

# 基于核数分层的 AS 关系拓扑建模

郭 虹<sup>1,2</sup>, 兰巨龙<sup>1</sup>, 汪 涛<sup>2</sup>, 刘洛琨<sup>2</sup>

(1. 国家数字交换系统工程技术研究中心, 河南郑州 450002;

2. 信息工程大学信息工程学院通信工程系, 河南郑州 450002)

**摘 要:** 基于真实 AS 级互联网拓扑实测数据进行层次特性和 AS 关系特性分析, 提出了一种基于核数分层的 AS 关系拓扑动态演化模型. 建模和仿真评估表明该模型在整体上具与实际网络相似的拓扑宏特性, 还再现了 AS 商业关系的规律, 是一种能够逼真模拟互联网 AS 级拓扑宏观、微观特性的动态演化模型, 并且作为一种 AS 关系推测算法, 还能为任意的 AS 级无向图提供带 AS 关系标注的拓扑图, 拓展了模型的应用.

**关键词:** 自治域级拓扑建模; 带 AS 关系标注的拓扑图; 带标记的度分布; 层次性;  $k$ -core 分解

**中图分类号:** TP393.4      **文献标识码:** A      **文章编号:** 0372-2112 (2011) 11-2627-08

## Modeling AS Relationships Based on $k$ -core Decomposition

GUO Hong<sup>1,2</sup>, LAN Ju-long<sup>1</sup>, WANG Tao<sup>2</sup>, LIU Luo-kun<sup>2</sup>

(1. National Digital Switching System Engineering and Technological Research Center, Zhengzhou, Henan 450002, China;

2. Department of Communication Engineering, Institute of Information Engineering, Zhengzhou, Henan 450002, China)

**Abstract:** Based on the in-depth analysis of the massive real AS-level topology data, we abstracted the multi characteristics describing AS-level topology's hierarchy and AS relationships, and gived material modeling ways about network's hierarchical partition and business relationships between ASs. Then, we proposed a novel algorithm for generating synthetic graphs annotated with AS relationships named ASRTM based on the  $k$ -core decomposition. Finally, the numerical simulations and analysis results compared with the real AS-level topology data show that most of macroscopic statistics closed to real networks', including power-law degree distribution and AS relationships properties. Otherwise, ASRTM, as a conferring algorithm about AS relationships, can give a graph annotated with AS relationships for given undirected topology graph.

**Key words:** modeling on Internet AS-level topology; AS topology annotated with AS relationships; Annotation-degree distributions; network hierarchy; the  $k$ -core decomposition

## 1 引言

互联网作为世界的信息基础设施,是名副其实的复杂巨系统.研究互联网拓扑结构的科学意义和应用价值正受到学术界、工程界和军事部门的普遍重视<sup>[1]</sup>.互联网拓扑依不同粒度可分为:路由器(Router-Level, RL)级拓扑和自治系统(Autonomous System, AS)级拓扑. AS 级拓扑以自治域为节点,以域间的逻辑连接关系为边,各自治域由不同的运营商(ISP)运营管理.域间通过边界路由器运行 BGP4 协议实现彼此互连,域内运行不同的内部路由协议.基于安全性和商业利益的考量, BGP 采用基于策略的路由,允许各自治域灵活选择路由策略,各自治域间依据商业协定配置策略,导致 AS 间存在某种商业关系.

AS 关系的存在极大地影响了对实际互联网结构的解释,其中最重要的一点就是:连通性不一定意味着路由可达.对 AS 关系进行研究,准确掌握 AS 间的商业关系,建立能反映真实互联网商业关系特性的 AS 级拓扑模型对互联网的研究和应用具有重要意义.例如,建立逼真的互联网仿真环境、选取更优的提供商、设计具有流量均衡功能的路由协议,理解整个互联网的演化过程等<sup>[2]</sup>.

近年来,对互联网拓扑结构和模型的研究取得了长足发展.从幂律特性<sup>[3]</sup>的提出,到对互联网演化规律和内在机理的探讨<sup>[4]</sup>,期间涌现出大量有影响的 AS 级拓扑模型,如: BA、Tiers、Transit-stub、Inet、AB、BRITE、GLP、DP、PFP、TANG、GLRG、CMU、GRG、PA 等.不过,以往的模型,对 AS 关系建模的很少,人们多基于无向拓扑图来

研究网络,存在先天的不足和与生俱来的困难,尤其在研究 AS 最短合法路径问题<sup>[5]</sup>上。

由于 ISPs 将 AS 间关系视为商业机密,加之至今对 AS 关系推测研究成果<sup>[6~10]</sup>不多,获取并准确掌握 AS 间商业关系是非常困难的事.而对 AS 级拓扑结构和 AS 关系的研究则有助于域间路由系统基本问题的解决.鉴于目前对含 AS 关系拓扑分析和建模的不足<sup>[5,11]</sup>,本文基于 CAIDA<sup>[12]</sup>提供的含 AS 关系的互联网实测数据,进行了拓扑层次和 AS 关系特性分析,提出了一种基于核数分层的 AS 关系拓扑动态演化模型 (ASRTM),并通过计算机建模仿真和对比分析说明其合理性及有效性。

## 2 相关工作

### 2.1 AS 关系与带标注的拓扑图

AS 间商业关系在经济上就是付费问题,可分为三类: Provider-Customer、Peer-Peer 和 Sibling-Sibling<sup>[6]</sup>。

(1) Provider-Customer (简称  $p2c$ ): 若节点 A 到 B 间具  $p2c$  关系,则 A 是 B 的提供商, B 为 A 的客户, A 为 B 提供 Internet 接入服务, B 向 A 付费;同时, B 到 A 为 Customer-Provider 关系 (简称  $c2p$ )。真实网络中绝大多数 AS 间都是此关系,这是互联网商业模式的基础。

(2) Peer-Peer (简称  $p2p$ ): 若节点 A 和 B 间具  $p2p$  关系,则 A 和 B 多为规模相似的 ISP, A、B 无偿地访问对方的网络资源,不向对方收费。

(3) Sibling-Sibling (简称  $s2s$ ): 若节点 A 和 B 间具  $s2s$  关系,则 A 和 B 通过对方访问 Internet, A、B 间也不付费.通常  $s2s$  关系发生在同一个管理实体下的多个 AS 之间。

如果在拓扑图中用有向边标注出 AS 间的关系,则称之为带标注的拓扑图<sup>[6]</sup>。 $p2c$  和  $c2p$  用单向边表示,方向从 Customer 指向 Provider;  $p2p$  用无向边表示;  $s2s$  用双向边表示;箭头均指向能提供 Internet 接入服务的节点,如图 1 所示。

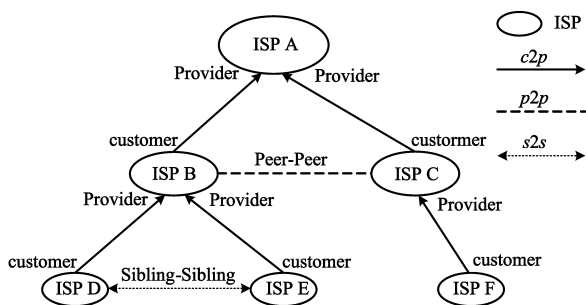


图1 带AS关系标注的拓扑图

### 2.2 AS 关系推测问题

AS 关系推测问题研究的目的在于提供带标注的拓扑图.由于大部分 ISP 视 AS 间关系为商业机密,其获取困难重重,只能依靠方法进行推测<sup>[6,9,13,14]</sup>。目前 AS 关

系推测方法主要有两类:完全依据 AS 路径的启发式推测算法和依据 Internet 路由注册机构的推测算法. Gao<sup>[6]</sup>的工作开创了 AS 关系推测研究的先河,她把合法 AS 路径的特征概括为 vally-free,并基于节点度,提出了一种根据 AS 路径集合推测 AS 间关系的启发式算法,其过程是:为输入的 AS 路径集合中的每一条 AS 连接依据合法 AS 路径的约束和上、下行路径分类,为其从  $c2p$ 、 $s2s$ 、 $p2p$ 、 $p2c$  关系中指定一种. Subramanian 等人<sup>[13]</sup>将 AS 关系推测问题抽象为 TOR (Type-Of-Relationship) 组合优化问题,即为一个从 BGP 路径集合  $P$  导出的无向图  $G(V, E)$  中的每一条边指定  $c2p$ 、 $p2p$ 、 $p2c$  关系中的一种,使得  $P$  中不合法的 AS 路径条数最小化;由于 TOR 是 NP 完全问题,他们提出一种启发式算法 SARK,该算法以各 AS 在图中的核数为度量进行 AS 关系推测.而 Battista<sup>[14]</sup>则认为在 TOR 定义下,无法准确对  $p2p$  关系进行推测. Dimitropoulos<sup>[11]</sup>研究也发现,在 TOR 框架下,  $p2p$  关系推测的准确度相对较低,并指出忽略  $s2s$  关系会导致更多的推测错误,提出了一种综合了 Gao 算法和 TOR 算法优点的多目标优化解决方法,尽可能使合法 AS 路径条数最多,同时使边尽量指向节点度高的节点。

### 2.3 AS 级网络拓扑特性

网络科学研究兴起于 20 世纪末,互联网结构研究继承了复杂网络领域的研究成果.目前,人们认识到 AS 级网络拓扑具有随机性、小世界性、层次性、幂律性、聚集性、异配性、富人俱乐部现象、商业关系、鲁棒和脆弱性以及自相似结构、向心性结构等特性<sup>[1,15]</sup>。这些特性是人们尝试从不同角度认识互联网的结果,各特性间具有内在联系.例如幂律性和向心性结构从不同角度体现了互联网拓扑的层次特性;而商业关系与层次性间又存在千丝万缕的联系等。

## 3 AS 关系拓扑图特性

### 3.1 拓扑数据的获取

CAIDA (Cooperative Association for Internet Data Analysis) 是对 Internet 结构及数据进行获取、分析和研究的国际合作机构.本文直接从 CAIDA 获取了 2004 年至 2010 年各季度首月份互联网 AS 关系拓扑数据,最新的是: as-rel.20100120.a0.01000.txt<sup>[13]</sup>。该 AS 拓扑图含 33508 个节点,15002 条有向边.这些数据源自 RouteViews 项目对 BGP 路由表信息的融合,其中 AS 关系部分采用的是 Dimitropoulos 等人的 AS 关系推测算法<sup>[10,11]</sup>结合 Internet 路由注册机构的公开信息,真实又权威。

### 3.2 AS 关系的拓扑特性

#### 3.2.1 各 AS 关系比例

对各实测数据分析  $c2p$ 、 $p2c$ 、 $p2p$ 、 $s2s$  关系的数量

及比例,发现 AS 级拓扑中占主导的是  $p2c$  和  $c2p$  关系,两者相加超过 90%; $p2p$  关系所占比例在 8% 左右; $s2s$  关系所占比例不足 0.5%. 进一步分析 AS 关系随时间的演化情况,发现尽管 AS 级网络不断演化,不同时间片各 AS 关系数量波动起伏较大,但由于商业关系的约束,相对网络边的总数,各商业关系的比例基本维持恒定.

需要注意的是,出于商业利益的考虑,AS 并不会将 Peer-Peer 连接通告给自己的 Provider 或其他 Peer,仅从外部采集 BGP 信息会系统地丢失掉部分 Peer-Peer 连接;同时偏好采样能较好地发现 Provider-Customer 和 Customer-Provider 连接;而采样点的缺乏则会导致 Sibling-Sibling 连接被误推测为 Provider-Customer 或 Customer-Provider 连接<sup>[10,16]</sup>. CAIDA 的数据说明中明确表明,可能存在误将低度 AS 节点相对高度 AS 节点推测为  $p2c$  关系.故,真实 AS 级互联网中  $p2c$  和  $c2p$  关系所占比例,应略低于表 1 中的数据; $p2p$  和  $s2s$  关系所占的比例,应略高于表 1 中的数据;不应忽略  $s2s$  关系.

### 3.2.2 带标记的度分布与补累积分布

度分布是表征拓扑结构最直观的度量之一,能反映网络的整体结构.文献[5]认为在 AS 级拓扑建模中,还应考虑带标记的度分布,用于反映 AS 关系的总体情况.

**定义 1(带标记的度)** 某 AS 节点的 Customer 邻居个数称为其的 Customer-degree,记作  $d_{p2c}$ ;同理,它的 Provider 邻居个数称为其的 Provider-degree,记作  $d_{c2p}$ ;它的 Peer 邻居个数称为其的 Peer-degree,记作  $d_{p2p}$ ;它的 Sibling 邻居个数称为其的 Sibling-degree,记作  $d_{s2s}$ .上述统称为带标记的度.显然,某 AS 节点的度值为 4 个带标记的度值之和.

**定义 2(带标记的度分布  $p(k)$ )** 表示网络中任何带标记的度值为  $k$  的概率.共有 4 种带标记的度分布.

图 2 给出 20100120 拓扑图各关系的带标记的度分布和补累积分布,缩小图为同一拓扑图的节点度分布和补累积分布,可看出:

(1)节点度分布是 4 个带标记的度分布之和,补累积分布同理;

(2)Customer-度分布在形态上与普通分布最为近似,服从幂律,说明在整个网络中,很多节点拥有 1~2 个 Customer 邻居,只有很少的节点才拥有较多的 Customer 邻居;

(3)Provider-度分布说明在 AS 级网络中,大多数低度节点( $k=1,2$ )拥有 1~2 个 Provider,网络中几乎不存在拥有 10 个以上 Provider 邻居的节点.这符合互联网的商业实际,只需选择有限的 ISP,即可接入 Internet;

(4)Peer-度分布说明大多数情形下,高度 AS 节点也仅为其它节点提供 1~2 条 Peer-Peer 连接.类似的,

Sibling-度分布说明,低度 AS 节点也仅为其它节点提供 1~2 条 Sibling 连接.

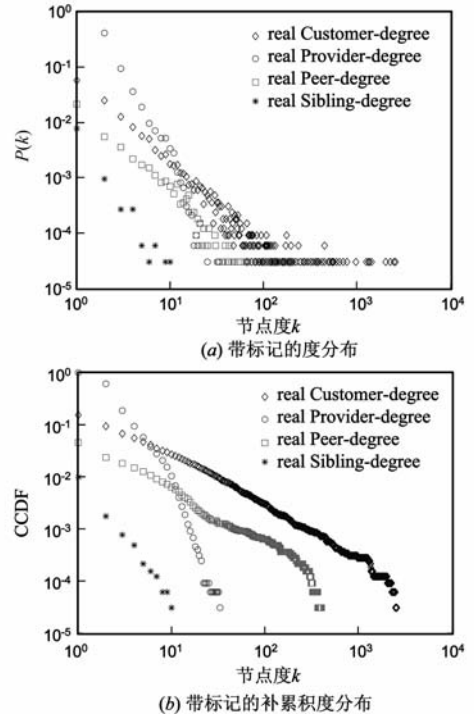


图 2 20100120-AS 级拓扑的带标记的度分布

### 3.2.3 带标记的度相关性

带标记的度分布并不能揭示出特定节点不同标记度间的关系.例如,大规模的 tier-1 AS 节点,其通常有大量的 Customer 邻居,即高的  $d_{p2c}$  值;几乎没有 Provider 邻居,即接近零的  $d_{c2p}$  值;有一点 Peer 邻居,即较小的  $d_{p2p}$  值.事实是,不同标记度间因商业关系的约束一定存在某种相关性.文献[5]提出一种联合的标记度分布(JADD)以描述带标记的度相关性:

$$P(d_{p2c}, d_{c2p}, d_{p2p}) = n(d_{p2c}, d_{c2p}, d_{p2p}) / N \quad (1)$$

其中  $n(d_{p2c}, d_{c2p}, d_{p2p})$  为拓扑中具有  $d_{p2c}, d_{c2p}, d_{p2p}$  度值的节点数,  $N$  为全网节点总数.显然, JADD 的边缘分布即为各带标记的度分布,含有不同标记的度相关性.但是, JADD 作为多维函数,不够直观,不易应用,还忽略了 Sibling-degree 的信息.为探究并直观展示 AS 级网络节点不同标记度间的相关性,本文统计了低度节点( $k=1,2,3$ )和高度节点(0.25%、1%)不同标记度的情形.发现:核心节点拥有超过 80% 的 Customer 邻居和 10% 的 Peer 邻居;低度节点其邻居 85% 以上是 Provider.所以不难得出结论:在互联网 AS 级拓扑中,不同标记度间存在相关性,与节点的属性和地位有关.

## 3.3 层次性与 AS 关系

### 3.3.1 层次性

互联网的规划建设和商业模式使其在组织架构上呈现出基于服务供应商的层次结构.文献[17]认为如果

网络中存在层次性,就会呈现明显的负的簇度相关性.对 AS 级实测数据作簇度分布,发现  $C(k)$  与  $k$  之间存在高度的负相关性,并随着  $k$  的增加而递减,递减斜率大致在  $-0.75$  左右.

为对 AS 级拓扑层次建模,文献[18]提出基于节点度值将 AS 级互联网大致划分为 4 个层次;文献[19]则以核数为度量,将 AS 级拓扑粗略分成 3 个部分:最深核,极大连通子图和孤立连通块;文献[4]发现了网络核心与边缘的不同演化规律,提出一种核心-边缘模型.

考虑到基于度的层次划分存在偏差和主观性,而核数(Coreness)是衡量节点中心度的参数之一,节点核数与其地位、规模正相关.所以核数能够较好地标识出节点在拓扑图中的中心度,较节点度具有更客观的层次刻画能力.本文基于对真实 AS 级拓扑数据的  $k$ -core 分解,提出一种基于核数划分 AS 级互联网层次的方法,如表 1 所示.

表 1 基于节点核数值划分的互联网层次

等级	$kc$	节点比例	所拥有连接的比例	层次属性	层次的物理意义
C1	$\geq 28$	0.25%	48.18%	网络核心层	国际或国家级主干
C2	10~27	1.92%	32.80%	网络传输层	大型的地区级 ISP
C3	4~9	8.04%	37.20%		小型的地区级 ISP
C4	3	10.29%	18.13%	终端客户层	小型的城域网、校园网、园区网
C5	2	42.15%	38.70%		
C6	1	37.35%	16.69%		

6 层次的划分依据是:对各实测数据进行  $k$ -core 分解后发现,不同尺寸的  $k$ -核( $kc$ )有着稳定的统计特性;最高核数  $kc_{\max}$  虽有波动,但相对于网络规模的变化  $kc_{\max}$  自 2004 年后逐渐趋于 28;最高核内所含节点数占全体总节点数的比例一直很稳定,在 0.142%~0.340% 之间;外部节点与最高核的相关连接占了全网总连接的很大一部分(41.42%~48.21%).充分说明,真实 AS 级网络中最高核很少突然出现或消失,一直很稳定,而且核心影响着全网的整体性能;各核层节点倾向于与最高核节点连接,最高核的影响力渗透到全网,非常重要.所以,与文献[19]类似,将所有最高核节点单独分层,对应网络核心层. $k$ -core 分解后,各核子图尺寸虽有波动,但相较全网规模各核节点比例几乎稳定,各核所拥有连接占全网总连接的比例也基本稳定,只在  $kc=1,2,3$  处波动大一点,说明网络中不稳定的部分主要是低核数节点,且此部分节点占总节点数的大部分(20100120:89.79%),而其余核层( $kc>3$ )节点数较少.虽然层次划分越细则拟合越好,但过拟合会带来模型复杂度的提升,兼顾模型的准确度和复杂度,将网络大致划分为三个大层: $kc\leq 3$  的节点对应终端客户层,由于各核层变化较大,分层刻画; $kc=4\sim 27$  合并处理,以 10 为界限划分成两个子层.所以,依核数高低将 AS 级拓扑划分为 6 个层次.

### 3.3.2 层次性与 AS 关系

为探究层次性与 AS 关系的内在联系,对各实测数据分别进行基于核数 6 层次划分的 AS 关系的统计分析,发现基于核数的层次划分使得 AS 关系呈现聚类性和逻辑性:

(1)网络核心层的节点拥有绝对的  $p2c$  边(84.13%)和部分  $p2p$  边(14.57%),基本不含  $s2s$  边(0.12%);

(2)终端客户层的节点拥有大量的  $c2p$  边(85.89%)和一点  $s2s$  边(0.43%),基本不含  $p2p$  边(0.29%);

(3)网络传输层作为终端客户层与核心层的过渡层,在与网络核心层的连接中还包含  $p2p$  边;在与终端客户层的连接中还包含  $s2s$  边;

(4)由低层到高层的连接以  $c2p$  边为主;反之,高层到低层的连接以  $p2c$  边为主;

(5) $p2p$  边主要出现在与第 1、2、3 相关的层次中; $s2s$  边主要出现在与第 4、5、6 相关的层次中.

上述结论与实际互联网的商业关系情况一致.所以,正是 AS 节点的层次归属决定了其间的商业关系;反之 AS 间的商业关系为互联网的层次性提供了新的佐证和内在解释.

## 4 基于核数分层的 AS 关系拓扑模型

基于实测数据 AS 关系的层次分析,本文提出了一种新的基于核数分层的 AS 关系拓扑动态演化模型 AS-RTM(Autonomous-System Relationship Topology Model).与文献[5]模型不同的是,前者通过对带标记的度分布以及 JADD 的拟合实现建模;本文所提模型则是在网络动态演化过程中,依据边缘节点的层次归属依概率确定边之间的商业关系.

### 4.1 基于核数的层次模型

层次模型的思想是:当新节点进入网络,依层次选择概率( $H_{cp}$ )选择相应的层次加入;再依各层间连接概率( $H_p$ )选择目标节点的层次;再依偏好择优概率从目标层次中选择相应的宿主节点进行连接,实现网络的动态演化增长.模型结构建立后,依据各实测数据层次统计分析结果确定模型参数(表 2).

表 2 ASRTM 层次模型参数

层次	$i, j$	$H_{cp_i}$ (%)	$\langle k \rangle$	$\bar{m}$	各层与其他层的连接概率 $H_{p_j}$ (%)					
					1层	2层	3层	4层	5层	6层
C1	1	0.253	436.704	419.041	4.35	15.15	21.81	13.33	35.66	9.71
C2	2	1.289	41.406	39.341	31.15	5.34	14.47	10.66	27.6	10.77
C3	3	7.522	11.505	10.511	28.43	8.91	9.46	11.38	28.66	13.16
C4	4	11.238	4.474	4.083	29.92	11.43	19.56	9.58	19.58	9.94
C5	5	45.137	2.194	2.043	39.75	14.78	24.61	9.76	7.41	3.69
C6	6	34.560	1.025	1.000	29.15	15.29	30.06	13.23	9.75	2.51

对实测数据各层节点比例从第 6 层到第 1 层数学进行拟合,发现基于核数的 6 层划分后各层节点比例大致服从瑞利分布(拟合曲线为式(2)向左平移 0.3),说明在层次模型中  $Hcp_i$  取值可适当偏离表 2 中所列参考值,但比例关系必须受瑞利分布约束。

$$f(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2}{2\sigma^2}\right), z \geq 0, \sigma = 1.3 \quad (2)$$

基于核数的层次模型充分体现了网络核心层的影响力渗透到网络各层.为在 ASRTM 模型中刻画影响全网的核心,将第 1 层设定为初始网络,在演化前完成核心层内节点的生成与边的连接及 AS 关系的指定。

### 4.2 AS 关系模型

AS 关系模型可抽象为:一个简单图  $G(V, E)$  加一个边到 AS 关系的映射函数  $\Psi: E \rightarrow \{c2p, s2s, p2p, p2c\}$ 。

表 3 ASRTM 模型 AS 层间关系示意和模型参数

	1	2	3	4	5	6
1	$c2p + p2p + p2c$ (19.07, 61.86, 19.07)	$p2p + p2c$ (51.62, 48.38)	$p2p + p2c$ (16.00, 84.00)	$p2p + p2c$ (2.42, 97.58)	$p2c$	$p2c$
2	$c2p + p2p$ (48.38, 51.62)	$c2p + p2p + p2c$ (27.66, 44.68, 27.66)	$c2p + p2p + p2c$ (3.61, 9.25, 87.14)	$p2p + p2c$ (3.02, 96.98)	$p2c$	$p2c$
3	$c2p + p2p$ (84.00, 16.00)	$c2p + p2p + p2c$ (87.14, 9.25, 3.61)	$c2p + p2p + p2c$ (28.81, 42.38, 28.81)	$c2p + p2p + p2c$ (2.41, 1.31, 96.28)	$p2c$	$p2c$
4	$c2p + p2p$ (97.58, 2.42)	$c2p + p2p$ (96.98, 3.02)	$c2p + p2p + p2c$ (96.28, 1.31, 2.41)	$c2p + s2s + p2c$ (48.47, 3.06, 48.47)	$c2p + s2s + p2c$ (0.60, 1.43, 97.97)	$s2s + p2c$ (1.59, 98.41)
5	$c2p$	$c2p$	$c2p$	$c2p + s2s + p2c$ (97.97, 1.43, 0.60)	$c2p + s2s + p2c$ (48.57, 2.86, 48.57)	$s2s + p2c$ (1.72, 98.28)
6	$c2p$	$c2p$	$c2p$	$c2p + s2s$ (98.41, 1.59)	$c2p + s2s$ (98.28, 1.72)	$c2p + s2s + p2c$ (48.44, 3.12, 48.44)

#### 4.2.1 模型算法

ASRTM 模型中节点被定义为二元组  $Node(n\_num, h\_num)$ ,其中  $n\_num$  为节点序号,  $h\_num$  为节点所归属层次号.以下是算法描述:

- 模型输入:  $N$
- 模型输出:网络拓扑的邻接矩阵  $A$  和含 AS 关系的邻接矩阵  $AR$
- 模型初始化:建立 6 个层集合  $\Omega_i$ ,并依  $Hcp_i$  确定各集合的大小  $|\Omega_i| = N_i = Hcp_i \times N, i = 1, 2, \dots, 6$ .
- 网络演化.

##### Step1 生成初始网络

$m_0 = N \times 0.253\%, e_0 = N \times \bar{k} \times 2.094\%, \bar{k} = 4.355$ , 初始网络节点间随机连接,边的 AS 关系依  $p_{11} = [c2p, p2p, p2c] = [19.07, 61.86, 19.07]/100$  随机指定,并记录到  $A$  和  $AR$  中,形成  $\Omega_1$ .因  $e_0$  较  $m_0$  而言稠密得多,能确保每个节点的度至少为 1.

##### Step2 网络的增长演化

- (1)在每一个时间步增加一个新节点  $vs$ .
- (2)依概率  $Hcp_i$  选择新节点加入的层次.当且仅当  $|\Omega_i| \leq N_i$  时,  $vs$  加入  $i$  层,  $vs.h\_num = i, vs \in \Omega_i$ ; 否则

由 ASRTM 模型结构和实测数据的 AS 关系层析分析结果,AS 关系建模只需在网络演化过程中对新添加边完成 AS 关系的指定即可.AS 关系模型的思想是:边的 AS 关系由端节点的层次归属依概率确定,当端节点由低层次指向高层次时以  $c2p$  关系为主;反之,由高层次指向低层次时以  $p2c$  关系为主; $p2p$  关系主要存在网络核心层和传输层间; $s2s$  关系主要在网络终端客户层间,尤其是与第 4 层相关的连接中.具体建模时要注意:在相应层次适当放大  $s2s$  和  $p2p$  关系所占比例,以反映“实际的 AS 级互联网可能存在比推测统计略低的  $p2c$  和  $c2p$  连接和比推测统计更多的  $p2p$  和  $s2s$  连接<sup>[2]</sup>”的事实;不仅要刻画层间的关系,还要刻画层内的关系;刻画时还要“突出主要关系,忽略次要关系”.具体 AS 关系模型见表 3.

回到(2).

(3)当  $1 \leq i \leq 5$  时,为新节点增添  $m$  条连接.从  $vs$  出发,引出  $m$  条边与目标节点  $vd_l(l = 1, \dots, m)$  相连.对目标节点  $vd_l$  的选择为:依概率  $Hp_j$  先选择  $vd_l$  归属的层次  $j, vd_l \in \Omega_j, vd_l.h\_num = j$ ; 如果  $|\Omega_j| \leq 3$ , 则视目标层次集合为空,即刻为  $\Omega_j$  增加一个节点  $vn = vd_l \in \Omega_j, vn.h\_num = j$ , 并依表 3 概率在  $AR$  中连接  $vs$  与  $vn$ , 同时更新  $A$  中的连接标记,然后转入(4).否则,从  $\Omega_j$  中依式(3)非线性择优出目标节点  $vd_l$ , 并依表 3 概率在  $AR$  中连接  $vs$  与  $vd_l$ , 同时更新  $A$  中的连接标记;对  $vs$  节点重复此操作  $m$  次.更新当前网络内的节点数.

(4)为一致对新加入节点的操作,还需对  $vn$  增加  $m - 1$  条边.

重复上述过程,至网络增长到期望规模  $N$ .

#### 4.2.2 $m$ 的取值与概率

ASRTM 模型中,为新节点分配的连接数为  $m$ .如果  $m$  的取值为常数值,则将导致各层次节点度值过于平均,  $m$  的取值应与节点归属的层次有关且为变值.为探究不同层次节点连接数  $m$  的分布规律,对各实测数据各层节点的度分布进行分析,以 20100120 数据为例,第

6层有 97.84% 的节点度值为 1, 不过该层也含有少量度大于 1 的节点. 说明加入第 6 层的节点应以不同的概率拥有不同的连接数, 其他层的情况类似. 对各实测数据各层节点的度分布分别进行曲线拟合, 得到  $m$  的数学模型, 如式(3)所示. 据此处理, 能确保各层节点的平均连接数接近期望值.

$$p(m) = m^{-\gamma} \times 10^{\alpha}$$

C2:  $\gamma = 1.9 \quad \alpha = 1.0 \quad m = 10 \sim 232$   
 C3:  $\gamma = 2.3 \quad \alpha = 0.9 \quad m = 4 \sim 35$   
 C4:  $\gamma = 3.5 \quad \alpha = 1.3 \quad m = 3 \sim 11$   
 C5:  $\gamma = 4.3 \quad \alpha = 1.0 \quad m = 2 \sim 6$   
 C6:  $\gamma = 5.5 \quad \alpha = 0 \quad m = 1 \sim 3$  (3)

### 4.2.3 择优连接概率

大多拓扑模型都以优先连接理论为基础, 不同的是各模型的优先连接概率的计算公式. ASRTM 也采用优先连接方法选择目标节点, 择优连接概率如式(4). 以往的优先连接概率多以节点度值为计算基数, 式(4)略有不同的是基数选取为节点度值与节点归属层次各层平均连接数  $\bar{m}$  (表 2) 之差. 这是因为, 网络演化之初, 各层内各节点的连接数相差不大, 使得一般的择优概率难以发挥作用, 故要去除各层相应的  $\bar{m}$  值后再进行择优. 为确保选择概率不为 0, 选择二者之差 + 1 作为计算基数.

$$P(i) = \frac{(k_i - \bar{m} + 1)^{1+\delta \lg(k_i - \bar{m} + 1)}}{\sum_j (k_j - \bar{m} + 1)^{1+\delta \lg(k_j - \bar{m} + 1)}}, \delta \in [0, 1], \delta = 0.35$$
 (4)

## 5 计算机建模仿真

由于网络动态演化过程中各节点选择层次和目标

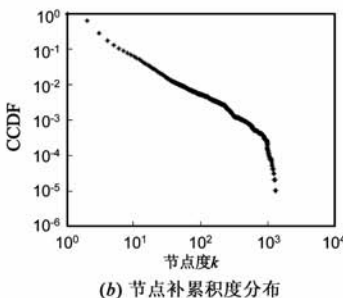
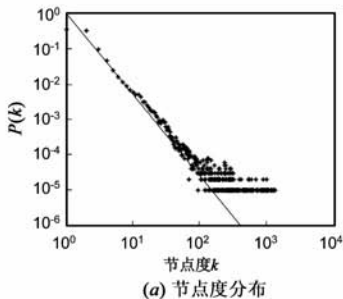


图3 ASRTM模型的节点度分布与补累积度分布

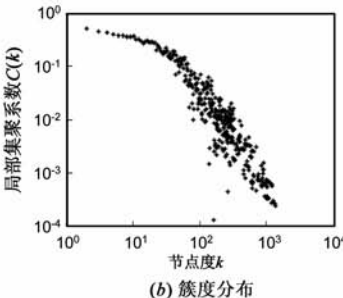
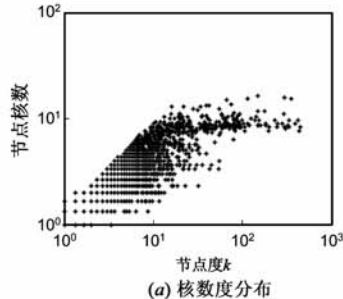


图4 ASRTM模型的节点核数度分布与簇度相关性

节点进行 AS 商业关系的连接均依概率发生, 每次仿真得到的网络拓扑有所不同, 本文采取多次实验取平均的方法来统计 ASRTM 模型拓扑的宏特征、层次特性和 AS 关系特性. 文中所有仿真结果均为十次独立实验取平均, 网络节点数为 20000.

### 5.1 模型的宏特征

表 4 中同时列出真实 AS 级互联网的宏特征, 均为实测数据的算术平均, 可看出 ASRTM 模型在平均节点度  $\langle k \rangle$ 、平均最短距离  $\langle d \rangle$ 、平均集聚系数、网络的同配系数  $r$ 、1% 最高度值节点间的 Rich-club 系数  $\varphi(1\%)$ 、度为 1, 2, 3 的节点在网络中所占的比例方面与真实的 AS 级拓扑的宏特征接近.

图 3 给出 ASRTM 模型的节点度分布和补累积分布. 可看出 ASRTM 模型的节点度分布呈幂律, 且幂律指数并不随网络规模的增大而变化, 幂指数在 2.3 附近, 很好地刻画出 AS 级拓扑的幂律特性. 充分说明 ASRTM 模型在幂律特性方面有突出表现, 从另一个侧面说明, 增长与优先连接并非唯一决定网络拓扑幂律特性的重要因素, 层次性也影响着网络的幂律特性.

表 4 ASRTM 模型和 Internet AS 拓扑宏特征对比

	$\langle k \rangle$	$\langle d \rangle$	$\bar{C}$	$r$
Real AS	4.355	3.775	0.257	-0.202
ASRTM	4.746	3.05	0.273	-0.271
	$\varphi(1\%)$	$Per_{k=3}$	$Per_{k=2}$	$Per_{k=1}$
		(%)		
Real AS	0.236	10.24	41.95	33.36
ASRTM	0.326	9.740	39.908	34.569

### 5.2 模型的层次特性

ASRTM 模型拓扑图的核数约为 22 ~ 27, 与模型层

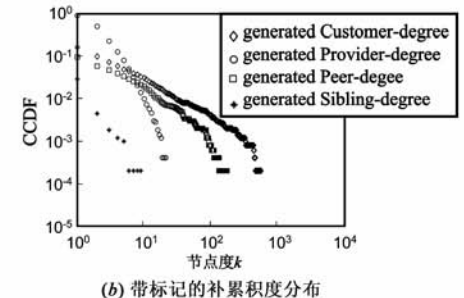
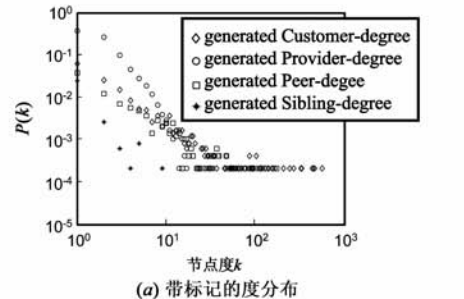


图5 ASRTM模型的带标记的度分布与补累积度分布

数和演化规模有关,比真实的 AS 级拓扑图核数(28)略低,但接近。

图 4 给出模型的节点核数-度分布和簇度分布图。可看出,ASRTM 模型呈现类似真实网络的节点核数-度分布趋势和负的簇度相关性且斜率大致在  $-0.7$  左右。

### 5.3 模型的 AS 关系特性

统计 ASRTM 模型输出的 AR 商业关系比例:  $Per_{c2p} = 44.423\%$ ,  $Per_{p2c} = 44.423\%$ ,  $Per_{p2p} = 10.911\%$ ,  $Per_{s2s} = 0.243\%$ 。图 5 给出各带标记的度分布与补累积度分布图。可知,ASRTM 模型中各关系比例与真实互联网的实际比例相当;并较好地模拟出带标记的度分布与补累积度分布,略有差异的原因在于 AS 关系建模时采取了“抓大放小”的策略,只在第 1、2、3 层进行  $p2p$  关系连接,在第 4、5、6 层进行  $s2s$  关系连接,但 4 种 AS 关系的度分布与补累积分布大的形态与趋势上与真实网络的带标记的度分布与补累积度分布一致,说明参数的误差不影响模型的正确性和有效性。

### 5.4 AS 关系推测准确性验证

ASRTM 模型也属于一种 AS 关系推测算法,其变体可对实际的或模拟生成的互联网无向拓扑图进行 AS 关系推测。通过验证推测算法的准确性来辅证 AS 关系模型的有效性。

AS 关系推测算法准确性验证,一种方法是通过实际调查一些 AS 系统间的真实关系,来检验推测算法在抽取样本上的准确性,该方法可以检验任何推测算法的有效性,但需要依赖特定网络管理员提供的正确信息,同时只能检验一定数量样本的准确性;另一种方法是通过计算合法 AS 路径的比例,该方法不需要依赖第三方提供的信息,但只能用于 TOR 问题解答的检验,不能用来衡量推测算法的准确度<sup>[2]</sup>。本文采用“类第一种”的验证方法,从未参与 AS 关系建模的数据集中选取 20080910 用于验证、评价推测算法的准确度。20080910 数据集构造的无向图 A 包含 29579 个 AS 和 66969 条无向边,将本文推测算法应用于 A 推测其 AS 关系,形成 AR1(原始的 AS 关系有向图是 AR)。在 133938 条有向边中,一次推测实验结果如表 5 所示。为衡量推测算法的准确度,将 AR1 与 AR 逐条比对,并进行分层统计,验证结果见表 5 中匹配率部分。

由表 5 可以得出:整体上 AR1 中各 AS 关系比例与 AR 中各 AS 关系比例相当,且呈现出比 AR 略低的  $c2p$  和  $p2c$  连接和更多的  $p2p$  连接;分层看,随着 AS 层次的降低,对 AS 商业关系推测准确度随之降低,即对从高层 AS 节点出发的边的推测准确度要高于对从底层 AS 节点出发的边的推测准确度;各层中  $p2c$  连接匹配最好;而  $p2p$  连接在第 1 层的匹配率近一半,  $s2s$  连接在

这一次实验中,仅第 3 层有匹配,匹配率仅为 10%。本文 AS 关系推测准确度与经典的 AS 关系推测算法相当。

表 5 20080910 数据集 AS 关系推测验证结果统计(%)

	<i>N</i>	<i>E</i>	<i>c2p</i>	<i>s2s</i>	<i>p2p</i>	<i>p2c</i>
AR	29579	133938	44.596	0.317	10.491	44.596
AR1	29579	133938	43.849	0.248	12.053	43.849
		各层	<i>c2p</i>	<i>s2s</i>	<i>p2p</i>	<i>p2c</i>
匹配率 %	C1	79.238	16.667	-	47.536	87.899
	C2	90.221	18.639	-	25.000	94.681
	C3	87.279	25.206	-	17.910	91.988
	C4	85.824	50.378	10.256	-	90.093
	C5	65.504	50.840	0.000	-	69.028
	C6	48.159	47.619	0.000	-	49.587
	总的		86.987	1.887	46.342	86.987

对 ASRTM 模型全局、局部的分析表明,该模型不仅较为准确地刻画出网络的幂律特性,还较好地刻画了真实网络的 AS 关系;该模型依据实测数据优化模型参数,是一种能够逼真模拟互联网 AS 级拓扑宏观、微观特性的动态演化模型。

## 6 结束语

AS 间商业关系的存在给互联网拓扑建模带来新的需求和挑战。本文基于真实 AS 级互联网拓扑实测数据进行层次特性和 AS 关系特性分析,提出了一种基于核数分层的 AS 关系拓扑动态演化模型(ASRTM)。仿真对比分析结果表明该模型不仅从整体上拥有与实际网络相似的拓扑宏特性,而且再现了真实互联网 AS 间商业关系的规律。同时,作为一种 AS 关系推测算法,能为任意的 AS 级无向图提供带 AS 关系标注的拓扑图,扩充了模型的应用。此外,模型参数的可测性为互联网 AS 关系和层次结构提供了一种内在的解释,对互联网的拓理论及应用具有一定的实际意义。

### 参考文献

- [1] 张国清. 互联网拓扑结构知识发现及其应用[J]. 通信学报, 2010, 31(10): 18-25.  
Zhang G Q. Discovery of internet topology knowledge and its applications[J]. Journal on Communications, 2010, 31(10): 18-25. (in Chinese)
- [2] 张国强. 互联网拓扑结构的研究[D]. 中国科学院研究生院博士学位论文, 2008.  
Zhang G Q. Research on Internet Topology[D]. Doctorate Dissertation, Graduate University of the Chinese Academy of Sciences, 2008. (in Chinese)
- [3] 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.
- [4] Zhang G Q, Zhang G Q, Yang Q F, et al. Evolution of the Internet and its cores[J]. New Journal of Physics, 2008, 10(12): 123027.
- [5] Dimitropoulos X, Riley G. Modeling autonomous-system relationships[A]. Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation[C]. Piscataway, NJ: IEEE, 2006. 143 – 149.
- [6] Gao L X. On inferring autonomous system relationships in the Internet[J]. IEEE/ACM Transactions on Networking, 2001, 9(6): 733 – 745.
- [7] Dimitropoulos X, Krioukov D, Huffaker B, et al. Inferring AS relationships: Dead end or lively beginning? [A]. Proceedings of the 4th Workshop on Efficient and Experimental Algorithms (WEA05)[C]. Berlin: Springer, 2005. 113 – 125.
- [8] Xia J, Gao L. On the evaluation of AS relationship inferences [A]. Proceedings of IEEE GLOBECOM[C]. Piscataway, NJ: IEEE, 2004. 1373 – 1377.
- [9] Qiu Jian, Gao Lixin. AS path inference by exploiting known AS paths[A]. Proceedings of GLOBECOMp06[C]. Piscataway, NJ: IEEE, 2006. 1 – 5.
- [10] Dimitropoulos X, Krioukov D, Fomenkov M, et al. AS relationships: Inference and validation [J]. ACM SIGCOMM Computer Communications Review, 2007, 37(1): 29 – 40.
- [11] Dimitropoulos X, Krioukov D, et al. Graph annotations in modeling complex network topologies [OL]. [http://arxiv.org/bbs/0708\(2007\)](http://arxiv.org/bbs/0708(2007)).
- [12] CAIDA. AS links annotated with AS relationships Dataset [OL]. [2010-01-20]. <http://www.caida.org/data/active/as-relationships/index.xml>.
- [13] Subramanian L, Agarwal S, Rexford J, et al. Characterizing the Internet hierarchy from multiple vantage points[A]. Proceedings of INFOCOM[C]. USA: IEEE Press, 2002. 618 – 627.
- [14] Battista G D, Patrignani M, Pizzonia M. Computing the types of the relationships between autonomous systems[A]. Proceedings of INFOCOM[C]. USA: IEEE Press, 2003. 156 – 165.
- [15] 张珂, 黄永峰, 李星. 一种基于适应度和节点聚类的 P2P 拓扑建模方法[J]. 电子学报, 2010, 38(7): 1634 – 1640.
- Zhang Ke, Huang Y F, Li X. A model for topology of P2P network based on fitness and node clustering[J]. Acta Electronica Sinica, 2010, 38(7): 1634 – 1640. (in Chinese)
- [16] 陈松, 王珊, 周明天. 基于实时分析的网络测量抽样统计模型[J]. 电子学报, 2010, 38(5): 1177 – 1180.
- Cheng S, Wang S, Zhou M T. Network data measurement and statistics model based on real-time analysis[J]. Acta Electronica Sinica, 2010, 38(5): 1177 – 1180. (in Chinese)
- [17] Ravasz E, Barabási A L. Hierarchical organization in complex networks[J]. Physical Review E, 2003, 67(2): 026112.
- [18] Govindan R, Reddy A. An analysis of internet inter-domain topology and route stability[A]. Proceedings of IEEE INFOCOM[C]. USA: IEEE Press, 1997. 850 – 857.
- [19] Carmi S, Havlin S, Kirkpatrick S, et al. A model of internet topology using  $k$ -shell decomposition[A]. Proceedings of the National Academy of Sciences[C]. USA: NAS, 2007. 11150 – 11154.

#### 作者简介



郭虹女, 1975年2月出生, 云南玉溪人. 1997年、2000年在信息工程学院获工学学士、工学硕士学位. 信息工程学院博士生、讲师, 研究方向为复杂网络、网络拓扑建模.

E-mail: guo\_hong@163.com



兰巨龙男, 1962年3月出生, 河北张家口人. 博士, 解放军信息工程大学信息工程学院教授, 博士生导师, 主要研究方向为路由技术、宽带信息网络.

E-mail: ndsclj@163.com

汪涛男, 1978年出生, 湖北黄石人. 博士, 解放军信息工程大学信息工程学院讲师, 主要研究方向为无线宽带通信网络.

刘洛琨男, 1963年出生, 河南洛阳人. 博士, 解放军信息工程大学信息工程学院教授, 主要研究方向为无线宽带通信网络.