

路由级 Internet 宏拓扑结构的谱密度分析^{*}

徐野¹ 赵海² 张文波¹

(沈阳理工大学信息科学与工程学院 沈阳 110168)¹ (东北大学信息科学与工程学院 沈阳 110004)²

摘要 针对路由级 Internet 宏拓扑结构进行了谱密度分布分析与无符号拉普拉斯谱(SLS)分布分析。首先通过对拓扑结构各异的 5 种采样拓扑图,分别进行谱密度-特征值分布分析,发现 5 组分析结果表现出高度的一致性,证明了 Internet 拓扑结构的自相似性,也证明了采样拓扑可以再现局部 Internet 拓扑结构特征。然后通过采样拓扑图的谱密度-特征值分布函数与 ER 图、BA 无尺度网络拓扑图进行比较,发现三者具有明显区别,得出谱密度-特征值分布函数可作为分辨图谱拓扑结构有效方法的结论。最后通过 4 组 3000 点采样拓扑进行 SLS 分布分析,发现尽管 4 组 3000 点采样路由与连接互不相同,但 SLS 谱分布却非常相似,4 组采样拓扑在特征值 $\lambda=1$ 处重数均较高,重数次高的特征值都群聚在 $\lambda=2$ 处。在特征值从 2~103 变化过程中表现出较明显的幂律分布特性,其幂指数值保持在 3.2813 至 3.8013 之间,特征指数接近。该结论为 Internet 宏观拓扑结构建模研究提供了量化判据,是 Internet 拓扑建模研究的重要内容。

关键词 Internet 拓扑建模,谱密度,谱密度-特征值分布,路由级 Internet 拓扑,无符号拉普拉斯谱分布

Spectrum Density Analysis on Router-level Internet Macroscopic Topology

XU Ye¹ ZHAO Hai² ZHANG Wen-bo¹

(College of Information Science and Engineering, Shenyang Ligong University, Shenyang 110168, China)¹

(School of Information Science and Engineering, Northeastern University, Shenyang 110004, China)²

Abstract Analysis of spectrum density and SLS(signless Laplacian spectra) on router-level macroscopic topology was performed. Firstly, we found in experiments that five analysis results of spectrum density on five sampling topologies showed highly similarity proving that Internet is a system of self-similarity as well as the ability for sampling topology to resemble the whole Internet topology. Secondly, obvious difference between the spectrum density analysis results of sampling topology, ER graph and BA graph indicated that spectrum density analysis is a good way to distinguish graphs. Finally, analysis of SLS on four groups of sampling topology with 3000 nodes showed that SLS distribution results were very much similar with each other though the topology samples were quite different. All four analysis results had high tuples at $\lambda=1$ and second high tuples at $\lambda=2$. Besides, principle of the power law distribution was observed from SLS analysis when eigenvalue in SLS ranged from 2 to 103, proving in another aspect that Internet topology had property of self-similarity. The research results above could be regarded as quantitative judgements during the research of modeling on Internet macroscopic topology.

Keywords Internet topology modeling, Spectrum density, Spectrum-eigenvalue distribution, Router-level Internet topology, SLS

1 引言

1.1 谱密度分析

一个无向图 G 可以用一个对称的邻接矩阵 A 来表示。图 G 中的节点 i 和 j 之间有边关联,则 $A_{ij}=A_{ji}=1$, 否则 $A_{ij}=0, A_{ji}=0$ 。一个图 G 的特征值就是指其邻接矩阵的特征值,也是特征多项式的根 $\lambda_1, \lambda_2, \dots, \lambda_n$ 。代数图论理论研究发现,图的特征值与图的许多基本拓朴性质密切相关。因此,考察图的特征值分布特征对于了解图的拓朴性质具有重要意义。

一个图 G 的谱(spectrum)就是该图邻接矩阵 A 的特征值及它们重数的集合,可以将谱记为:

$$Spec(G) = \begin{pmatrix} \lambda_1 & \dots & \lambda_n \\ m_1 & \dots & m_n \end{pmatrix} \quad (1)$$

其中 m 为特征值的重数^[1,2]。

图的谱密度就是该图邻接矩阵 A 的特征值密度,一般记为 $\rho(\lambda)$ 。对于一个有限系统, $\rho(\lambda)$ 可以写成关于特征值的 δ 函数的和^[1-3]:

$$\rho(\lambda) = \frac{1}{N} \sum_{i=1}^n \delta(\lambda - \lambda_i) \quad (2)$$

其中, λ_i 是图 G 邻接矩阵 A 的第 i 个特征值, N 为特征值总数。当 $N \rightarrow \infty$ 时, $\rho(\lambda)$ 趋近于一个连续函数。

1.2 将谱密度分析用于路由级 Internet 宏拓扑结构

将路由级 Internet 宏拓扑结构抽象成一个由节点与边连接而成的图,那么可以将谱密度分析方法应用于其研究。至今为止, Internet 拓扑建模的研究大体上可按时间顺序分为三代^[4,5]: 第一代 20 世纪 80 年代的随机图(ER 图)产生器;

^{*} 基金项目: 国家高技术研究发展计划(2001AA415320)。徐野 副教授, 博士, 主要研究方向为数字通信、模式识别与非线性系统; 赵海 教授, 博士生导师, 主要研究方向为复杂网络、嵌入式技术。

第二代为 20 世纪 90 年代的结构产生器;第三代始于 1999 年 Faloutsos 等人发现 Internet 拓扑结构中存在幂律(power-law)^[6],也即类似 BA 模型生成的无尺度网络,从而产生基于网络节点度的拓扑模型与拓扑产生器。

上述三种类型的模型均可构建出结构复杂的大型网络,但针对建模结果的评测方法,大多只有定性描述,而在严格的定量分析与评价方法方面,目前常用的是谱密度(包括规格化拉普拉斯谱与无符号拉普拉斯谱)方法,这是因为谱密度反映的是图特征值的分布规律^[1,3],因此可以反映图的拓扑特征规律。在使用谱密度方法分析网络拓扑结构特征之前,先要验证该方法分辨拓扑图的能力。

1.3 谱密度对拓扑图的分辨能力

已有研究表明,ER 随机图的谱密度收敛于一个半圆,且半圆的底边部分呈指数分布^[7],其谱密度图谱如图 1 所示^[3]。

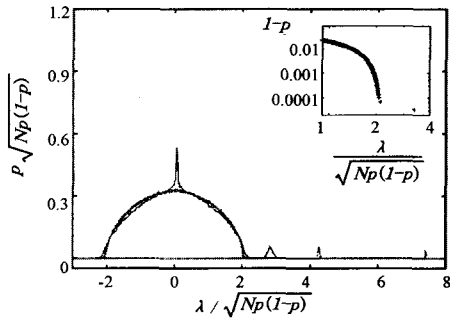


图 1 ER 随机图谱密度分析图

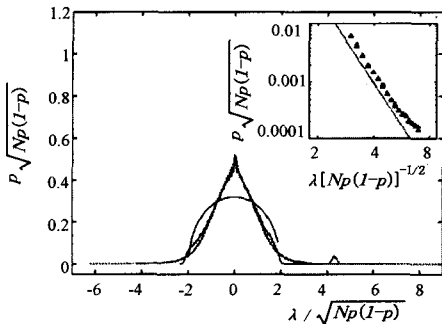


图 2 BA 无尺度网络拓扑图的谱密度分析图

部分研究^[3,8]表明,采用 Barabási-Albert(BA)模型建模过程^[9-11]生成的无尺度(scale-free)图的谱密度收敛于一个对称的连续函数,该函数的中部呈三角形,两侧边缘部分呈幂律分布,其谱密度图谱如图 2 所示^[3,8]。

由图 1 和图 2 可以明显看到,不同拓扑图的谱密度分布结果表现出明显的差别,由于谱密度分布表现的是拓扑图的最基本元素——特征值的分布情况,因此可以被视为分辨不同网络拓扑结构差异的有效手段。本文下面将对基于 CAIDA¹ 21 个 Internet 网络监测节点的 Internet 拓扑测量数据做谱密度分析,以揭示其拓扑结构的特征。

2 采样算法

Internet 网络规模宏大,如果将其抽象成拓扑图,在 21 个

监测节点测量结果下,其由路由器表示的拓扑节点数高达 1145841,由路由器连接表示的拓扑边高达 2907638。经同名 IP 解析后的拓扑节点数与边数分别为 29367 和 190280^[12]。

由于数据量巨大,直接计算上述拓扑图的谱密度是不可行的。因此要对 Internet 拓扑图进行点与边的采样,以从源数据中提取可进行谱密度计算的,且可以表征 Internet 拓扑结构特征的邻接矩阵。

为避免在采样过程中丢失 Internet 拓扑特征,采样算法的设计必须遵循以下几点算法思想与要求。

- 采样必须是随机的。采样的初始点必须通过随机方式选取,以便选取 Internet 拓扑任何可能区域的拓扑结果。
- 采样结果拓扑必须是连通图。
- 采样结果要尽可能覆盖较大范围网络。由于 Internet 拓扑中存在着许多度值非常大的节点,因此采样时如果将其所有直接连接的节点都采到,那么邻接矩阵中所表现出的网络拓扑,大多只是某一热节点及其相邻一跳距离的多个子节点,而不能很好地反映更广泛范围内网络的特征。因此在采样时,需要根据采样目标网络大小,严格控制某一节点直接连接节点的入队列数量。本文算法中对度值大的节点,按 $D=N \times \alpha$ 取值控制节点入队列,其中 N 为目标邻接矩阵长度, α 为控制参数,本文取 0.2。即凡是节点度值大于 D 的节点,从其直接连接节点中随机选取 D 个节点入队列,进入下一轮循环。

• 采到的点,其连接一定采到。采样算法的终止条件是队列中的总点数达到目标邻接矩阵大小的要求。这时对队列中剩余未处理的点,同样按照上述循环方式处理,但不再做节点入队列操作。此外,对与其连接但不在队列中的节点,不做连接记录,只记录那些与已经存在在队列中的节点之间的连接。这即保证了不会采到更多的节点,也保证了凡采到的点,其连接也都采到。

• 采样结果要生成邻接矩阵。由于邻接矩阵是对称矩阵,其上、下三角矩阵的内容完全相同。因此采样算法通过连接记录生成邻接矩阵的上三角后,将上三角内容拷贝到下三角部分,形成完整的邻接矩阵,进行输出。

根据上述算法思想与要求,本文通过 C 语言在 Windows NT 平台上实现了采样算法,算法初始值为一个 Internet 测量拓扑中的随机 IP 地址,然后以该 IP 地址为核心,逐跳地采样出相连的 IP 与连接,最后形成一个邻接矩阵并输出。

限于篇幅,本文不对采样算法做细致描述。其实现细节与上述算法思想与要求相一致。

3 采样拓扑的谱密度分布

3.1 不同点数的采样拓扑对 Internet 拓扑特征的代表能力

为验证采样拓扑对整体 Internet 拓扑结构的表征能力,本文先从 Internet 拓扑测量结果中采出三份拓扑样本——ip 数为 30、边数为 29,ip 数为 300、边数为 536 与 ip 数为 500、边数为 753 的三个拓扑图,计算得其特征值与谱密度如表 1 所示。

¹ CAIDA (The Cooperative Association for Internet Data Analysis),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共有 30 余家,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。亚洲仅有 3 家,其中两家在日本东京,另外一家在中国东北大学复杂网络研究中心。

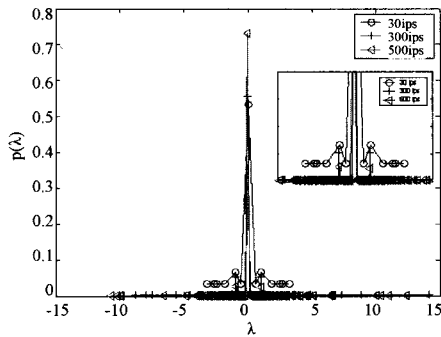
表1 采样拓扑图特征值与谱密度计算结果

30 ips		300 ips		500 ips	
$\lambda(13)_1$	$\rho(\lambda)$	$\lambda(104)_1$	$\rho(\lambda)$	$\lambda(112)_1$	$\rho(\lambda)$
-3.2196	0.0333	-8.7818	0.0033	-10.7058	0.0020
...
-1.0000	0.0667	-0.2887	0.0033	-0.3483	0.0020
-0.5663	0.0333	-0.1767	0.0033	-0.2635	0.0020
-0.0000	0.5333	-0.0000	0.5567	-0.0000	0.7320
0.5663	0.0333	0.1479	0.0033	0.1113	0.0020
1.0000	0.0667	0.2520	0.0033	0.1910	0.0020
...
3.2196	0.0333	14.1650	0.0033	12.3570	0.0020

注1:括号内数字为该拓扑图邻接矩阵的特征值个数。

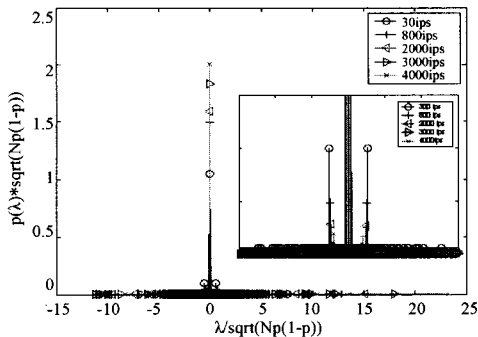
观察表1中谱密度与特征值,发现三组采样拓扑的特征值(分别为13个、104个与112个)都具有对称性质。这与文献[3,8]对ER图拓扑与BA无尺度网络拓扑的谱密度分析结果所表现出的对称性相一致。

根据表1谱密度计算结果,对 $\rho(\lambda)$ 与 λ 做图,如图3所示。从图中可以明显看到,尽管采样拓扑图中的节点数不同,但其谱密度分析结果却表现出惊人的一致性,即在 $\lambda = \pm 1.0000$ 处存在两个小峰,在 $\lambda = 0.0000$ 处谱密度取最大值,为全图最高峰值。特征值 λ 在其他处时 $\rho(\lambda)$ 值均小且平缓,无明显波动起伏。从数据上看,在非波峰处,多处 λ 对应的 $\rho(\lambda)$ 值是相等的。可见,对采样算法来说,尽管随机起始IP地址不同,选择IP数目也不同,但采样所得拓扑图的谱密度分布情况却表现出良好的一致性,从图中看到,300个IP拓扑图与500个IP图的差异几乎是微不足道的。这也部分地证明了采样算法对基于21个监测点的Internet拓扑测量结果的采样是客观有效的。



图中子图为将横坐标区间限定为[-5,5],纵坐标区间限定为[0,0.2]后的放大图,以便更清楚观察图形。

图3 3种采样拓扑图的谱密度-特征值分析图



图中右侧子图为将横坐标区间限定为[-3,3],纵坐标区间限定为[0,0.15]后的放大图,以便更清楚观察图形。

图4 5种采样拓扑图(点数分别为300,800,2000,3000与4000)的谱密度-特征值分析图

进一步研究谱密度与特征值变化关系,本文采用文献[1,3]尺度坐标轴的方法,对谱密度分析图的两个坐标轴都使用 $\sqrt{Np(1-p)}$ 重新标度。将y轴标度为 $\rho(\lambda) \sqrt{Np(1-p)}$,将x轴标度为 $\lambda / \sqrt{Np(1-p)}$,其中N是图中的节点个数, Np 是图的平均度。

此外,为进一步分析采样拓扑对Internet全局拓扑结构的刻画程度,在重新标度坐标轴的谱密度分析图中,本文将增加800个ip、2000个ip、3000个ip及4000个ip采样拓扑分析结果,修改后的谱密度分析结果如图4所示。由图4可以看到,尽管采样拓扑ip节点数目相差较大,但标度后的谱密度-特征值关系却表现出了明显的相似性,说明图谱的分布状态一致。5种采样拓扑均在 $\lambda = 0$ 处取 $\rho(\lambda)$ 极大值,观察子图,5种采样拓扑在 $\lambda = 0.5$ 左右取第二极值。根据采样算法从不同的随机IP地址开始,从21个CAIDA监测点测量的Internet拓扑图中采取点数、连接数各异的5种采样拓扑图,表现出上述的一致性,即说明了采样算法的有效性,既采样拓扑良好地再现了局部Internet拓扑结构特征,又证明了Internet拓扑结构的自相似性。

3.2 谱密度分析法对采样Internet拓扑及ER,BA拓扑图的分辨能力

此外,将图4与图1,2做比较(均做重新标度处理),发现采样Internet拓扑图的谱密度-特征值分布函数与ER图、BA无尺度网络拓扑有着明显的区别,而与文献[1]对中国地区教育网路由级拓扑研究的结论基本一致,由此可以得出结论:谱密度-特征值分布函数确实可作为分辨图谱拓扑结构的可行方法。

3.3 相同点数的采样拓扑对Internet拓扑特征的代表能力

由于Internet是一个动态网络,其拓扑结构不总是一成不变的。而即使是本文研究的一份Internet拓扑结构快照——全球21个CAIDA监测点在2006年01月30日测量Internet拓扑结果,在对测量拓扑做邻接矩阵时,初始点选择不同,其谱密度计算结果也不尽相同。尤其对采样算法来说,由于采样初始点为一个随机路由IP地址,如果IP地址分布位置不同(可能位于亚洲、欧洲或美洲某处),那么尽管采样IP地址点数相同,采样结果也必然不相同。

而采样点数控制着采样拓扑的范围,点数越多,所采的局部Internet拓扑范围越大。上文据图4已经分析到,不同初始IP地址、不同采样点数的五份采样拓扑表现出良好的一致性,说明不同采样结果均能够表征局部Internet拓扑特征,而基于Internet拓扑自相似性原理,可以推论出采样拓扑也能够表征Internet的拓扑结构特征。但不同采样拓扑在表征Internet拓扑特征的程度有所区别。我们知道,采样范围越大,采样结果越逼近真实网络,即采样算法对Internet拓扑采样点数越多,越能够更真实地表现Internet拓扑特征。推论至极端情况,采样算法将Internet拓扑中所有点数全部采到,那么得到的采样拓扑就是真实测量的全局Internet拓扑图。但限于计算能力,本文不可能将Internet中路由IP地址全部采样。考虑到图4结果表现出不同采样拓扑谱密度分布的一致性,本文将只对Internet拓扑测量结果进行部分采样,并将采样结果作为Internet拓扑实例进行研究。

由于谱密度-特征值可以有效地表征网络拓扑特征规律[1,3,13],而观察图3与图4可知,谱密度-特征值分布图中最显著的即为谱密度的两个极值点,虽然本文未对这两个极值点具体反映了拓扑图的哪些属性进行研究,但这两个极值点

无疑是分辨目标网络拓扑图的有力判据。因此考查图 4 中不同采样拓扑的两个谱密度极值,发现在第一极值点处,由 300 个 ip 采样拓扑图的 1.0460 开始,先以较大幅度(0.4488)增长至 800 个 ip 采样拓扑图,然后增长幅度减缓(分别为 0.092,0.2445)增长至 3000 个 ip 采样拓扑图的 1.8313,最后到 4000 个 ip 的采样拓扑图中极值增长为 0.17344,即拓扑图谱密度极大值由 3000 点图到 4000 点图增长到 2.0047。直观看来,谱密度极大值在采样拓扑 ip 数到达 2000 后,增长幅度逐渐变缓;第二极值随采样拓扑 ip 数增加而逐渐降低,从 300 点的 0.1002 开始,在 2000 点时取值 0.0269,在 3000 点与 4000 点时分别取值 0.0069 与 0.0193,虽有些振荡,但振荡幅度较小,可以判断增长幅度在 2000 点后逐渐变缓。

为进一步分析从 3000 点开始谱密度第一极值的增长、第二极值的降低幅度,本文分别再采取 3000 点,4000 点的拓扑各 3 份,并将其极值变化列在表 2 中。

表 2 3000,4000 点采样拓扑图谱密度极值

采样拓扑 ip 数	谱密度第一极值		谱密度第二极值	
	λ	$\rho(\lambda)$	λ	$\rho(\lambda)$
3000(组 1)	0	1.8313	± 0.4354	0.0069
3000(组 2)	0	1.5446	± 0.4918	0.0168
3000(组 3)	0	1.9167	± 0.4043	0.0148
3000(组 4)	0	1.8196	± 0.4453	0.0135
4000(组 1)	0	1.7928	± 0.4551	0.0049
4000(组 2)	0	2.0047	± 0.3887	0.0193
4000(组 3)	0	1.8594	± 0.4138	0.0133
4000(组 4)	0	1.9751	± 0.4004	0.0112
统计结果 ₁		6.8108%		-6.3462%

注 1:对 3000 点、4000 点的四组统计均值,然后用 4000 点谱密度均值与 3000 点均值求差后得到关于 4000 点谱密度均值的百分比。

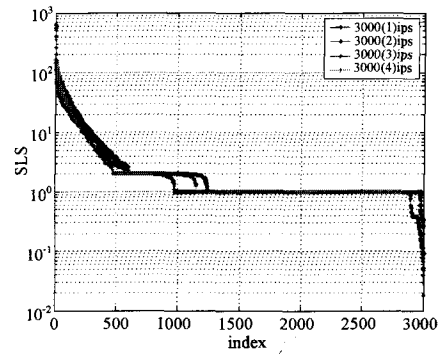
由表 2 可以看到,谱密度的第一极值与第二极值在 3000 点和 4000 点拓扑中相差不大,都在 6%左右,考虑到 Internet 拓扑的复杂性,可以认为 6%差别并不显著,是一个平缓的变化。因此可以推论出,3000 点的采样拓扑与 4000 点采样拓扑对 Internet 拓扑结构的表征结果相仿,考虑 Internet 拓扑结构的自相似性,3000 或 4000 点采样拓扑均可以作为 Internet 拓扑的采样结果进行研究。

考虑可计算能力,本文选择了 3000 点采样拓扑作为 Internet 测量拓扑实例数据样本,并以其为基础进行宏拓扑结构的建模研究。为更确定 3000 点采样拓扑对 Internet 测量拓扑特征的表征能力,本文下面将通过在区分相同规模拓扑结果效果更佳的无符号拉普拉斯谱(SLS)^[1,14]对 3000 点采样拓扑进行进一步的分析。

4 无符号拉普拉斯谱(SLS)

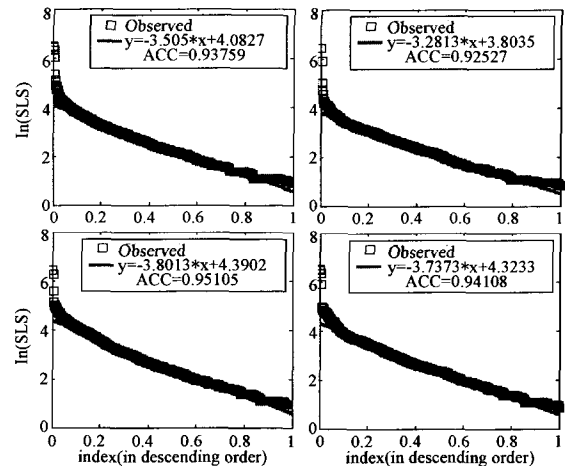
一个图 G 的无符号拉普拉斯矩阵 $|L|$ 定义为 $|L| = D + A$,其中 D 是描述 G 的节点度的对角矩阵, A 是 G 的邻接矩阵, A 与 D 的点序相同^[14]。无符号拉普拉斯谱(SLS)就是 $|L|$ 的特征值集合。部分代数图论理论的研究表明^[14],在 D 、 A 、单位矩阵 I 、全 1 矩阵 J 的各种线性组合(A 的系数非零)中,到目前为止,SLS 是区分规模相同、拓扑互异图效果最好的谱^[1,14]。SLS 对 4 组 3000 个 IP 地址采样拓扑分析结果如图 5 所示。

从图 5 可以看到,尽管 4 组 3000 点采样路由与连接互不相同,但 SLS 谱分布却非常相似,这再一次地证明了 Internet 拓扑结构的自相似性。



图中纵坐标是对数坐标,横坐标为拉普拉斯矩阵 $|L|$ 的特征值按降序排列的序列值

图 5 4 组 3000 个 IP 地址采样拓扑图的无符号拉普拉斯谱-特征值降序序列分析图



图中对纵坐标数据做了对数处理,横坐标为拉普拉斯矩阵 $|L|$ 的特征值按降序排列的序列值(按 $(i-1)/(N-1)$ 做归一化处理,其中 i 为序列号, N 为当前分析中点数)

图 6 4 组 3000 个 IP 地址采样拓扑图的 SLS 特征值(>2)与降序序列的幂律分析图

4 组采样拓扑在特征值 $\lambda=1$ 处重数均较高,重数次高的特征值都群聚在 $\lambda=2$ 处。在特征值从 $2 \sim 10^3$ 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。

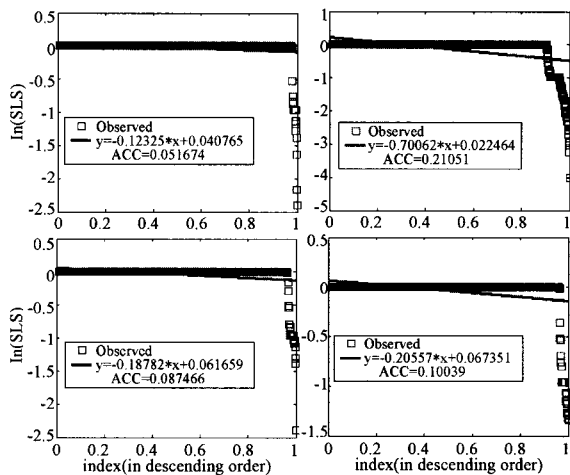
为进一步分析两部分可能存在的幂律分布情况,本文按 $SLS > 2$ 与 $SLS < 1$ 为判据,从 4 组 3000 点采样拓扑 SLS 分析结果抽取两份数据,分别进行幂律拟合分析。分析结果如图 6 与图 7 所示。

观察图 6,从图中可以明显看到,无符号拉普拉斯谱(SLS)特征值与其降序序列之间符合幂律分布,且拟合直线与观测数据拟合程度甚好,其相关系统数绝对值 ACC(absolute value of the correlation coefficient)都在 0.92 以上。

此外,观察 4 个子图所代表的 4 组 3000 点采样拓扑的拟合结果,发现尽管 4 组 3000 点采样拓扑表现的是完全不同区域的 Internet 拓扑结构,但在 SLS 谱密度分析结果(见图 5)中表现出高度的相似之后,在 SLS 特征值(>2)与特征值降序序列的幂律分析结果中同样非常相似。该结论也支持了 Internet 网络拓扑结构的自相似性。

4 个子图中幂指数值保持在一 3.2813 至一 3.8013 之间,特征指数很接近,其关于幂指数均值的标准差为 0.1146。因此可以推论出 SLS 特征值(>2)与特征值降序序列关系符合

幂律分布,这与文献[1]对中国教育网 ISP 拓扑结构的研究结论一致。



图中对纵坐标数据做了对数处理,横坐标为拉普拉斯矩阵 $|L|$ 的特征值按降序排列的序列值(按 $(i-1)/(N-1)$ 做归一化处理,其中 i 为序列号, N 为当前分析中点数)

图7 4组3000个IP地址采样拓扑图的SLS特征值(<1)与降序序列的幂律分析图

图7是图5中SLS <1 部分SLS与特征值降序序列关系的幂律拟合分析结果。观察图7,首先看到4个子图中相关系统数绝对值ACC均较低,最大值仅为0.2105,说明幂律拟合效果差。其次观察四子图中观测值(Observed)的分布情况,发现纵坐标数据集中分布在 $y=0$ 的一条直线上,在 $x>0.9$ 处存有少量数据。将上述两处数据分开处理,显然可以发现如下结论:对 $x<0.9$ 的数据,其幂律拟合结果将为一条水平线,即幂指数为0的一条直线。而由于该部分数据所占比重较大,因此可以推断出SLS特征值(<1)与特征值降序序列间不符合幂律分布;在 $x>0.9$ 处数据,可以看到其基本上呈现出一条非常陡直线的分布状态,如果对其进行幂律拟合,其幂指数将趋近于无穷大。

综上所述,SLS特征值(>2)与特征值降序序列关系符合幂律分布,而在 <1 处并不符合幂律分布规律,上述研究结论与文献[1]对中国教育网ISP拓扑的SLS分析结论是一致的。

5 分析结果

本节首先通过对比ER随机图、无尺度网络与Internet测量拓扑的谱密度分布情况,发现本文实际测量拓扑图不是ER随机图(节点度呈指数分布,谱密度呈半圆形,边界部分呈指数衰减),也不是符合BA模型的无标度图(谱密度中部呈三角形,边界部分呈幂律衰减)。而与文献[1]对中国地区教育网路由级拓扑研究的图谱基本一致,由此可以得出结论:谱密度-特征值分布函数可作为分辨图谱拓扑结构的可行方法。然后,本节对拓扑结构各异的5种采样拓扑图,分别进行谱密度-特征值分布分析,发现5组分析结果表现出高度的一致性,证明了采样拓扑可以有效地表征Internet拓扑特征,也证明了Internet拓扑结构的自相似性。

而对4组3000点采样拓扑进行SLS分布分析,发现尽管4组3000点采样路由与连接互不相同,但SLS谱分布却非常相似,4组采样拓扑在特征值 $\lambda=1$ 处重数均较高,重数较高的特征值都群聚在 $\lambda=2$ 处。在特征值从 $2\sim 10^3$ 变化过程中

表现出较明显的幂律分布特性,而在特征值小于1的数据中不具有幂律分布特征。同样从另一个角度证明了Internet拓扑结构的自相似性。

对SLS谱做进一步幂律分布分析,按SLS >2 与SLS <1 为判据,从4组3000点采样拓扑SLS分析结果抽取两份数据,分别进行幂律分布函数的拟合分析,发现无符号拉普拉斯谱(SLS >2)特征值与其降序序列之间符合幂律分布,且拟合直线与观测数据拟合程度甚好,其相关系统数绝对值ACC(absolute value of the correlation coefficient)都在0.92以上。这与文献[1]对中国教育网ISP拓扑结构的研究结论一致。

最后,由于采样拓扑的谱密度分布在 $\lambda=\pm 1$ 附近处有两个小峰,与BA模型的无标度图不同。这些结果说明,Internet路由级拓扑结构的发展可能遵循一种不同于文献[11]中所描述的BA模型的生长过程,也就是说Internet拓扑结构并非完全遵循BA模型(当前Internet拓扑建模中的常用模型)的生长过程。因此,谱密度分析方法不仅从定量角度分析了Internet拓扑结构特征,也是评价拓扑建模结果优良与否的重要方法,是Internet拓扑建模研究中的重要研究内容。本文研究与Chen等人^[15]、姜誉等人^[1]的研究从不同的角度得出了类似的结论。

参考文献

- [1] 姜誉,方滨兴,胡铭曾,等.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
- [2] West D B.图论导引[M].机械工业出版社,2006:1-47,339-348
- [3] Farkas I J,Derényi I,Barabási A,et al.Spectra of 'real-world' graphs:Beyond the semicircle law[J].Physical Review E,2001,64(2):1-12
- [4] 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
- [5] 汪小帆,李翔,陈关荣.复杂网络理论及其应用[M].清华大学出版社,2006:49-70
- [6] Faloutsos M,Faloutsos P,Faloutsos C.On power-law relationships of the Internet topology[J].ACM SIGCOMM Computer-Communication Review,1999,29(4):251-262
- [7] Mehta ML.Random Matrices.2nd ed[M].New York:Academic Press,1991
- [8] Goh KI,Kahang B,Kim D.Spectra and eigenvectors of scale-free networks[J].Physical Review E,2001,64(5):1-5
- [9] Barabasi A L,Eric Bonabeau.Scale free networks[J].Scientific American,2003:50-59
- [10] Barabási AL,Albert R.Emergence of scaling in random networks[J].Science,1999,286(5439):509-512
- [11] Albert R,Barabási A.Statistical mechanics of complex networks [J].Reviews of Modern Physics,2002,74(1):47-97
- [12] 徐野.Internet路由级宏拓扑结构的TL模型[D].沈阳东北大学,2006:36-38
- [13] Gkantsidis C,Mihail M,Zegura E.Spectral analysis of Internet topologies[C]//Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'03).IEEE Computer Society,New York,2003:364-374
- [14] Dam E,Haemers WH.Which graphs are determined by their spectrum[J].Linear Algebra and its Applications,2003,373:241-272
- [15] Chen Q,Chang H,et al.The origin of power-laws in Internet topologies revisited. Http://topology.eecs.umich.edu/data.html,2002