

一种基于多项复杂特征的 Internet 路由级拓扑建模方法

张 昕, 赵 海, 李 超

(东北大学信息科学与工程学院, 辽宁沈阳 110004)

摘 要: 通过对 CAIDA 多监测点获得的测量数据的分析, 提取出表征 Internet 路由级拓扑规律的多项特征值. 在特征值分析的基础上, 综合考虑了路由级拓扑的动态变化, 拓扑的层次性, 以及节点间非线性连接的特点, 提出了面向 Internet 路由级拓扑的 DNLH(Dynamic Non Linear Hierarchy) 模型, 并给出建模算法. 以实际数据为依据, 结合遗传算法优化了节点间非线性连接参数. 通过实验分析, 证明了该建模方法的有效性.

关键词: Internet; 路由级拓扑; 幂律分布; 遗传算法

中图分类号: TP277 **文献标识码:** A **文章编号:** 0372-2112(2008)01-0057-07

A Model for Router-Level Topology of Internet Based on Complex Characters

ZHANG Xin, ZHAO Hai, LI Chao

(School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China)

Abstract: The measuring data which were got by CAIDA multi monitors were analyzed, then multi eigenvalues describing the router level topology of Internet were extracted from analysis results. Considering the dynamic change, the hierarchy and the non linear connecting character between the nodes in the router level topology of Internet, the DNLH(dynamic non linear hierarchy) model and the algorithm of the model were put forward. According to factual data, the non linear connecting parameters between the nodes were optimized by genetic algorithm. Through quantitative experimenting analysis, this model is found valuable.

Key words: Internet; router level topology; power law distribution; genetic algorithm

1 引言

Internet 作为当今人类社会信息化的标志, 其规模正飞速增长, Internet 拓扑建模研究就是寻找在这个看似混乱的网络之中是否包含有一定的客观规律^[1]. 发现 Internet 拓扑的内在机制是认识 Internet 的必然过程, 是在更高层次上开发利用 Internet 的基础. 然而, Internet 与生俱来的异构性、动态性、发展的非集中性以及如今庞大的规模都给拓扑建模带来巨大挑战^[2]. 在海量授权数据的统计分析的基础上, 针对现有 Internet 模型的不足, 提出一个对真实 Internet 路由级拓扑符合程度较高的建模方法, 并通过分析和实验说明其合理性及有效性. 本文其余部分按下述方式组织: 第 2 节介绍研究现状及背景知识, 介绍本文数据来源, 并依实际数据给出 Internet 路由级拓扑的部分统计规律; 第 3 节介绍本文建模思想及建模算法, 并给出模型中各待定参数的确定方法; 第 4 节为仿真实验部分, 定性定量的分析了本文模型的优

势所在; 第 5 节为全文总结, 并对未来工作做一定程度的展望.

2 相关工作

2.1 Internet 拓扑建模研究现状

Internet 拓扑建模的研究一直是学术界的一个热点, 在早期对随机图产生器的研究之后, 目前的建模方法主要分为两类, 一类是基于广域/局域网层次结构设计思想的拓扑生成方法, 如 Tiers(等级)模型^[3]与 Transit-stub 模型^[4], 此类模型建模效果较第二类方法要差^[5]; 另一类是于 Faloutsos 等人发现 Internet 拓扑结构中存在幂律(power law)规律^[6]之后, 从而产生基于网络节点度分布的拓扑建模方法. 此类模型中 Inet^[7]是一个考虑了节点度分布层次的静态模型, 它不能反映 Internet 动态变化情况; 其余动态模型如 AB 模型^[8]、GLP 模型^[2]、DP 模型^[5]等均采用节点间线性优先连接的方法, 不能充分体现 Internet 增长过程中高连接度优先的情

况;BRITTE 模型^[9]注重更好的交互性,但模型提供的最优参数使其接近早期的 BA 线性模型^[10],劣势明显;TANG 模型^[11]采用节点间非线性优先连接的方法,主要面向 AS 级拓扑,与路由级拓扑在多项指标上均有所不同,适应性有限.因此可以看出,研究 Internet 路由级拓扑,需充分考察实际测量数据的特性后,方能给出有效模型.

2.2 Internet 拓扑分析背景知识

Internet 作为一个高度复杂的系统,其拓扑结构不能由简单的统计特征值(如节点数、连接数、平均最短路径长度等)来直接刻画.文献[6]发现 Internet 中节点度分布呈现高度扭曲,并近似服从幂律分布的规律.该文献还提出了多种幂律分布形式,包括 frequency-degree 幂律与 degree-rank 幂律,来描述 Internet 拓扑结构特征.文献[12]指出,由于 frequency-degree 幂律相当于考察的是度值 d 的频率的概率密度函数(probability density function, 简称 PDF),而累积分布比概率密度的统计对于分析样本的鲁棒性更好.因此,文献[12]考察了频率的互补累积分布函数(complementary cumulative distribution function, 简称 CCDF),发现其与 d 也呈幂律关系,该幂律被称为 CCDF(d)-degree 幂律.表 1 给出文中使用的变量符号及定义,表 2 给出 3 种幂律分布函数的描述.

表 1 变量符号定义表

| 符号变量 | 定义 |
|-----------|----------------------------------|
| N | 拓扑图中节点数 |
| E | 拓扑图中边数 |
| $degree$ | 节点度值, 简记为 d |
| \bar{d} | 拓扑图中节点平均度, $\bar{d} = 2E/N$ |
| PDF | 度值等于 $degree$ 的节点的频数 |
| $CCDF$ | 度互补累积分布函数, 指度数大于 $degree$ 的节点累积数 |
| $rank$ | 在度降序排列中, 度值等于 $degree$ 的节点的排序(秩) |

表 2 三种幂律分布函数

| 幂律分布形式 | 函数形式 | 幂指数 |
|--------------------|--------------------|----------|
| frequency-degree | $PDF \propto d^R$ | R_{fd} |
| degree-rank | $d \propto rank^R$ | R_d |
| CCDF(d)-degree | $CCDF \propto d^R$ | R_{cd} |

2.3 Internet 路由级拓扑统计规律

对拓扑结构进行建模分析,最重要的一点是要得到其实际的统计规律,进而指导模型设计.本文以 CAIDA* 对 Internet 拓扑的测量结果为基础,给出其路由级节点度的 frequency-degree 幂律分布、CCDF(d)-degree 幂律分布、degree-rank 幂律分布、节点平均度以及叶子节点的比例等统计规律.因单点探测可导致的隐藏三角路由问题,从而会影响节点度相关数据的统计分析.为避免这种情况对分析结果的影响,本文选取基于 21 个监测点在同一时间段的拓扑测量结果,共计 7GB 数据,约 3 千万条记录,合并后进行统计分析.文献[12]指

出,Internet 路由级拓扑的各统计值随时间变化缓慢,且变化量微小,因此下述统计值可以作为进一步分析的基础数值.

• frequency-degree 幂律分布分析

Internet 路由级节点度的 frequency-degree 幂律分布情况及拟合曲线如图 1 所示,横坐标为节点度值 d ,纵坐标为节点频数 PDF,二者均取自然对数值.拟合幂指数 R_{fd} 为 2.141,很明显,在叶子节点部分以及节点度较高时拟合曲线偏差较大.

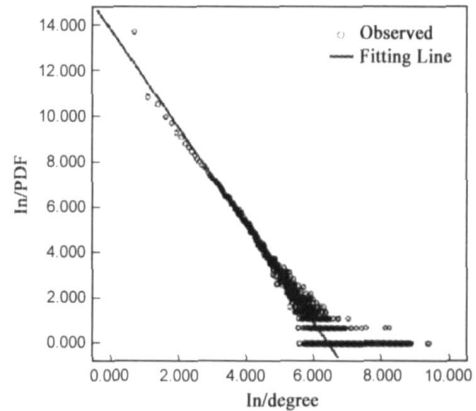


图 1 Internet 路由级节点度 frequency-degree 幂律分布

• CCDF(d)-degree 幂律分布分析

Internet 路由级节点度的 CCDF(d)-degree 幂律分布情况及拟合曲线如图 2 所示,横坐标为节点度值 d ,纵坐标为度数大于 d 的节点累积数 CCDF,二者均取自然对数值.拟合幂指数 R_{cd} 为 1.256,与图 1 类似,在叶子节点部分以及节点度较高时拟合曲线偏差较大,但整体拟合效果明显好于图 1,符合 CCDF(d)-degree 幂律分布对数据偏差鲁棒性较好的特点.

• degree-rank 幂律分布分析

Internet 路由级节点度的 degree-rank 幂律分布情况及拟合曲线如图 3 所示,横坐标为度降序排列中节点的排序(秩) $rank$,纵坐标为节点度值 d ,二者均取自然对数值.拟合幂指数 R_{dr} 为 0.831,与图 1、2 类似,在叶子节点部分以及节点度较高时拟合曲线偏差较大.

• 节点平均度及叶子节点比例

经计算,路由级 Internet 节点平均度 \bar{d} 为 4.868,路由级 Internet 拓扑中叶子节点占总节点数比重(N_{leaf}/N) 为 17.79%.

* CAIDA (The Cooperative Association for Internet Data Analysis, <http://www.caida.org>), 是一个对全球范围 Internet 结构及数据进行研究的国际合作机构,研究的主要内容包括 Internet 网络的产生、发展及演化趋势,以及 Internet 网络行为、动力、网络传播特征和 Internet 宏观拓扑结构的变化规律. CAIDA 在世界范围内的参与者主要分布在北美洲、欧洲的许多国家中的研究所、军事机构及高等学府中,其中东北大学复杂网络研究中心为中国 CAIDA 第一节点.

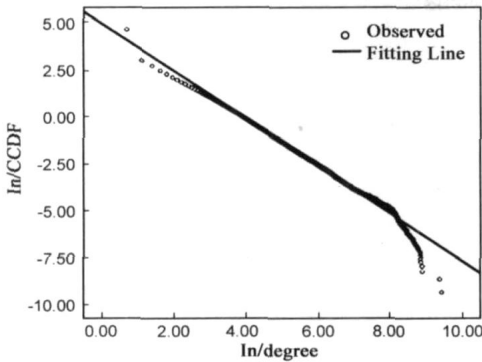
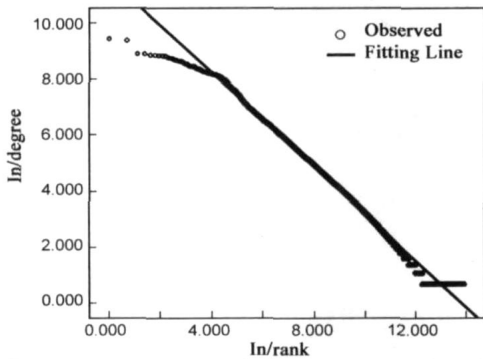
图 2 Internet路由级节点度 CCDF(d)-degree 幂律分布

图 3 Internet路由级节点度 degree-rank 幂律分布

由幂律分析结果可以看出, Internet 路由级拓扑节点度分布都具有明显的层次性, 可以分为 top 节点、主体节点以及叶子节点三层. top 节点是指拓扑中度值较大的部分节点, 其个数记做 N_{top} ; 主体节点是指拓扑图中除 top 节点以及叶子节点余下的部分, 其个数记做 N_{main} ; 叶子节点是指度为 1 的节点, 其个数记做 N_{leaf} . 需要说明一点, 该层次性指的是根据节点度值大小的不同, 将其划分为若干层次, 与文献[3, 4]中的广域/局域网层次, 以及文献[13]中的上下级层次关系不同.

文献[6, 7, 12, 14]指出, AS 级拓扑的节点度分布呈现一定的幂律规律, 并且节点度分布具有明显的层次性, 这与本节统计的 Internet 路由级拓扑相似. 但由文献中还可以得知, AS 级拓扑的节点平均度约为 6.3, 叶子节点比重约为 30%, 且 top 节点个数较少. 因此可以看出, 路由级拓扑与 AS 级拓扑存在相当程度的差别, 路由级建模算法可以部分借鉴 AS 级建模的方法, 但不能直接沿用.

3 DNLH 模型

为体现 Internet 路由级拓扑的动态变化, 同时考虑 2.3 节分析得出的实际测量数据呈现层次性的结论, 进一步结合计算得到的多个统计特征值, 并以统计数据对文献[11]提出的节点间非线性连接方法进行优化, 最后提出 DNLH(Dynamic Nonr Linear Hierarchy) 模型.

3.1 动态性增长及非线性优先连接

Internet 路由级拓扑是动态增长的, 故使节点度按照幂律方式增长的动态无尺度网络模型即是扑模型的基本核心. 传统的 BA 模型^[10] 生长过程中的连边概率, 即一个新节点与一个已经存在的节点 i 相连接的概率与节点 i 的度 k_i 、节点 j 的度 k_j 之间的关系为线性关系, 满足式(1):

$$\Pi_i = k_i \setminus \sum_j k_j \quad (1)$$

将其修正为式(2):

$$\Pi_i = k_i^{1+\varepsilon} \setminus \sum_j k_j^{1+\varepsilon} \quad (2)$$

即为非线性连接. 其中 ε 为待确定参数, 据文献[11]所述, 当 ε 取值范围在 $[0.1, 0.3]$ 时, 其模型生成拓扑幂律指数分布区间为 $[2.1, 2.5]$, 覆盖 Internet 路由级节点度分布的幂指数.

3.2 度分布层次性

由第 2 节的幂律分析的结果可知, 节点度分布呈现一定的层次性, 故分段对其分布曲线进行拟合. 由于 degree-rank 幂律分布所得结果比较直观, 易于参与节点度值调整, 所以选择其作为分析目标. 图 4 给出对 top 节点部分与主体节点部分 degree-rank 幂律曲线的二段拟合结果, 其中叶子节点度值单一, 不对其做拟合分析.

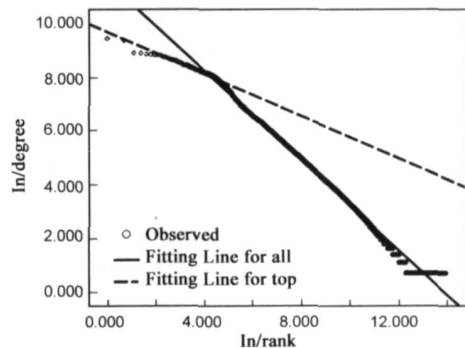


图 4 Internet路由级节点度的 degree-rank 幂律分布二段拟合曲线

图 4 中 top 节点部分, 近似符合幂指数为 0.393 的 degree-rank 幂律分布. 幂律分布函数为:

$$y = 12386.97 \cdot x^{-0.393} \quad (3)$$

并且 top 节点在节点整体中所占比例(N_{top}/N)为 3.54%.

3.3 建模算法

DNLH 模型所需参数如表 3 所示. 其中新增非叶子节点的连接数为 m , 考虑到真实拓扑图中的节点平均度为 4.8 左右, 由表 1 得知 $\bar{d} = 2E/N$, 故每个新增节点平均增加 $\bar{d}/2$ 条边, 其值约为 2.4. 由于叶子节点的存在, 为保证整体节点平均值稳定而向上取整, 因此参数 m 的取值为整数 3. 另外参数 T 的存在目的是避免算法

调节过于频繁, 其值对算法实质并无影响, 故直接指定, 不作讨论.

文献[12]指出, 理想状态下 $R_{cd} = R_{fd} - 1$. 为使建模结果更加精确, 算法中节点度的幂指数 R_{fd} 确定为 2.3 节所计算的 R_{cd} 值加 1, 即 2.256.

模型以式(2)为基本公式, 其中节点动态增长, 节点度分布由 ε 控制, 保持幂指数稳定在 R . 增长过程

中, 对非叶子节点按度值降序排列, 将前 w_{top} 的节点划分为 top 节点, 依照式(3)对其进行度值 degree rank 幂律调整. 调整后根据 \bar{d} , 计算 E 的变化量, 在主体节点部分进行边数的修正, 以保证度平均值在 degree rank 幂律调整前后保持一致. 整个过程中, 无论增加节点、增边、减边, 已有叶子节点均保持不变, 以维持 w_{leaf} 稳定. DNLH 模型生成算法详细表述如表 4 所示.

表 3 DNLH 模型主要参数定义表

| 参数 | 取值 | 定义 |
|---------------|-------------------|--|
| N | 动态增加 | 图中节点总数, 即拓扑的整体规模 |
| \bar{d} | 4.868 (2.3 节统计值) | 图中节点平均度 |
| m | 3 (由平均度分析得出) | 每新增非叶子节点的连接数, 用于稳定节点平均度 |
| w_{leaf} | 17.79% (2.3 节统计值) | 图中叶子节点比重, 控制节点度层次分布中叶子层比例 |
| R | 2.256 (2.3 节统计值) | 主体节点 frequency degree 幂指数 R_{fd} , 控制拓扑整体节点度分布取向 |
| T | 0.1 (指定) | 幂律调整阈值, 即节点数每增加 T , 才对网络拓扑做 degree rank 幂律调整 |
| R' | 0.393 (需优化) | 式(3)参数, top 节点 degree rank 幂指数 R_{t} , 控制拓扑 top 层节点度分布取向 |
| ε | [0.1, 0.3] (需优化) | 式(2)参数, 用于调整幂指数 R_{fd} |
| w_{top} | 3.542% (需优化) | 图中 top 节点比重, 控制节点度层次分布中 top 层比例 |

表 4 DNLH 模型生成算法

| 步骤 | 操作内容 |
|-----|--|
| (1) | 新增一个节点, 按概率 w_{leaf} 判断其是否为叶子节点, 对其执行步骤(2)(3) |
| (2) | 对每个新节点, 将其连接到网络中已经存在的非叶子节点上, 新增非叶子节点与 m 个节点相连; 一个新节点 j 与一个已经存在的节点 i 相连接的概率 Π_{ij} 与节点 i 的度 k_i 、节点 j 的度 k_j 之间满足式(2) |
| (3) | 判断节点数增长率, 如果在上一次 degree rank 幂律调整后, 增长率超过 T , 则执行步骤(4), 完毕后返回步骤(1) |
| (4) | 对图中所有节点的度按降序排列, 按照式(3)计算度值最大的 w_{top} 节点(top 节点)度的理论值. 对目前所有 top 节点执行步骤(5), 完毕后执行步骤(6) |
| (5) | 如果节点 i 当前度值大于度的理论值, 则执行减边操作. 根据式 2 从主体节点中选择节点 i 的邻居节点并去除二者之间的连接, 直至节点 i 度值减为理论值; 如果节点 i 当前度值小于度的理论值, 则执行增边操作. 根据式 2 从主体节点中选择非邻居节点 j 并增加二者之间的连接, 直至节点 i 度值等于理论值; 注意, 以上增、减边操作均不涉及已有叶子节点 |
| (6) | 统计现有边数, 参照 \bar{d} , 计算出所需调节的边数 m' , 按式(2)在主体节点部分选择节点对, 进行去除连接或增加连接的操作. 重复执行 m' 次, 将 m' 条边重新加入或从网络拓扑中减掉 |

3.4 参数确定遗传算法

3.3 节给出了 DNLH 模型的建模算法流程, 其中有部分主要参数取值需要进一步确定, 包括 ε 、 R' 与 w_{top} . 参数 ε 决定生成拓扑的 R_{fd} 整体趋势, 但二者之间的数学关系并不明确, 且由算法流程可知, 建模过程中对主体节点部分有两次边数调整, 将导致 R_{fd} 有所变化, 进而影响值的确定. 参数 R' 和 w_{top} 是由图 4 所示拟合直线得出, 但其节点数量少, 拟合优度并不如主体节点部分, 并且这二者的取值将影响到幂律调整动作的大小. 为使算法生成拓扑更加符合真实情况, 对以上三个参数做进一步优化, 给出更为合理的取值.

考虑到参数选择对拓扑影响的复杂性, 本文采用浮点型遗传算法^[15, 16]对模型参数进行优化, 参数 ε 、 R'

与 w_{top} 为遗传基因. 为防止遗传基因在杂交或变异时发生突变, 影响遗传算法收敛, 本文设定参数 R' 与 w_{top} 的限制区间大小为正负 10%, 参数 w_{top} 区间设定方法与此相同, ε 则据 3.1 节分析限定在区间 [0.1, 0.3] 之内. 考虑到主体节点为拓扑中最主要也是数量最大的部分, 其分布最为稳定, 本文选择第 2.3 节所得的 Internet 路由级拓扑的节点度分布幂指数, 即拓扑主体所符合的幂律分布为评价指标. 节点总数设为 $N = 10000$. 根据上述分析, 确定本文遗传算法如下所示.

(1) 遗传基因(个体)编码

算法中的遗传基因由待优化的参数组成的向量表示, 其编码结果由式 4 表示:

$$x = [\varepsilon, R', w_{top}] \quad (4)$$

(2) 适应度(评估)函数

待定参数的确定应使 DNLH 模型的幂律分布与 Internet 拓扑的实际情况拟合最好, 即两组拓扑幂指数的差值最小. 因此建立适应度函数为:

$$f(x) = \sum_{R=R_{fd}, R_{dr}, R_{cd}} |R - R_0| \quad (5)$$

其中 R_0 为对应幂指数的统计值, 具体值参见 2.3 节.

(3) 生成初始群体

设群体规模为 L (本文 L 取值为 30), 包含 L 个遗传基因, 即要生成 L 个 10000 点的 DNLH 模型生成拓图, 并根据其主体节点部分 R_{fd} 、 R_{dr} 与 R_{cd} 值, 计算适应度函数值.

(4) 杂交

对杂交池 MP 中 L 个个体, 随机选择个体 x_i , 与最优个体 x_j 杂交. 以参数 R' 为例, 杂交操作为:

$$\begin{aligned} R'_i{}^{new} &= R'_i(1 - \alpha) + \beta R'_j \\ R'_j{}^{new} &= R'_j(1 - \alpha) + \beta R'_i \end{aligned} \quad (6)$$

其中 $\alpha(0 < \alpha < 1)$, $\beta(0 < \beta < 1)$ 为随机数, $R'_i{}^{new}$ 与 $R'_j{}^{new}$ 为杂交操作后的新参数, 进入下一轮遗传计算. 杂交操作之后, 遗传算法对遗传个体的参数进行编码修正, 防止参数超出限定区间.

(5) 变异

对长度为 $L \times l$ (其中 L 为个体数目, l 为单个遗传基因的编码长度) 的变异池, 对每个参数以 0.05 的概率 (考虑待定参数数量而选择) 进行变异操作. 以参数 R' 为例, 其变异操作为

$$\begin{aligned} R'_i{}^{new} &= R'_i(1 + \alpha) & \text{if } \gamma \geq 0.5 \\ R'_i{}^{new} &= R'_i(1 - \alpha) & \text{if } \gamma < 0.5 \end{aligned} \quad (7)$$

其中 $\alpha(0 < \alpha < 1)$, $\gamma(0 < \gamma < 1)$ 为随机数. 变异操作后对个体再次进行参数修正操作, $R'_i{}^{new}$ 为变异操作后的新参数.

(6) 选择

杂交与变异操作之后, 根据产生的新参数, 重新生成 L 个遗传基因. 为减小遗传计算的复杂程度, 并保留最优个体, 选择当前种群中的最优个体, 与上一代最优个体比较, 选择适应度高的保存. 若上一代最优个体被选中, 则之以替代当前种群中的最差个体, 群体总数保持不变.

(7) 终止条件

对新群体中的所有个体, 循环进行适应度函数计算、变异、杂交与选择等操作, 直到种群中的最优个体的适应度函数在最近 K 轮遗传操作中没有增长. 考虑 DNLH 模型生成算法耗时较大, 本文 K 取值为 500.

(8) 优化结果

通过遗传计算, 算法对遗传基因优化结果如表 5 所示, DNLH 模型生成算法将使用优化后参数创建生成拓扑图.

表 5 遗传算法优化结果

| 符号变量 | 原值 | 优化值 |
|---------------|------------|---------|
| ε | [0.1, 0.3] | 0.106 |
| R' | 0.393 | 0.314 |
| w_{top} | 0.03542 | 0.03931 |

4 实验分析

DNLH 模型是面向 Internet 路由级拓扑的模型, 而现有模型均为面向 AS 级拓扑, 故彼此之间无法进行量化分析. 从方法上来说, DNLH 模型综合考虑了拓扑增长的动态性、节点间非线性连接特性, 以及节点度分布的层次性, 并以实际海量数据优化了关键参数, 克服了已有模型的不足, 其合理性是显然的. 下面主要验证其有效性.

由 3.3 节给出的 DNLH 模型的建模算法流程可知, 生成拓扑的边数严格按照节点平均度的实际统计值与当前拓扑中的节点数控制, 故生成拓扑的节点平均度必然与实际拓扑一致, 不需验证. 类似的, 叶子节点所占比例也无需验证. 又由于算法是依据测量值的 degree rank 幂律分布进行的节点度调整, 故生成结果必然与实际情况高度一致, 因此也不必对其做 degree rank 幂律分析.

考虑到 CCDF(d)-degree 幂律分布对数据偏差鲁棒性较好的特点, 对生成拓扑进行 CCDF(d)-degree 幂律分布分析, 以考察建模算法与测量值吻合的精确程度. 依表 4 所述算法生成两组节点个数为 10000 的拓扑图与实际拓扑作对比, 算法中非优化参数按表 3 所述选取, 优化参数按表 5 所述选取.

文献[17]指出, 即使具有相同幂律分布的无标度图也可以有完全不同的结构. 因此, 仅依靠节点度的幂律分布来判定生成拓扑与真实拓扑吻合还不够, 需要与其他度量结合起来. SLS(拉普拉斯谱)是目前区分拓扑特征量化分析最好的一种手段^[18], 因此本文使用 SLS 谱分析来评价 DNLH 生成拓扑的有效性. 限于计算能力, 本文对实际 Internet 拓扑进行采样, 选择 4000 个连通节点组成的子拓扑进行对比分析(与文献[18]所用实例规模相近). 由文献[18]可知, 子拓扑的拉普拉斯谱可以代表整个拓扑的特征. 与前述类似, 依表 4 所述算法生成两组节点个数为 4000 的拓扑图与实际拓扑作对比, 算法中非优化参数按表 3 所述选取, 优化参数按表 5 所述选取.

4.1 CCDF(d)-degree 幂律分布分析

DNLH 模型生成拓扑图的 CCDF(d)-degree 幂律分

布及拟合结果如图 5 所示, 其中横坐标为节点度值 d , 纵坐标为度数大于 d 的节点累积数 CCDF, 二者均取自然对数值. 拟合幂指数 R_{cd} 分别为 1.128(图 5(a)) 和 1.122(图 5(b)).

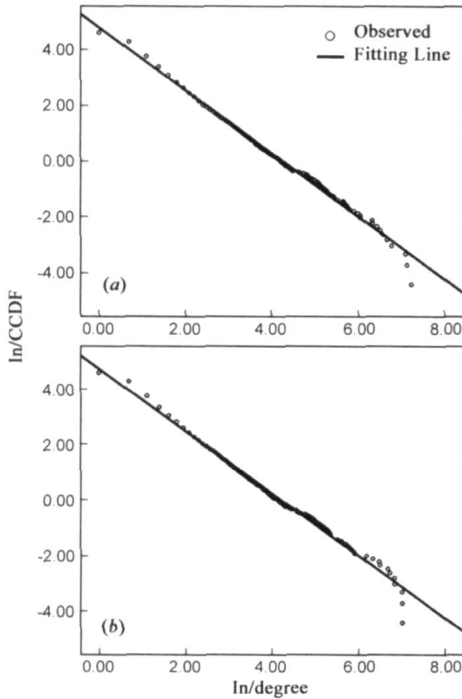


图 5 DNLH 模型生成拓扑图 CCDF(d)-degree 幂律分布分析及拟合结果

图 5 所示两幅拓扑的直观表现, 与图 2 给出的 Internet 实际路由级拓扑 CCDF(d)-degree 幂律分布曲线十分相似, 无差异明显的噪声点. 两组拓扑图均表现出明显的幂律分布特征, 且图中点分布情况与图 2 相似, 即与拟合直线相比, top 节点部分有明显的度值偏小趋势. 同时, 两幅生成拓扑幂律分布表现十分相近, 说明 DNLH 模型生成算法收敛情况较好, 生成结果稳定可靠.

生成拓扑的拟合幂指数 R_{cd} 为 1.128 和 1.122, 接近 2.3 节得出的实际统计结果 1.256, 说明本文对指数的优化很好地完成幂律指数调节的任务.

4.2 SLS 谱分析

一个图 G 的无符号拉普拉斯矩阵 $|L|$ 定义为 $|L| = D + A$, 其中 D 是描述 G 的节点度的对角矩阵, A 是 G 的邻接矩阵. 无符号拉普拉斯谱(SLS)就是 $|L|$ 的特征值集合.

对特征值降序排列, 生成 SIS 谱分布图. 实际拓扑图与 DNLH 生成拓扑图所做的 SLS 谱分析结果如图 6 所示, 横坐标为特征值序号, 纵坐标为特征值. 图 6(a) 给出 Internet 实际采样拓扑的 SLS 谱分布结果, 图 6(b) 与图 6(c) 给出生成拓扑的 SLS 谱分布结果, 三幅图均采用半对数坐标(纵轴为对数坐标).

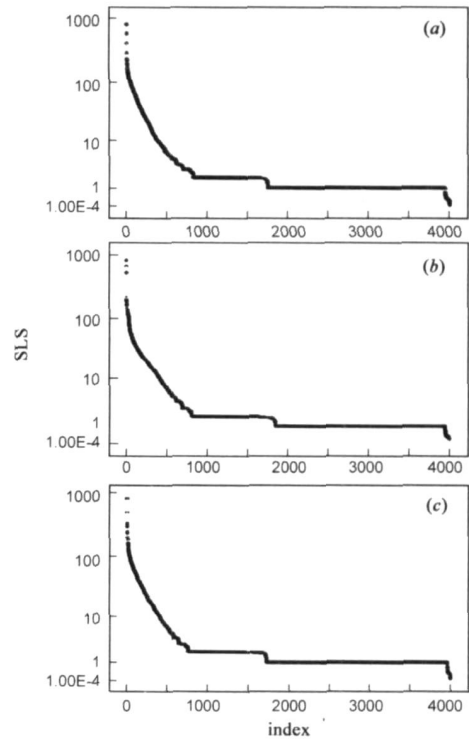


图 6 DNLH 生成拓扑与 Internet 测量拓扑的 SLS 谱分析图

从图中我们可以观察到生成拓扑均与 Internet 测量拓扑表现出较高的相似性, 其下降趋势基本一致. 三组拓扑均在特征值 $\lambda=1$ 处重数较高, 重数次高的特征值均群聚在 $\lambda=2$ 处. 在特征值从最大值下降到 2 的变化过程中表现出较明显的幂律分布特性, 在特征值小于 1 的数据中同样可以看到类似幂律的分布特征. 说明 DNLH 模型生成拓扑结构均较好地重现了 Internet 拓扑特征规律, 拓扑生成结果可接受. 另外, 与 4.1 节相同, 两组生成拓扑的 SLS 谱分布情况极其相似, 进一步说明了 DNLH 模型生成算法的稳定性较好.

5 总结与展望

本文分析了 CAIDA 测量 Internet 路由级拓扑数据, 统计共 21 个监测节点的测量数据, 分析了 Internet 路由级节点度分布的三种幂律规律, 并给出主要特征度量的计算结果. 在此基础上, 分析现有 AS 级拓扑模型的不足, 提出了 DNLH 模型, 其主要优点在于:

- 可以反映 Internet 拓扑的动态增长;
- 节点间以非线性优先方式连接;
- 体现节点度分布的层次性;
- 以海量实际测量数据为背景, 优化了模型参数, 对真实拓扑符合较好.

DNLH 模型虽然比较接近实际 Internet 路由级拓扑, 但其参数确定算法比较复杂, 收敛情况也未达到最佳, 且对统计值的依赖较强, 对未来 Internet(如 IPv6 网络)的

适应性有待验证. 算法整体复杂度较高, 生成超大规模网络所需计算代价较大. 因此, 以后进一步的工作应包括对参数快速确定算法的研究, 以及整体建模算法时间效率的提高, 使之更适用于快速发展的 Internet.

参考文献:

- [1] 张宇, 张宏莉, 方滨兴. Internet 拓扑建模综述[J]. 软件学报, 2004, 15(8): 1220– 1226.
Zhang Yu, Zhang Hongli, Fang Binxing. A survey on internet topology modeling[J]. Journal of Software, 2004, 15(8): 1220– 1226. (in Chinese)
- [2] T Bu, D Towsley. On distinguishing between Internet power law topology generators[A]. In Proc. of the IEEE INFOCOM 2002, Vol 2[C]. New York: IEEE, 2002. 638– 647.
- [3] M B Doar. A better model for generating test networks[A]. In Proc. of IEEE Global Internet[C], London: IEEE, 1996. 86– 93.
- [4] Calvert K, Doar M, Zegura E. Modeling internet topology[J]. IEEE Communication Magazine, 1997, 35(6): 160– 163.
- [5] S T Park, D M Pennock, C L Giles. Comparing static and dynamic measurements and models of the internet's topology[A]. In Proc. of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies[C]. 2004, 3: 1616– 1627.
- [6] M Faloutsos, P Faloutsos, C Faloutsos. On power law relationships of the internet topology[J]. ACM SIGCOMM Computer Communication Review, 1999, 29(4): 251– 262.
- [7] J Winick, S Jamin. Inet 3. 0: internet topology generator[R]. Technical Report, CSE TR-456 02, Ann Arbor: University of Michigan, 2002.
- [8] R Albert, A L Barabási. Topology of evolving networks: local events and universality[J]. Physical Review Letters, 2000, 85(24): 5234.
- [9] A Medina, A Lakhina, I Matta, J Buers. BRIT: an approach to universal topology generation[A]. In Proc. of the MASCOTS 2001[C], Washington, 2001. 346– 353.
- [10] A L Barabási, R Albert. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509– 512.
- [11] S Bar, M Gonen, A Wool. An incremental super linear preferential internet topology model[A]. In Proc. of the 5th Annual Passive and Active Measurement Workshop[C]. LNCS 3015, 2004. 53– 62.
- [12] 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.
- [13] S Jaiswal, A L Rosenberg, D Towsley. Comparing the structure of power law graphs and the internet AS graph[A]. In Proc. of the 12th IEEE International Conference on Network Protocols (ICNP' 04)[C]. 2004. 294– 303.
- [14] P Mahadevan, etc. The Internet AS Level Topology: Three Data Sources and One Definitive Metric[A]. In Proc. of the ACM SIGCOMM Computer Communications Review (CCR), Vol 36(1)[C]. 2006. 17– 26.
- [15] G Rudolph. Convergence properties of canonical genetic algorithms[J]. IEEE Trans. on Neural Networks, 1994, 5(1): 96– 101.
- [16] 孟伟, 韩学东, 洪炳. 蜜蜂进化型遗传算法[J]. 电子学报, 2006, 34(7): 1294– 1300.
Meng Wei, Han Xuedong, Hong Bingrong. Bee evolutionary genetic algorithm[J]. Acta Electronica Sinica, 2006, 34(7): 1294– 1300. (in Chinese)
- [17] L Li, D Alderson, W Willinger, J Doyle. A first principles approach to understanding the internet's router level topology[J]. ACM SIGCOMM Computer Communications Review, 2004, 34(4): 3– 14.
- [18] 姜誉, 方滨兴, 胡铭曾, 何仁清. 大型 ISP 网络拓扑多点测量及其特征分析实例[J]. 软件学报, 2005, 16(5): 846– 856.
Jiang Yu, Fang Binxing, Hu Mingzeng, He Renqing. An example of analyzing the characteristics of a large scale ISP topology measured from multiple vantage points[J]. Journal of Software, 2005, 16(5): 846– 856. (in Chinese)

作者简介:



张 昕 男, 1979 年出生于内蒙古赤峰, 东北大学信息科学与工程学院博士生. 主要研究方向为复杂网络与信息融合.

E-mail: zhangxin1979@hotmail.com



赵 海 男, 1959 年出生于辽宁沈阳, 东北大学信息科学与工程学院教授, 博士生导师. 主要研究方向为复杂网络、传感器网络、数据融合.



李 超 男, 1982 年出生于湖北洪湖, 东北大学信息科学与工程学院博士生. 主要研究方向为复杂网络.