下, VBA 模型的可调聚集程度特性优于 BA 模型。

根据以上对 VBA 模型的拓扑度量值的分析, 只要选择适当的参数, 该模型可以生成完好的符合 Internet 具有的幂律特征和小世界特性的网络, 并且对网络演化过程中的事件描述更切合实际的 Internet, 优于 BA 模型。

4 结束语

VBA 模型假设在时间 t , 以概率 q 断开的 m 条边是随机从已存在的点中选择。这个假设不完全符合真实 Internet 的情况, 对 AS 图的研究表明结点平均度数较低的 AS 出生或死亡的数量很大^[6], 而平均度数较大的 AS 几乎不发生变化。也就是说, 边倾向于从度数较小的结点断开, 而不是随机断开。这样可以对 VBA 模型的规则 1 做改进, 例如, 可以以概率 $\beta/k_i(t)^\alpha$ 选择要断开的边的端点, 这里, α, β 是需根据实际数据集的分析获得的参数。对模型的进一步改进工作有待继续研究和讨论。

参考文献:

- [1] Zhang Yu, Zhang Hong-li, Fang Bin-xing. A survey on Internet topology modeling[J]. Journal of Software, 2004, 15(8):1220-1226.
- [2] WU Jin-shan, DI Zeng-ru. Complex networks in statistical physics[J]. Progress in physics, 2004, 24(1):18-46.
- [3] Pastor-Satorras R, Vespignani A. Evolution and Structure of the Internet[M]. Cambridge: Cambridge university press, 2004.
- [4] Barabasi A L, Albert R. Emergence of Scaling in Random Networks[J]. Science, 1999(8):509-512.
- [5] Albert R, Barabási A L. Topology of evolving networks: local events and universality[J]. Physical Review Letters, 2000, 85(24):5234-5237.
- [6] Qian C, Chang H, Govindan R, et al. The origin of power laws in Internet topologies revisited[A]. in INFOCOM 2002, Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, Proceedings IEEE, Vol. 2 [C]. Los Alamitos, California: IEEE Computer Society Press, 2002. 608-617.

本刊自 2006 年更名为《计算机技术与发展》
欢迎投稿, 欢迎订阅。