

一种基于适应度和节点聚类的 P2P 拓扑建模方法

张 珂, 黄永峰, 李 星

(清华大学电子工程系, 北京 100084)

摘 要: 通过对 Gnutella 网络拓扑实测数据的分析, 得到了 Gnutella 网络的累积节点度分布、聚类系数、平均路径长度和 rich-club 系数等拓扑特征. 在拓扑特征分析的基础上, 通过引入适应度、邻居数限制和节点聚类等机制, 提出了面向非结构化 P2P 网络拓扑的 FCL (Fitness model with tunable Clustering and neighbours Limitation) 模型, 并给出了建模算法. 通过实验分析, 证明了该模型的有效性.

关键词: P2P 网络; 适应度; 节点聚类; 无尺度网络; 小世界网络

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2010) 07-1634-07

A Model for Topology of P2P Network Based on Fitness and Node Clustering

ZHANG Ke, HUANG Yong-feng, LI Xing

(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: The measuring topology data of Gnutella network were analyzed, then cumulative degree distribution function, clustering coefficient, average path length and rich-club coefficient of Gnutella network were extracted from analysis result. By introducing fitness, nodes with neighbours limitation and nodes clustering, the FCL (Fitness model with tunable Clustering and neighbours Limitation) model and the algorithm is proposed. Simulation indicated that the model is valid.

Key words: peer-to-peer (P2P) network; fitness; node clustering; scale-free network; small-world network

1 引言

P2P 被认为是改变 Internet 的新一代网络技术, 在应用领域和学术界取得了广泛的重视和成功, 目前已经占据了 Internet 超过一半的带宽资源^[1]. P2P 网络可以分为结构化与非结构化两种, 其中非结构化 P2P 具有易于部署和鲁棒性强等特点, 得到了大规模应用, 目前流行的 PPLive、BitTorrent、KaZaA、eMule 和 Gnutella 都属于此类^[2]. 本文主要研究非结构化 P2P 网络, 论文中出现的 P2P 均特指非结构化 P2P 网络.

发现 P2P 网络的拓扑特征并建立相应的演化模型是该领域的一个重要研究内容, 也是对 P2P 技术进行优化与改进的基础. 本文结合复杂网络^[3]中的相关概念, 对 P2P 网络 (Gnutella) 的实测拓扑数据进行分析, 得到其拓扑特征及演化规律. 在分析结果的基础上, 进一步提出基于适应度和节点聚类的 P2P 网络拓扑建模方法, 并通过对该模型的仿真实验分析, 验证模型的有效性.

2 相关工作

2.1 P2P 网络拓扑分析的相关复杂网络概念

本文将利用复杂网络理论中的节点度、聚类系数、平均路径长度和 rich-club 系数等概念对 P2P 网络拓扑进行分析, 因此首先对这些概念进行简单介绍. P2P 网络可以定义为一个无向无权简单网络 $G(V, E)$, 其中 V 和 E 分别为网络的节点集和边集^[3].

通常认为度可用于描述节点的“重要性”, 度越大的节点越重要. 对于 V 中的任意节点 v_i , 其节点度 k_i 定义为网络 G 中与该节点相连的节点数目^[4]. 节点度分布函数 $P(k)$ 可以描述网络中各种节点所占的比例, 定义为网络中节点度为 k 的节点数目与网络总节点数 $N = |V|$ 之比^[4~6]. 实测的网络拓扑数据包含有统计噪声, 而节点度分布对这些噪声的鲁棒性较差. 因此, 文献^[5, 7, 8]提出了利用累积节点度分布函数 $P^G(k)$ 描述网络的节点度分布, 定义为:

$$P^G(k) = \sum_{k_0=k}^{\infty} P(k_0) \quad (1)$$

实际网络中一个节点的两个邻居节点很可能也互相为邻居,这种属性称为网络的聚类属性.节点 v_i 和网络 G 的聚类属性由聚类系数 C_i 和 C 描述.目前有多种聚类系数的定义被提出^[9~11],使用最广泛的是 Watts 和 Strogatz 提出的如下定义^[9]:

$$C_i = \frac{2T_i}{k_i(k_i - 1)} \quad (2)$$

其中, T_i 为节点 v_i 的邻居节点之间实际存在的边数. C 定义为所有节点聚类系数 C_i 的平均值.

网络 G 中两个节点 v_i 和 v_j 之间的距离 d_{ij} 定义为这两个节点的最短路径上的边数^[5].网络 G 的平均路径长度 L 定义为 G 中所有节点对之间距离 d_{ij} 的平均值^[5].

实际网络中 rich-club 现象的存在极大地缩短了平均路径长度,文献[12]提出了 rich-club 系数 $\varphi(r)$ 用于量化 rich-club 现象,定义为由网络 G 中度最高的比例为 r 的节点组生成的子图中实际的边数与可能最多边数之比.

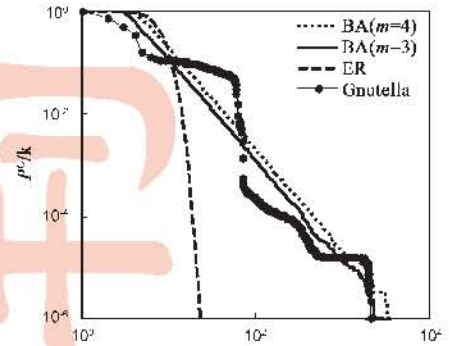
2.2 Gnutella 网络的拓扑特征

Gnutella 是最有代表性的非结构化 P2P 网络之一,本文首先对 Gnutella 网络的实际拓扑进行分析,作为进一步设计 P2P 网络模型的主要依据.本文以 ION P2P 项目*对 Gnutella 网络拓扑的测量结果为基础,给出其累积节点度分布、聚类系数、平均路径长度和 rich-club 系数等拓扑特征.为了便于分析,本文分别选取了 2004-9-27 和 2005-2-2 两个时间的拓扑数据,共计约 200MB 数据,约 40 万条记录.由于实测拓扑数据存在误差,导致网络的不连通性,所以本文对网络的最大连通子图进行分析.这两个网络的节点规模分别为 928486 和 1307360,最大连通子图规模分别为 928440 和 1307275.舍弃的节点数的比例小于 0.1%,并且都是 k 很小的节点,所以,最大连通子图完全可以代表原始网络的拓扑特征.

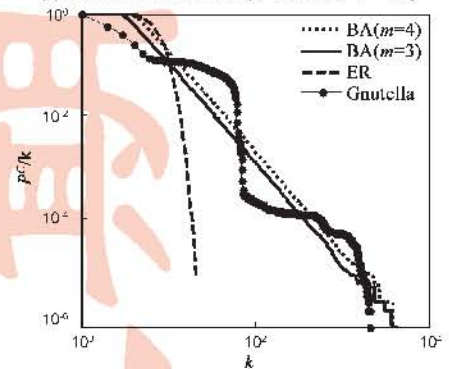
2.2.1 节点度分布

Gnutella 网络及同规模的 BA 无尺度网络^[3,6]和 ER 随机网络^[3]的累积节点度分布如图 1 所示,横坐标为节点度 k ,纵坐标为公式(1)的计算结果,二者均取常用对数值.本文所说的同规模是指网络的节点数 N 相等并且平均节点度 $\langle k \rangle$ 近似相等.从图 1 可以看出, Gnutella 网络的累积节点度分布 $P^G(k)$ 具有重尾的特征,但不是幂率函数. $P^G(k)$ 围绕幂率函数上下波动,在 $k = 70$ 附近存在一个较大的负向阶跃.网络最大节点度 $\max(k)$ 也可以印证 Gnutella 网络的累积节点度分布的重尾特征,图 1(a)中 Gnutella 网络、ER 随机网络和 2 个 BA

无尺度网络的最大节点度分别为:2157、23、3093 ($m = 3$) 和 3681 ($m = 4$),图 1(b)中 4 个网络的最大节点度分别为:2079、21、4010 ($m = 3$) 和 4358 ($m = 4$).



(a) 2004-9-27 的拓扑数据 ($N=928440, \langle k \rangle=6.9$)



(b) 2005-2-2 的拓扑数据 ($N=1307275, \langle k \rangle=6.75$)

图 1 Gnutella 网络的累积节点度分布

2.2.2 小世界特征

小世界网络同时具有较高的聚类系数和较小的平均路径长度^[9],为了研究 Gnutella 网络是否具有小世界特征,需要计算其聚类系数 C 和平均路径长度 L .然而,对于 Gnutella 这样的 10^6 节点规模的网络来说,依目前计算机的处理能力,不论使用何种算法,计算平均路径长度都是不可能的.因此,本文将依据定理 1 计算网络平均路径长度的上界.

定理 1 假设 v_k 是网络 $G(V, E)$ 中的任意一点,由 v_k 到网络中所有其它节点的最短路径长度的平均值记为 $d_k = (\sum_j d_{kj}) / (N - 1)$,则 G 的平均路径长度 L 具有上界 $2d_k$,即:

$$L \leq 2d_k, \quad \forall v_k \in V \quad (3)$$

$$\begin{aligned} \text{证明} \quad L &= \frac{1}{N(N-1)} \sum_i \sum_j d_{ij} \\ &\leq \frac{1}{N(N-1)} \sum_i \sum_j (d_{ik} + d_{kj}) \end{aligned}$$

* ION P2P 是由俄勒冈大学的多媒体与互联网研究组开展的研究项目,该项目研究如何进行精确的 P2P 系统测量,相关数据可以在如下网址下载: <http://mirage.cs.uoregon.edu/P2P/>.

$$= \frac{1}{N-1} \sum_i d_{ik} + \frac{1}{N-1} \sum_j d_{kj} = 2d_k$$

证毕.

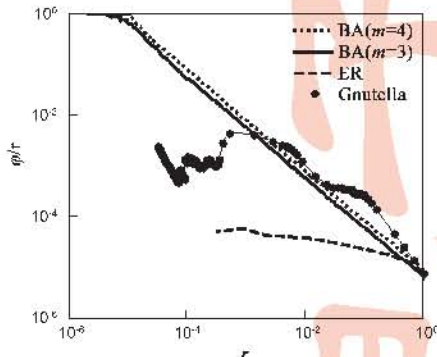
在实际的分析工作中,本文选择节点度最高的节点作为 v_k 来计算 L 的上界. Gnutella 网络及同规模 ER 随机网络和 WS 小世界网络的聚类系数 C 和平均路径长度上界 $\sup(L)$ 如表 1 所示. 从表中可以看出, Gnutella 网络的 C 远远大于 ER 随机网络, 而 $\sup(L)$ 则小于 ER 随机网络. 通过调整 WS 小世界模型的重连概率 p , 可以得到与 Gnutella 网络近似的聚类系数, 此时的 $\sup(L)$ 基本与 Gnutella 网络同数量级. 上述分析说明 Gnutella 网络是一个小世界网络.

表 1 Gnutella 网络的聚类系数与平均路径长度

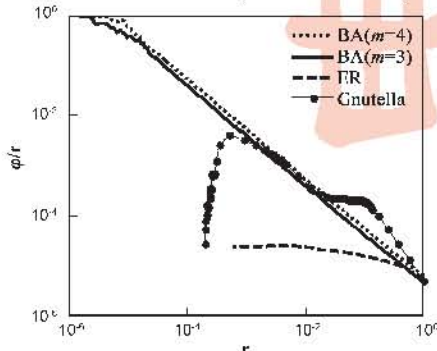
| 网络 | 2004-9-27 ($N = 928440$) | | 2005-2-2 ($N = 1307275$) | |
|--------------|----------------------------|-----------|----------------------------|-----------|
| | C | $\sup(L)$ | C | $\sup(L)$ |
| Gnutella | 0.02 | 9.928 | 0.0098 | 6.9152 |
| | 7.4×10^{-6} | 13.2524 | 5.2×10^{-6} | 13.9158 |
| WS ($K=6$) | $(p=0.68)$ | | $(p=0.746)$ | |
| | 0.02 | 15.3018 | 0.0098 | 16.6174 |
| WS ($K=8$) | $(p=0.687)$ | | $(p=0.752)$ | |
| | 0.02 | 13.1807 | 0.0098 | 14.1102 |

2.2.3 rich-club 现象

Gnutella 网络及同规模 ER 随机网络和 BA 无尺度网络的 rich-club 系数如图 2 所示, 横坐标为节点比例 r , 纵坐标为 rich-club 系数 $\varphi(r)$, 二者均取常用对数值. 从



(a) 2004-9-27的拓扑数据($N=928440, \langle k \rangle=6.9$)



(b) 2005-2-2的拓扑数据($N=1307275, \langle k \rangle=6.75$)

图 2 Gnutella 网络的 rich-club 系数

图中可以看出, Gnutella 网络的 $\varphi(r)$ 在 $(0.0004, 1]$ 的范围内远高于 ER 随机网络, 甚至还高于 BA 无尺度网络. 然而, 在 $(0, 0.0004]$ 范围内, Gnutella 网络的 rich-club 系数急剧下降. 以上分析说明 Gnutella 网络具有不完全的 rich-club 现象, 最核心节点之间连通性较差.

从上述三方面的分析结果可以看出, Gnutella 网络的节点度具有重尾分布的特征, 累积节点度分布围绕幂率函数上下波动, 并且存在一个较大的负向阶跃; Gnutella 是小世界网络; Gnutella 网络具有不完全的 rich-club 现象, 网络的 rich-club 系数较高, 但最核心节点之间的连通性较差. Gnutella 网络拓扑的这些统计特征与经典的复杂网络理论^[3,6,9]发现的拓扑特征并不十分吻合, 从而导致现有网络模型不能正确刻画 Gnutella 网络的拓扑. 因此, 有必要根据 P2P 网络的内在机制, 对现有复杂网络模型进行改进, 得到更精确的 P2P 网络模型.

3 FCL 模型

为体现 P2P 网络拓扑的演化过程, 同时结合 2.2 节得出的实际测量数据呈现的多项统计特征, 以文献[13]提出的适应度模型为基础, 提出 FCL (Fitness model with tunable Clustering and neighbours Limitation) 模型.

3.1 基于适应度的优先连接

BA 无尺度网络模型证明了实际网络中包含的节点规模增长和优先连接的机制是产生网络节点度幂率分布的内在原因^[3,6]. 但是, 根据 2.2.3 节的结论, BA 网络的 rich-club 系数与 Gnutella 网络并不完全一致, 导致这个差异的原因分析如下.

BA 网络在 t 时刻的边密度为:

$$\rho = \frac{mt}{t(t-1)/2} = \frac{2m}{t-1} \quad (4)$$

这说明, 老节点之间的连通性比新节点好, 进而导致老节点的度比新节点高. 又由于 BA 模型的连接概率为^[6]:

$$\Pi_i = \frac{k_i}{\sum_j k_j} \quad (5)$$

则老节点将获得更多的优先依附连接. 最终, 网络中连通性很好的初始节点成为网络的最核心的部分, 从而形成了图 2 所示的 rich-club 现象.

根据文献[12]的研究结果, BA 模型可以较好地刻画 Internet-AS 网络中的 rich-club 现象. Internet-AS 网络的最核心节点都是 Transit AS, 它们为了向其它 Stub AS 提供数据转发服务, 通常都会刻意保持很好的连通性, 从而形成较高的 rich-club 系数. 然而, P2P 是纯粹的自组织网络, 没有相应机制保持核心节点之间的连通性, 因此产生了不完全的 rich-club 现象.

在 P2P 网络中, 节点的度及其增长速度不仅与该

节点的年龄有关,还与该节点的内在性质相关,这些内在性质包括节点的计算性能、网络接口带宽和节点存储内容等.文献[13]提出的适应度模型(fitness model)可以很好地描述 P2P 网络节点的内在性质,该模型的构造算法与 BA 网络基本一样,但是将优先连接概率修改如下:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \quad (6)$$

其中, η_i 为节点 v_i 的适应度,所有节点的适应度满足同一个概率分布 $P(\eta)$. 本文将以适应度模型为基础,建立 P2P 网络模型 - FCL 模型.

3.2 节点的邻居数限制

根据 2.2.1 节的分析结果, Gnutella 网络的累积节点度分布 $P^c(k)$ 并不是幂率函数,而是围绕幂率函数上下波动,并且在 $k=70$ 存在一个明显负向阶跃. $P^c(k)$ 的负向阶跃说明网络中包含了大量节点度在 70 附近的节点,导致这一现象的原因是 Gnutella 软件的一个主要版本 LimeWire 对节点的邻居超级节点数进行了限制(默认值为 30,用户可自行修改),以防止个别节点的负担过重.

在实际的各种 P2P 网络中,节点的邻居数(即节点度 k)通常都会受到限制,一方面是 P2P 软件实现算法的人为限制,另一方面是节点自身能力的客观限制.为了在模型中体现这一特征, FCL 模型将对适应度模型进行改进,向模型中引入受限节点的概念.在 FCL 模型中,对于每一个新加入网络的节点,首先依概率 p 确定其是否为受限节点,如果是,则依概率分布 $P(k_{\max})$ 为其分配节点度上界 k_{\max} . 在网络的后续增长过程中,如果受限节点的度达到了 k_{\max} ,它将不能够再被新节点连接.

3.3 节点聚类行为建模

2.2.2 节的分析结果已经证明 Gnutella 是一个小世界网络,这印证了文献[14]的研究结论.大量研究表明,多数 P2P 都是小世界网络,都具有较高的聚类系数^[15]. 聚类系数描述了网络聚类程度的高低,为了研究 P2P 网络中高聚类系数产生的机制,本文将对节点聚类行为建模.

从表 1 也可以看出 WS 网络模型完全可以生成符合 Gnutella 小世界特征的网络.然而 WS 模型的高聚类系数继承自初始的最近邻耦合网络^[9],实际网络显然不可能从规则网络继承高聚类系数.文献[16]利用配置网络的方法,生成聚类系数可调的任意度分布网络,但这类配置网络模型并不能揭示出高聚类系数产生的演化机制.文献[17,18]利用局域优先连接机制设计了聚类系数可调的无尺度网络模型,可同时产生实际网络

的多项拓扑特征.这些模型虽然各不相同,但它们的本质都是通过调整网络中的三角形的数目达到调整网络聚类系数的目的.本文将节点聚类行为建模为切入点,通过节点聚类调整网络中的三角形数量,进而产生通用的聚类系数可调网络模型.

节点聚类行为是指网络中具有某种共同或相似特征的节点之间建立连接的行为.比如在 P2P 文件共享网络中,用户为了下载自己感兴趣的文件而连接那些拥有目标文件的节点.此时网络中的文件分布在一定程度上反映了节点之间的相似度,这种基于节点相似度的连接行为都是节点聚类行为.

节点聚类行为模型的具体算法如下(以下叙述均规定网络为简单图):

(1) 从初始网络开始:初始网络可以为任意网络.

(2) 节点聚类:依均匀概率从网络中选取一个节点 v_i , 并从与该节点相连的边中任选一条,重连该边到 v_i 的一个间接邻居节点 v_j (邻居节点的邻居节点). v_j 在随机化重连中被选择的概率为:

$$\Pi_j = \frac{n_{ij}^a}{\sum_{v_j \in V_i} n_{ij}^a} \quad (7)$$

其中, n_{ij}^a 为 v_i 和 v_j 之间的公共邻居节点数, V_i 为 v_i 的间接邻居节点集.

(3) 重复 S 次节点聚类.

节点聚类行为模型中仅考虑了最简单的节点聚类行为:随机化重连间接邻居.采取随机化重连而非随机化加边的方式演化网络是为了保持固定的平均节点度 $\langle k \rangle$, 便于模型之间的比较.重连过程中仅考虑间接邻居节点是对节点聚类行为的一种抽象,因为大多数实际网络中的节点仅仅了解自己周围的局部信息,而非全局信息.此外,模型中用两个节点之间的公共邻居节点数 n_{ij}^a 度量节点相似度,这与实际网络是相符的.比如在 P2P 文件共享网络中,因为两个节点都对公共邻居节点上的文件感兴趣,所以公共节点越多说明两个节点的兴趣相似度越高.

3.4 FCL 模型算法

通过前 3 节的分析,建立基于适应度和节点聚类的 P2P 网络拓扑模型 - FCL 模型,该模型可以生成同时具备节点度重尾分布、小世界特征和不完全 rich-club 现象的网络.

FCL 网络模型的算法如表 2 所示.考虑到实际 P2P 网络中大多数节点都是计算性能、网络接口带宽和节点存储内容等内在性质近似的普通节点,只有少数节点具有特殊性质, FCL 网络模型采用正态分布描述节点适应度 η , 即 $P(\eta) = N(\mu_1, \sigma_1^2)$. 同样,考虑到大多数用户都不会修改 P2P 软件的最大邻居总数限制, FCL 模型

仍然采用正态分布描述受限节点的节点度上界 k_{\max} , 即 $P(k_{\max}) = N(\mu_2, \sigma_2^2)$.

FCL网络模型的主要参数如表3所示. 从FCL网络拓扑的构造算法容易看出, 当 N 足够大时, 网络的平均节点度 $\langle k \rangle \approx 2m$. σ_1 描述了节点适应度的异质性, 该参数越大, 节点之间的适应度差异越大. μ_2 描述了节点的度上界的平均值, 与实际P2P软件的默认值对应. σ_2 描述了用户修改默认最大邻居总数的意愿, 该值越大, 修改默认参数的用户越多. α 描述了节点聚类的微观属性, 该值越大, 则在每一次节点聚类行为中, 公共邻居数大的间接邻居节点被连接的概率越大, 即相似度高的间接邻居节点被连接的可能性越高. S 描述了节点聚类行为的宏观属性, 该值越大, 网络中的节点聚类行为越多, 节点的活跃程度越高. 当 $p=0$ 且 $S=0$ 时, FCL模型退化为适应度模型; 当 $p=0$ 、 $S=0$ 并且 $\sigma_1=0$ 时, FCL模型退化为BA无尺度网络模型.

表2 FCL网络模型的拓扑构建算法

| 步骤 | 名称 | 操作 |
|----|---------|---|
| 1 | 增长 | 从一个具有 m_0 个节点的全连接拓扑开始, 每次引入一个新的节点. |
| 2 | 分配适应度 | 依正态分布 $N(\mu_1, \sigma_1^2)$, 为新节点分配一个适应度 η . |
| 3 | 分配最大邻居数 | 首先依概率 p 确定新节点是否为受限节点, 如果是, 则依正态分布 $N(\mu_2, \sigma_2^2)$ 为其分配一个最大邻居数 k_{\max} . |
| 4 | 优先连接 | 在新节点与 m 个节点度尚未达到上界 k_{\max} 的老节点之间建立连接, 已经存在的节点 v_i 被选中的概率由式(6)计算得到. |
| 5 | 重复增长 | 重复步骤1~4, 直到网络的节点规模达到 N . |
| 6 | 节点聚类 | 以步骤5得到的网络为初始网络, 按照3.3节的算法进行节点聚类. |

表3 FCL网络模型的主要参数

| 参数 | 说明 |
|------------|---------------|
| N | 网络的节点数. |
| m | 每个新节点带入的连接数. |
| μ_1 | 节点适应度分布的均值. |
| σ_1 | 节点适应度分布的标准差. |
| p | 网络中受限节点的比例. |
| μ_2 | 受限节点的度上界的均值. |
| σ_2 | 受限节点的度上界的标准差. |
| α | 节点聚类行为的偏向性参数. |
| S | 节点聚类行为的发生次数. |

4 实验分析

FCL模型从适应度和节点聚类等方面对P2P网络进行建模, 为了验证其有效性, 并分析模型参数对网络结构的影响, 本文对FCL模型进行仿真实验分析. 对FCL模型的累积节点度分布、rich-club系数、聚类系数和

平均路径长度等参数进行分析, 所有结果均为100次仿真实验结果的平均值.

4.1 累积节点度分布

FCL模型的累积节点度分布如图3所示, 坐标系的意义与图1相同. 从图中可以看出, FCL网络的累积节点度分布是重尾分布, 并且可以产生负向阶跃. 影响阶跃位置与幅度的参数有3个, 分别是 p 、 μ_2 和 σ_2 . 参数 μ_2 是FCL网络中受限节点的度上界的均值, $P^C(k)$ 函数就在 $k = \mu_2$ 附近发生负向阶跃. 参数 σ_2 是FCL网络中受限节点的度上界的标准差, σ_2 值越大, 受限节点的度上界越分散, $P^C(k)$ 函数在 $k = \mu_2$ 附近的曲线也就越平滑. p 描述了FCL网络中限制邻居的比例, 值越大, $P^C(k)$ 函数的负向阶跃幅度越大.

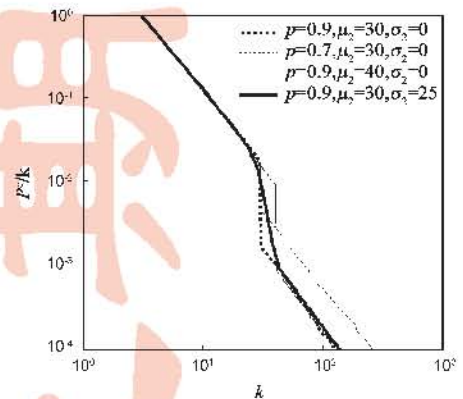


图3 FCL网络的累积节点度分布($N=10^4, m_0=7, m=3, \sigma_1=0, S=0$)

4.2 rich-club系数

FCL模型的rich-club系数如图4所示, 坐标系的意义与图2相同. 当 $p=0$ 并且 $\sigma_1=0$ 时, FCL模型退化为BA无尺度网络模型, 因此可以通过图4对FCL模型和BA模型进行比较. 从图中可以看出, FCL网络中最核心节点之间的连通性比BA网络差, 而影响这一连通性的参数是 p 和 σ_1 . p 越大, 网络中的高节点度节点越少, 网络的优先依附特征越不明显, 导致核心节点之间的连通性越差. σ_1 越大, 网络的异质性越大, 适应度对节点度

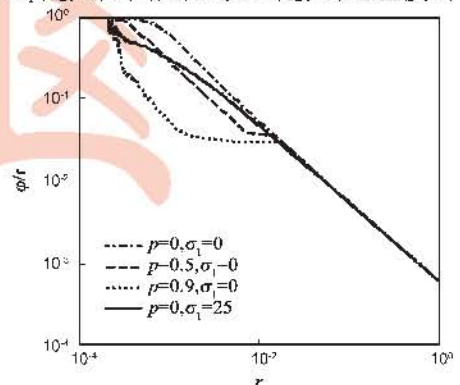


图4 FCL网络的rich-club系数($N=10^4, m_0=7, m=3, \mu_1=10, \mu_2=30, \sigma_2=0, S=0$)

的影响越大,节点年龄对节点度的影响也就越小,核心节点之间的连通性也就越差.由于图 3 和图 4 是多次实验结果的平均值,因此显得比图 1 和图 2 更加平滑,实际上单次实验结果的函数曲线与实际网络一样参差不齐.

4.3 聚类系数

FCL 模型的聚类系数 C 与参数 α 和 S 之间的关系如图 5 所示,横坐标为节点聚类的次数 S ,纵坐标为网络聚类系数 C .从图中可以看出,网络聚类系数随着 α 和 S 的增大而增加. α 越大或者 S 越大,分别说明节点聚类的偏向性越强和节点聚类次数越高,网络中新产生的三角形也越多,网络的聚类系数就越高.当 S 在 $[10^4, \infty)$ 范围逐步增大时,网络聚类系数能逐步增大至 0.8 以上,然而考虑到下面两个原因,图 5 并没有给出这部分结果.第一,在 $[10^4, \infty)$ 范围内网络的节点度分布将发生明显变化;第二,实际 P2P 网络的聚类系数通常不会超过 0.1.

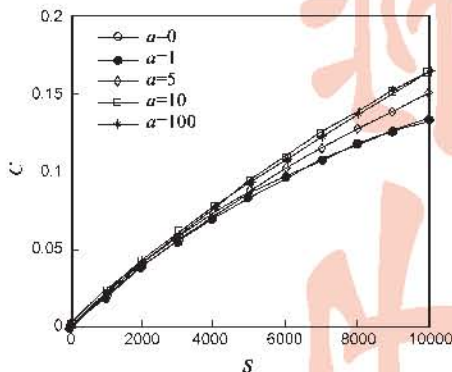


图5 FCL网络的聚类系数
($N=10^4, m_0=7, m=3, \mu_1=10, \sigma_1=1, p=0.9, \mu_2=30, \sigma_2=1$)

4.4 平均路径长度

FCL 模型的平均路径长度 L 与参数 α 和 S 之间的关系如图 6 所示,横坐标为节点聚类的次数 S ,纵坐标为网络平均路径长度 L .从图中可以看出 L 随着 S 的增大而增大,随着 α 增大先增大后减小.然而, L 的增减

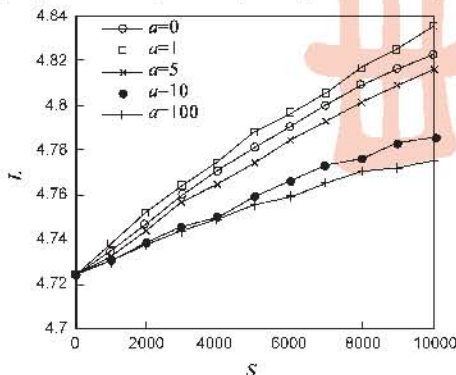


图6 FCL网络的平均路径长度
($N=10^4, m_0=7, m=3, \mu_1=10, \sigma_1=1, p=0.9, \mu_2=30, \sigma_2=1$)

幅度很小,并没有数量级性质的变化.综合 4.3 节的分析结果,FCL 模型可以生成小世界网络.

4.5 网络连通性

WS 小世界模型中当随机化重连概率较大的时候,网络会出现不连通的情况^[3],仿真实验表明 FCL 小世界模型也存在同样的问题,当参数 α 和 S 都很大时,FCL 网络可能是不连通的.为了克服这一缺点,可以结合 NW 小世界模型思路,采用随机化加边的方式描述节点的聚类行为,从而得到连通的小世界网络.

5 结束语

本文分析了 Gnutella 网络的实测拓扑数据,给出了累积节点度分布、rich-club 系数、聚类系数和平均路径长度上界等计算结果.在此基础上,通过引入适应度、邻居数限制和节点聚类机制,提出了 FCL 模型.仿真实验表明,FCL 模型可以生成更加符合实际 P2P 网络统计特征的拓扑,该模型对进一步研究 P2P 网络拓扑的演化机制具有重要意义.

参考文献:

- [1] 陈贵海,李振华.对等网络:结构、应用与设计[M].北京:清华大学出版社,2007.
Guihai Chen, Zhenhua Li. Peer-to-Peer Network: Structure, Application and Design[M]. Beijing: Tsinghua University Press, 2007. (in Chinese)
- [2] Kin-Wah Kwong, Danny H K Tsang. Application-aware topology formation algorithm for peer-to-peer networks[A]. In Proc. IEEE ICC'07[C]. Glasgow: IEEE, 2007. 73 - 79.
- [3] Newman M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167 - 256.
- [4] Douglas B West. Introduction to Graph Theory (2nd Edition) [M]. New Jersey: Prentice Hall, 2000.
- [5] 汪小帆,李翔,陈关荣.复杂网络理论及其应用[M].北京:清华大学出版社,2006.
Wang Xiao-fan, Li Xiang, Chen Guan-rong. Complex network Theory and Application [M]. Beijing: Tsinghua University Press, 2006. (in Chinese)
- [6] Albert-László Barabási, Réka Albert. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509 - 512.
- [7] Romualdo Pastor-Satorras, Alessandro Vespignani. Evolution and Structure of the Internet: A Statistical Physics Approach [M]. Cambridge: Cambridge university press, 2004.
- [8] G Siganos, M Faloutsos, P Faloutsos, C Faloutsos. Power-laws and the AS-level internet topology[J]. IEEE/ACM Trans. on Networking, 2003, 11(4): 514 - 524.
- [9] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393(6684): 440 - 442.

- [10] M E J Newman, S H Strogatz, D J Watts. Random graphs with arbitrary degree distributions and their applications[J]. *Physical Review E*, 2001, 64:026118.
- [11] Sara Nadiv Soffer, Alexei Vázquez. Network clustering coefficient without degree-correlation biases[J]. *Physical Review E*, 2005, 71:057101.
- [12] Shi Zhou, Raul J Mondragon. The rich-club phenomenon in the Internet topology [J]. *IEEE Communications Letters*, 2004, 8(3):180 - 182.
- [13] Ginestra Bianconi, Albert-László Barabási. Bose-einstein condensation in complex networks[J]. *Physical Review Letters*, 2001, 86(24):5632 - 5635.
- [14] Ripeanu M, Foster I, Iamnitchi A, Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design[J]. *IEEE Internet Computing*, 2002, 6(1):50 - 57.
- [15] 刘琼, 徐鹏, 杨海涛, 彭芸. Peer-to-Peer 文件共享系统的测量研究[J]. *软件学报*, 2006, 17(10):2131 - 2140.
LIU Qiong, XU Peng, YANG Hai-Tao, PENG Yun. Research on measurement of peer-to-peer file sharing system[J]. *Journal of Software*, 2006, 17(10):2131 - 2140. (in Chinese)
- [16] M Angeles Serrano, Marian Boguna. Tuning clustering in random networks with arbitrary degree distributions[J]. *Physical Review E*, 2005, 72:036133.
- [17] Zhongzhi Zhang, Lili Rong, Bing Wang, Shuigeng Zhou, Jihong Guan. Local-world evolving networks with tunable clustering[J]. *Physica A*, 2007, 380:639 - 650.
- [18] Bing Wang, Huanwen Tang, Zhongzhi Zhang, Zhilong Xiu. Evolving scale-free network model with tunable clustering[J]. *International Journal of Modern Physics B*, 2005, 19(26):3951 - 3959.

作者简介:



张珂 男, 1978 年 3 月生于河南禹州. 2003 年在清华大学获得工学硕士学位. 现为清华大学电子工程系博士研究生. 主要从事复杂网络, 网络多媒体等方面的研究工作.
E-mail: zhke@ovi.com

黄永峰 男, 1967 年 12 月生于湖北赤壁. 现为清华大学电子工程系副教授. 获得华中理工大学计算机系统结构专业博士学位. 研究方向为计算机网络, 多媒体通信与网络安全.

李星 男, 生于 1956 年. 现任清华大学电子工程系教授, 博士生导师, 网络与人机语音通信研究所所长, 网络中心副主任. 1989 年获美国 DREXEL 大学电机与计算机工程系博士学位. 研究领域包括统计信号处理, 多媒体通信和计算机网络.