

基于 hops 的 Internet 复杂网络分割度分析

苏威积, 赵海, 徐野, 张文波

(东北大学 信息科学与工程学院, 辽宁 沈阳 110004)

摘要: 首先讨论了相同分割度对不同网络传播行为的影响, 并以小世界网络为例讨论了不同分割度对相同网络传播行为的影响; 定义了基于 hops 的 Internet 分割度, 根据 CAIDA 提供的全球范围 30 个主要 monitor 5 年以来采集的数据, 对 Internet hops 进行统计分析, 并计算出 Internet 分割度, 与国际该领域研究成果进行相比, 更进一步揭示了 Internet 小世界效应的规律; 针对 hops 与 RTT 进行了 10 种模型的回归分析, 得出 hops 与 RTT 更符合幂函数关系及其关键参数, 从统计上证明了 Internet 复杂网络传播行为的幂率关系, 并以此建立了 Internet 分割度时间敏感性模型 (ISTSDM); 建立了针对 Internet 分割度的时间序列随机过程模型 (MTSSPISD), 并以此讨论了 Internet 分割度的时间演化规律; 最后利用模型 ISTSDM 和模型 MTSSPISD 对 2008 年北京奥运期间 Internet 分割度和 IP 层数据平均传播时间进行了预测。

关键词: 分割度; 集群度; 特征路径长度; hops; RTT; 复杂网络; 小世界效应

中图分类号: TN91

文献标识码: A

文章编号: 1000-436X(2005)09-0001-08

Internet complex network separation degree analysis based on hops

SU Wei-ji, ZHAO Hai, XU Ye, ZHANG Wen-bo

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

Abstract: At the first the infection of the same separation degree to different network's transmission was discussed, then the infection of different separation degree to same network's transmission was discussed by giving an example of small world network; Internet separation degree based on hops was defined. Based on the data that were supported by 30 monitors of global Internet from CAIDA, Internet hops was analyzed by statistic, Internet separation degree was computed. Small world character of Internet was illustrated more amply than the other international research result in this field; Aim at hops and RTT, 10 kinds regression analysis were done and a conclusion was draw that it more accords with power function between hops and RTT, the key parameters were found, which proved Internet complex network transmission is power function. Internet separation degree time sensitivity model (ISDTSM) was created; The model of time series stochastic process of Internet separation degree (MTSSPISD) was created, and then the time evolvement rule of Internet separation degree was discussed; At last, the Internet separation and average transmission time of IP level data during 2008 Beijing Olympic Game were forecast by using model ISDTSM and model MTSSPISD.

Key words: separation degree; clustering coefficient; average path length; hops; RTT; complex network; small world phenomenon

1 引言

Internet 是一个具有高冗余度的复杂网络系统。

两百多年来, 对网络理论的研究经历了三个阶段: 18 世纪提出规则网络阶段、19 世纪中叶随机网络阶段、近代的复杂网络阶段^[1]。

收稿日期: 2005-01-26; 修回日期: 2005-06-06

基金项目: 国家级火炬计划资助项目 (2002EB010154)

Stanley Milgram 的著名的邮递员实验，得出全世界 60 亿人，任意两个人都可以通过平均 6 个熟人联系起来，这种现象称为“六度分离”。1998 年，Watts 和 Strogatz 在他们的“small-world”理论^[2]中发现小世界网络同时拥有较大的集聚程度 C 和较小的最小距离 L ，又称分割度。从此启发了计算机科学家用复杂网络理论研究网络，Yook S H 等人在域的层次上计算了 Internet 的小世界特征量^[3]。Adamic L A 等人在 WWW 层上计算了其小世界特征量^[4]。

由于对 Internet 节点与连接数量巨大、拓扑的动态变化、以及数据获取、统计等的复杂性，对 Internet 的静态物理量的统计和计算已经成为 Internet 复杂性研究的关键。本文首先分析了分割度对网络传播的影响，其次提出了以 *hops* 为统计量，定量地分析了 Internet 的复杂网络特征量——分割度，并进行了 10 种模型的回归分析。建立模型 ISDTSM 和 MTSSPISD 分别对 Internet 分割度和 IP 层数据平均传播时间进行了计算。

2 分割度对网络的影响

2.1 分割度的概念

复杂网络的静态几何量描述了复杂网络微观的构成，是复杂网络统计分析的基础。Internet 在 IP 层是无向网络，对于无向网络其静态几何量包括：度及其分布特征，度的相关性，集聚程度及其分布特征，最短距离及其分布特征，介数(*betweenness*)及其分布特征，连通集团的规模分布，而重点讨论分割度。

定义 1 分割度又称特征路径长度，假设某个普通的拓扑网络图 G 是连通图， G 中有 N 个节点和 k 条无权边，且 $k \ll \frac{N(N-1)}{2}$ ，即 G 是稀疏的。这里 G 的关联矩阵为 $N \times N$ 维的矩阵，用 $A = \{a_{ij}\}$ 表示。若点 i 与 j 有边相连，则 $a_{ij} = 1$ ，反之， $a_{ij} = 0$ 。利用图论与网络理论方法可得 G 的最短路径矩阵 $D = \{d_{ij}\}$ ，特征路径长度为

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} l_{ij} \quad (1)$$

其中， L 表示 G 中两点的平均距离， l_{ij} 表示 i, j 两点的实际距离。

分割度体现了网络的连通性，分割度越低，说

明数据在网络上平均传播所经过的中间点越少，传播时间越短。分割度越高说明数据经过的网络中间节点越多，传播时间越长。

2.2 相同分割度对不同网络的影响

图 1 是节点数为 200，分割度都为 4.27 的 4 种基本网络拓扑结构。它们的传播分布情况如图 2。

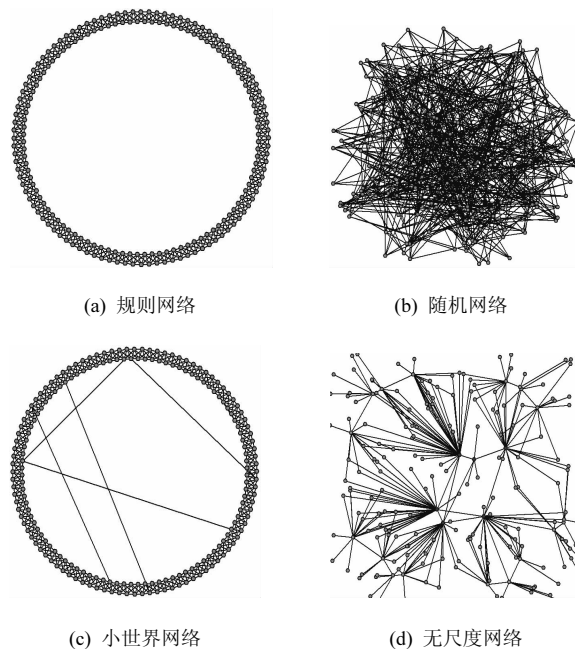


图 1 4 种网络的拓扑结构

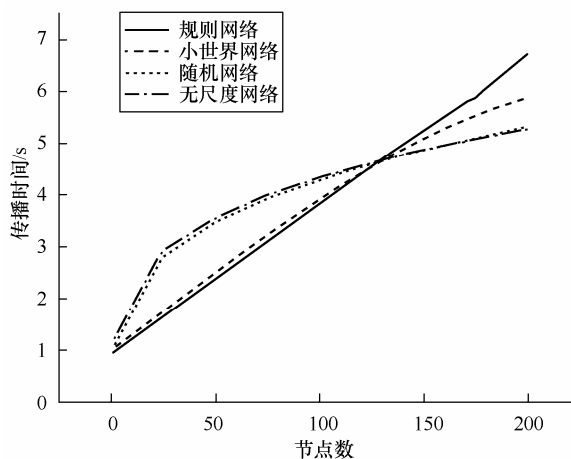


图 2 分割度相同的 4 种网络传播特征曲线

从图 2 可以看出规则网络服从线性函数规律 $Y = b_0 + b_1 X$ ，小世界网络、随机网络、无尺度网络呈现参数不同的幂率分布 $Y = b_0 X^{b_1}$ 。具体参数如表 1， b_0 为常数项， b_1 为幂指数。

表 1 3 种网络的幂率关系参数

网络	b_0	b_1
小世界网络	0.8804	0.3250
随机网络	1.0298	0.3095
无尺度网络	1.1696	0.2839

表 2 分割度相同的 4 种网络传播时间/s

节点数	规则网络	小世界网络	随机网络	无尺度网
1	0.9257	1.01441	1.02980	1.16960
25	1.6241	1.73700	2.78879	2.91684
50	2.3516	2.48700	3.45608	3.55120
75	3.0791	3.21575	3.91818	3.98444
100	3.8066	3.90450	4.28305	4.32352
125	4.5341	4.53450	4.58930	4.60628
150	5.2616	5.08700	4.85571	4.86098
175	5.8919	5.54325	5.09229	5.06799
200	6.7166	5.88450	5.30789	5.26380

在分割度相同的情况下，网络传播整体所需时间略有不同。如表 2 所示，由 200 个节点构成的 4 种网络中平均传播时间为 5.793s，规则网络所需时间高于 4 种网络中传播最快的无尺度网络 17.1%，小世界网络传播时间高于无尺度网络 11.79%。随机网络高于无标度网络 0.8376%，与无尺度网络传播时间几乎相同。

2.3 不同分割度对相同网络的影响

在小世界网络中由于长程连接数量的不同导致网络分割度的不同。图 3 是分割度不同的 4 种小世界网络。图 4 是 4 种小世界网络曲线的拟合结果。

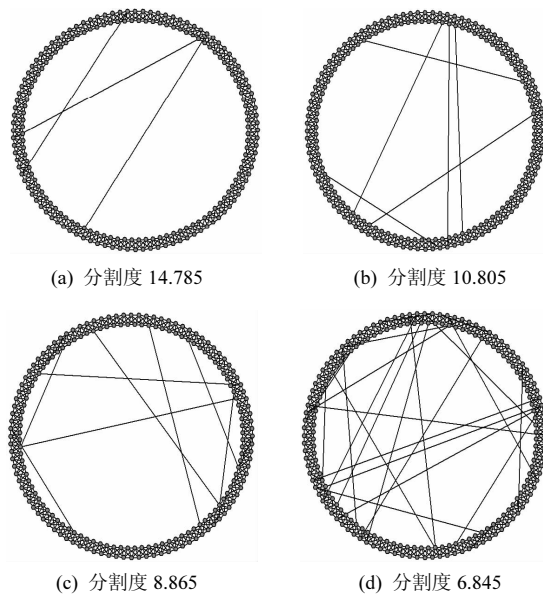


图 3 分割度不同的 4 种小世界网络

根据幂指数方程： $Y = b_0 X^{b_1}$ 。X 为节点数；Y 为传播时间。表 3 是分割度分别为 14.785，10.805，8.865，6.845 的小世界网络，表 4 是此 4 种小世界网络的传播时间。

可以看出这 4 个小世界网络都符合幂率分布，由于分割度不同，传播时间随着分割度的增大而增大，呈现近似线性关系。

表 3 4 种小世界网络幂函数参数

网络	b_0	b_1
a	0.6837	0.6662
b	0.8228	0.5619
c	0.8884	0.5060
d	0.9483	0.4414

表 4 4 种小世界网络传播时间

节点数	a	b	c	d
1	0.68370	0.82280	0.88840	0.94830
25	5.83678	5.02108	4.52862	3.92642
50	9.26231	7.41217	6.43113	5.33177
75	12.13480	9.30875	7.89568	6.37673
100	14.69820	10.94190	9.13290	7.24012
125	17.05400	12.40360	10.22460	7.98954
150	19.25650	13.74170	11.21270	8.65909
175	21.33920	14.98500	12.12230	9.26878
200	23.32450	16.15260	12.96970	9.83151

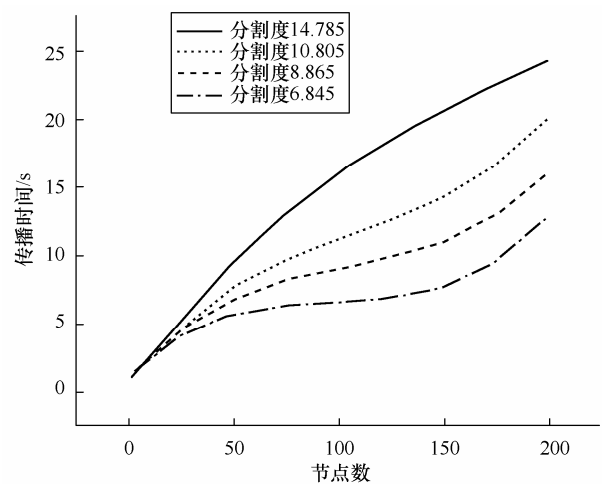


图 4 分割度不同的 4 种小世界网络传播特征曲线

分割度是研究复杂网络的基本尺度，体现了一个复杂网络系统的连通能力。分割度越高，网络中任意两个节点之间通信所经过的中间节点数越多，

平均通信时间越长；反之，分割度越低，网络任意两点通信所经过的中间节点越少，平均通信时间越短。从统计学上可以看出分割度体现了网络拓扑结构对网络传播能力的影响。

3 Internet 分割度分析

3.1 国外 Internet 分割度的研究

根据小世界网络的特性，Yook S H 等人在域的层次上计算了 Internet 的小世界特征量。Adamic L A 等人在 WWW 层上计算了其小世界特征量如表 5。

表 5 Internet 小世界特征表

Network	size	<i>l</i>	<i>C</i>
Internet ^[3] , domain level	3015~6209	3.7~3.76	0.18~0.3
WWW ^[4] , site level	153127	3.1	0.1078

列表中包括所研究网络的实际对象(network)，大小(size)，平均最短距离(*l*)，平均集聚程度(*C*)。通过对比实际网络与相应(相同顶点数和边数)随机网络的性质，可以发现 Internet 网络的小世界特征^[5]。

而 Faloutsos M、Faloutsos P、Faloutsos C、Govindan R、Tangmunarunkit H 根据无尺度网络的特征，分别在域和路由器的层次上统计 Internet 的特征。

表 6 Internet 无尺度特征

network	size	<i>l_{real}</i>	<i>l_{rand}</i>
Internet ^[6] , 域	3015~4389	6.3	6.3
Internet ^[6] , 路由器	3888	12.15	8.75
Internet ^[7] , 路由器	150000	11	12.8

列表中包括所研究网络的实际对象(network)，大小(size)，平均最短距离(*l_{real}*)，按随机网络计算的最短距离(*l_{rand}*)。可见无尺度网络一般具有小世界特征^[5]。

3.2 基于 hops 的分割度计算

TCP/IP 协议中 hops 代表在 Internet 上从源节点到目的节点 IP 包经过的路由器的个数，hops 从拓扑上体现了某一路径的网络路由效率和网络传输率。

定义 2 Internet 分割度定义为平均 hops，表示为 \overline{hops}

$$\overline{hops} = \frac{1}{N} \sum_{i=1}^N (hops)_i$$

其中 $(hops)_i$ 是第任意两个 IP 节点 A、B 构成的链路 *i* 的 hops，*N* 是总的链路数。

Internet 分割度 \overline{hops} 则反映了其宏观的网络特征。根据 CAIDA 全球 30 个节点 Monitor 在 2004 年 2 月 18 日监测的全球 Internet 所有固定 IP 的 hops，进行一维频数分析，统计结果如图 5。

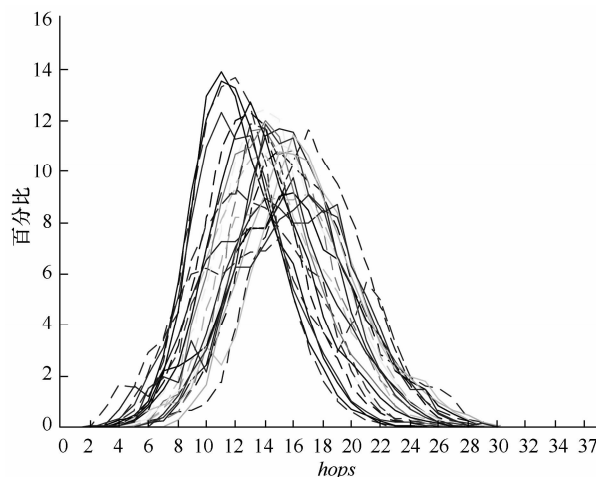


图 5 30 个 Monitor 捕获 hops 分布图

从此图可以看出不同 Monitor 捕获的 hops 在分布上有相似性。对全体数据进行一维频数分析，hops 的分布如图 6。统计结果见表 7，*N* 代表样本数量、*Mean* 为均值、*Variance* 为方差、*Skewness* 为正态分布的偏度、*Kurtosis* 为正态分布的峰度，及其标准差。*Skewness* 为 0.329 说明数据左偏，右尾较长，*Kurtosis* 为 0.111 说明变量值分布曲线略比标准正态分布高。由此说明 hops 分布符合均值为 14.44，方差为 16.07 的正态分布。Internet 的分割度为 14.44，说明全球任意两 IP 节点可通过平均 15 个中间节点相互通信。

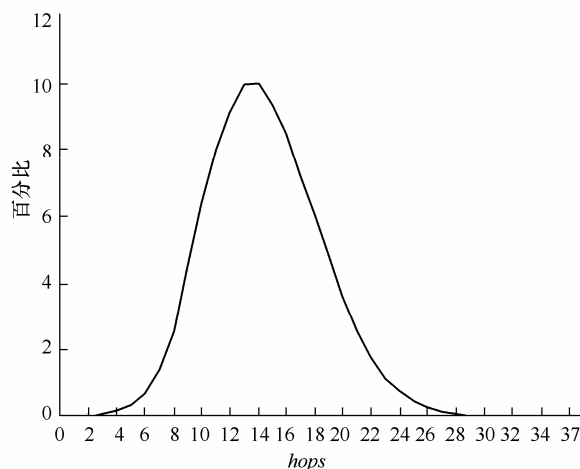


图 6 Internet hops 分布图

<i>N</i>	Mean	Variance	Skewness	Kurtosis
203662974	14.44	16.07	0.329	0.111

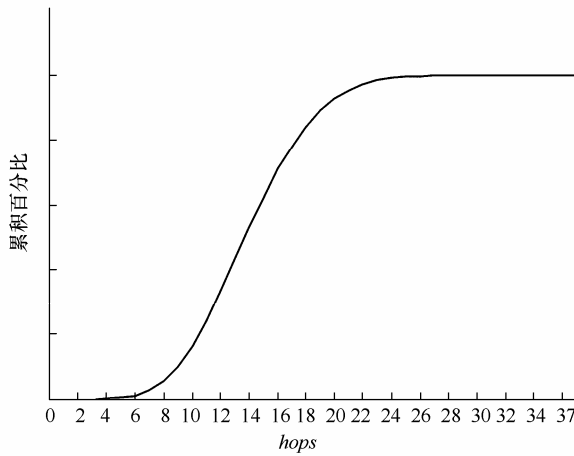


图 7 Internet hops 百分比分布图

图 7 表明全球 Internet 80% 的节点之间的通信可以在 18 跳之内完成，98% 的节点之间的通信可以在 24 跳之内完成。通过以上分析可知 Internet 具有小世界效应，样本比文献[3,4,6,7]更加具有代表性。

3.3 分割度对网络传播的影响

Watts 和 Strogatz 在文献[2]中提出对于从规则网络到随机网络，所有 $p \in (0,1)$ 的网络，其扩散时间刚好与最短路径一致。但并没有考虑到分割度的时间尺度，也就是通过集群度和分割度不能准确计算出网络的传播时间。如果要计算网络的传播只能通过其他物理量，并建立特定的传播模型。目前研究最为彻底，应用最为广泛的传播模型是 SIR 模型和 SIS 模型^[8~10]。

由于 Internet 其拓扑和协议的特殊性，可以通过 hops 与 RTT 相关分析，得出其分割度对传播时间影响。在全世界 30 个节点 Monitor 的 hops 与 RTT 统计结果如图 8 所示。为了确定 hops 与 RTT 之间确定性关系，对 hops 与 RTT 进行回归分析。首先，建立 10 个拟合曲线模型如下：

线性模型 Liner

$$Y = a + b_1X + b_2X \quad (2)$$

二次曲线模型 Quadratic

$$Y = b_0 + b_1X + b_2X^2 \quad (3)$$

复合曲线模型 Compound

$$Y = b_0 \times b_1^X \quad (4)$$

等比级数曲线模型 Growth

$$Y = e^{(b_0+b_1X)} \quad (5)$$

对数曲线模型 Logarithmic

$$Y = b_0 + b_1 \ln X \quad (6)$$

三次曲线模型 Cubic

$$Y = b_0 + b_1X + b_2X^2 + b_3X^3 \quad (7)$$

S 型曲线

$$Y = e^{(b_0 + \frac{b_1}{X})} \quad (8)$$

指数方程模型

$$Y = b_0 e^{b_1 X} \quad (9)$$

Inverse 曲线模型

$$Y = b_0 + \frac{b_1}{X} \quad (10)$$

幂率曲线模型 Power

$$Y = b_0 X^{b_1} \quad (11)$$

进行曲线拟合，拟合结果如图 9 所示。

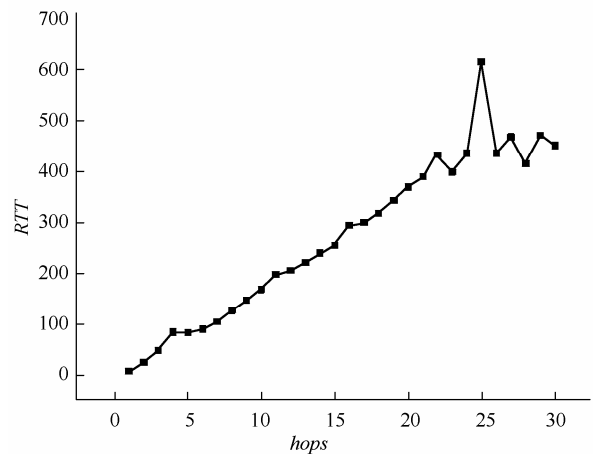


图 8 全世界 hops 与 RTT 相关分析

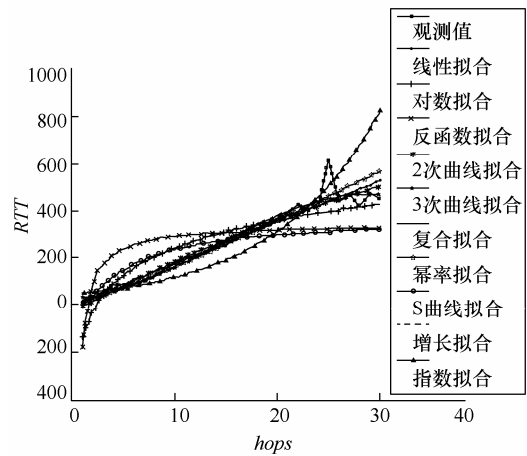


图 9 hops 与 RTT 回归分析

图 9 是图 8 的 10 种模型经过曲线拟合的结果，从图形上可以看出线性拟合、2 次曲线拟合、3 次曲线拟合、幂率拟合的拟合效果都很接近实际观测曲线。统计量对比分析，如表 8 所示。

表 8 曲线拟合检验表

Mth	R^2	$d.f.$	F	$Sigf$
Lin	0.934	28	393.91	0.000
Log	0.815	28	123.08	0.000
Inv	0.410	28	19.49	0.000
Qua	0.939	27	208.85	0.000
Cub	0.952	26	173.57	0.000
Com	0.765	28	91.03	0.000
Pow	0.976	28	1138.60	0.000
S	0.845	28	152.73	0.000
Gro	0.765	28	91.03	0.000
Exp	0.765	28	91.03	0.000

从实验结果来看，此 10 种模型均具有统计学意义，方差分析的 F 值概率值都很小，近似为 0.000。比较校正 R^2 值，幂率曲线模型的 $R^2=0.976$ 最大，拟合最好的是幂率曲线模型，参数为 $b_0=11.6179$ ， $b_1=1.1437$ 。说明 RTT 会随着 $hops$ 的增加幂指数增加

$$Y = 11.6179X^{1.1437} \quad (12)$$

称为 Internet 分割度时间敏感性模型 (ISTSDM)。

3.4 Internet 分割度的时间演化

我们利用分布在全球不同地理位置的 30 个主要路由器节点作为源节点，对整个 Internet 进行定时扫描，捕获路由信息。对从 2001 年 7 月到 2004 年 6 月全球 Internet 数据进行分析，分析结果如图 10。可以看出不同节点在不同时间的 $hops$ 都有着趋势性、时间敏感性、随机性。

图 10，图 11 中纵坐标代表 $hops$ (均跳数)，横坐标月数范围从 0 到 36，代表 2001 年 7 月统计到 2004 年 6 月，时间范围为 36 个月。

把 $hops$ 看成随机变量，不同节点不同时间的 $hops$ 可以看作是随机变量 $hops(t)$ 的不同试验，对于一个确定的时间 t ， $hops(t)$ 是一随机变量；而当 t 在某一时间段内 $[a, b]$ 变化时，就得到一族随机变量，记为 $\{hops(t), t \in [a, b]\}$ ，称为对于 $hops$ 时间敏感性演化的随机过程。

对于随机过程 $\{hops(t), t \in [a, b]\}$ ，其均值函数表

示为

$$\overline{hops}_t = E(hops(t)) \quad (13)$$

均值函数 \overline{hops}_t 表示随机过程 $hops(t)$ 在各个时刻的摆动中心，其变化趋势如图 12。

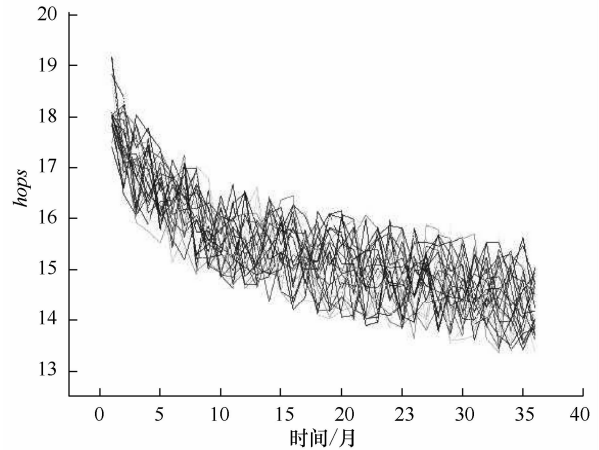


图 10 30 个节点连续 36 个月 hops 变化趋势曲线

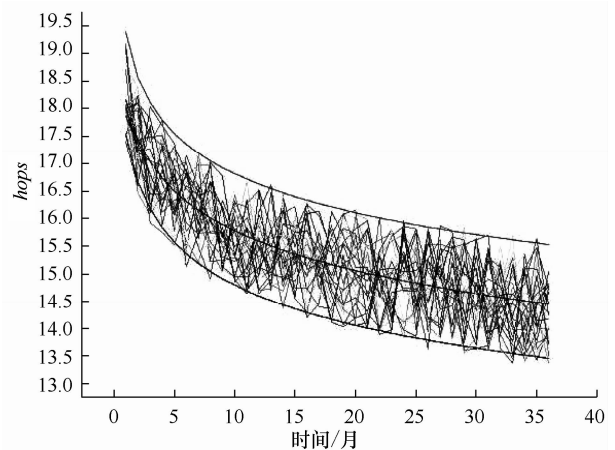


图 11 均值与方差函数的曲线变化趋势

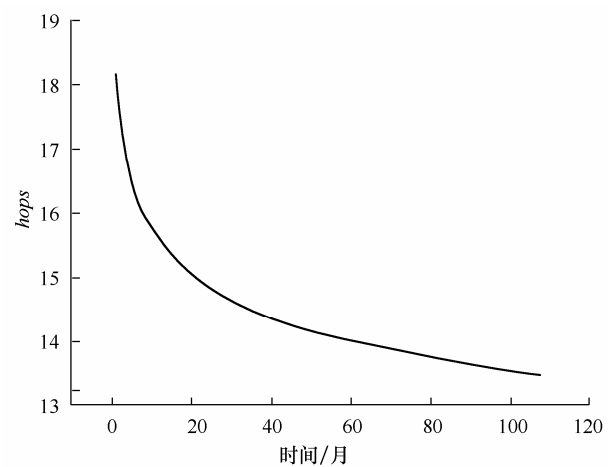


图 12 均值函数曲线

方差函数为

$$\text{var}(hops(t)) = E[(hops(t) - \overline{hops_t})^2] \quad (14)$$

方差函数 $\text{var}(hops(t))$ 表示随机过程 $hops(t)$ 对于均值 $\overline{hops_t}$ 的偏离程度，如图 13。

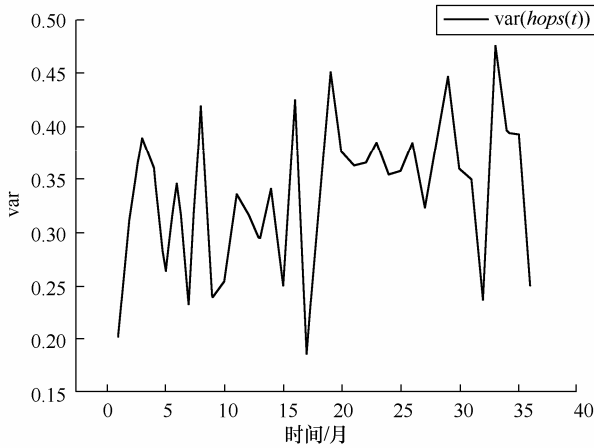


图 13 方差函数曲线

从图 12、图 13 可以看出随机过程 $hops(t)$ 是非平稳过程，平稳随机过程包括一阶平稳(要求 $EX(T) = m$ 为常数)，二阶平稳(要求 $\text{var}(T) = m$ 为常数)，广义平稳(协方差平稳)，显然数据显示其不满足平稳随机过程。很容易从图中看出其并不满足泊松分布，所以 $hops(t)$ 也非泊松过程。由于其将来的状态并不依赖于现在，故其也非马尔可夫过程。说明 Internet 分割度有着其自身的时间演化规律，不能用传统的时间序列模型进行描述。

经过计算，得出自 2001 年 7 月到 2004 年 6 月 36 个月的均值函数与方差函数的值。表 9 列举了其中 10 个月的价值。

表 9 10 个月均值函数、方差函数抽样值

时间(年.月)	$\overline{hops_t}$	$\text{var}(hops(t))$
2001.7	18.17370	0.2023450
2001.10	16.73290	0.3596300
2002.2	16.01250	0.4196660
2002.6	15.59110	0.3187560
2002.10	15.29210	0.4242880
2003.2	15.06020	0.3736800
2003.6	14.87070	0.3536870
2003.10	14.71050	0.3792790
2004.2	14.57180	0.2357890
2004.6	14.44930	0.2495680

利用回归分析得出 $\overline{hops_t}$ 的曲线函数符合幂函数关系

$$\overline{hops_t} = b_0 t^{b_1} \quad (15)$$

其参数分别为 $b_0=18.2979$, $b_1=-0.0652$, 则 $\overline{hops_t} = 18.2979t^{-0.0652}$, 称为 Internet 分割度的时间序列随机过程模型(MTSSPISD)。

而根据分割度时间演化模型(式(14)), 我们可以预计 2008 年 8 月北京奥运期间, 全球平均 hops 应为 13.7069 跳, 近似 14 跳。

进而根据 Internet 分割度时间敏感性模型 (ISTSDM) 公式, 可以预计 2008 年 8 月北京奥运期间, Internet 的 IP 层数据平均传播时间为 116ms。

4 结论

Internet 的异构性、动态性、发展的非集中性以及规模的巨大和拓扑的复杂性, 都促使人们用崭新的网络理论——复杂网络理论对其进行分析。

本文首先从不同角度讨论了分割度对复杂网络传播的影响。分别讨论了分割度相同的不同种类的 4 种网络: 规则网络、小世界网络、随机网络、无尺度网络的传播情况, 得出了分割度相同情况下的 4 种网络的传播比; 以小世界网络为例, 讨论了在分割度不同的情况下, 相同种类网络的传播特征, 得出同种网络传播特性与分割度近似呈线性关系的结论。

其次, 根据 Internet 的特殊性, 定义了 Internet 分割度, 确立了 Internet 分割度与 hops 之间的数学关系, 并根据 CAIDA 提供的全球范围 Internet 30 个主要节点 Monitor 5 年采集的数据, 对 hops 进行统计分析, 计算出 Internet 分割度为 14.44, 得出 Internet 小世界效应的规律。由于样本的巨大、真实性和可靠性, 本文所得的结果与国际该领域前人的结果相比, 更能代表 Internet 的真实情况。

第三, 针对 hops 与 RTT 进行了 10 种模型的回归分析, 得出 hops 与 RTT 更符合幂函数关系及其关键参数 $b_0=11.6179$, $b_1=1.1437$ 。从统计上证明了 Internet 复杂网络传播行为的幂率关系, 并以此建立了 Internet 分割度时间敏感性模型 (ISTSDM); 建立了针对 Internet 分割度的时间序列随机过程模型 (MTSSPISD), 并以此讨论了 Internet 分割度的时间演化规律。

最后利用模型 ISTSDM 和模型 MTSSPISD 对 2008 年奥运期间 Internet 分割度和 IP 层数据平均传播时间进行了预测。

参考文献:

[1] FLOYD S, PAXSON V. Difficulties in simulating the Internet[J]. IEEE/ACM Trans on Networking, 2001,9(4): 392-403.

[2] WATTS D J, STROGATZ S H. Collective dynamics of ‘small-world’ networks[J]. Nature, 1998,393(6684): 440-442.

[3] YOOK S H, JEONG H, BARABASI A L. Modeling the Internet’s large-scale topology[J]. PNAS 2002,99(13):13382-13386.

[4] ADAMIC L A, HUBERMAN B A. Evolutionary dynamics of the World Wide Web[J]. Nature, 1999,401(9): 131.

[5] ROMUALDO P S, ALBERT D G. Statistical Mechanics of Complex Networks[M]. Springer, 2003.

[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] GOVINDAN R, TANGMUNARUNKIT H. Heuristics for Internet map discovery[A]. Proceedings of IEEE INFOCOM[C]. Israel: Tel Aviv, 2000.1371.1-11.

[8] BAILEYN T. The Mathematical Theory of Infectious Diseases and Its Applications[M]. New York: Hafner Press, 1975.

[9] ANDERSON R M, MAY R. M. Infectious Diseases of Humans Dynamics and Control [M]. Oxford: Oxford University Press, 1992.

[10] HETHCOTE H W. The mathematics of infectious diseases[J]. SIAM Review, 2000, 42(4): 599-653.

作者简介:



苏威积 (1975-), 男, 辽宁抚顺人, 东北大学信息与工程学院博士生, 主要研究方向为信息融合、嵌入式技术、复杂网络。



赵海 (1959-), 男, 辽宁沈阳人, 东北大学教授、博士生导师, 主要研究方向为信息融合、嵌入式技术、复杂网络。



徐野 (1976-), 男, 辽宁沈阳人, 东北大学博士生, 主要研究方向为信息融合、嵌入式技术、复杂网络。



张文波 (1973-), 男, 黑龙江伊春人, 东北大学博士生, 主要研究方向为信息融合、嵌入式技术、复杂网络。