

因特网拓扑演化及其节点平均连接度的分形研究

张文波^{1,2}, 赵海², 孙佩刚^{2,3}, 徐野², 张昕²

(1 沈阳理工大学信息科学与工程学院, 辽宁沈阳 110168 2 东北大学信息科学与工程学院,
辽宁沈阳 110004; 3. 沈阳炮兵学院, 辽宁沈阳 110162)

摘 要: 针对因特网 (Internet) 的复杂网络特征, 通过计算 Internet 标准结构熵和时间的关系, 揭示了 Internet 的拓扑结构向着更有序方向演化的规律, 并利用协同学原理求出了影响其演化的序参量. 另外, 通过对 Internet 节点平均连接度时间序列进行相空间重构, 计算了其分形维数、最大 Lyapunov 指数和 Kolmogorov 熵, 从理论上分析了这组数据可预测的时间尺度. 最后, 根据分析的结果, 通过对时间序列进行预处理, 运用混沌预测方法对未来一段时间内 Internet 节点的平均连接度进行了预测.

关键词: 因特网; Internet 标准结构熵; 分形; 混沌预测

中图分类号: TP316 **文献标识码:** A **文章编号:** 0372-2112 (2006) 08-1438-08

Research on Internet Topology Evolution and the Fractal of Average Degree of Nodes

ZHANG Wen-bo^{1,2}, ZHAO Hai², SUN Pei-gang^{2,3}, XU Ye², ZHANG Xin²

(1 School of Information Science and Engineering, Shenyang Ligong University, Shenyang, Liaoning 110168 China;

2 School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004 China;

3 Shenyang Artillery Academy, Shenyang, Liaoning 110162, China)

Abstract In view of the complex network character of Internet, the computing result of the relation of Internet standard entropy with time shows that Internet evolves from low level order to high level one. Furthermore, the order parameter was also obtained by computing according to the synergetics theory. However, the time series data of average nodes degree were reconstructed in phase space. The fractal dimension, maximum Lyapunov parameter and Kolmogorov entropy were also computed. The predictable time scale of the serial data was got from the above computation. Finally, in means of the analytical results, the time series data were preprocessed, and the average nodes degree of Internet in near months was predicted by means of chaos predicting.

Key words Internet; Internet standard entropy; order parameter; fractal; chaos predicting

1 引言

目前, 对复杂网络演化和预测的研究已经取得了相当多的研究成果, 香港城市大学的陈关容教授及其合作者在时变复杂网络^[1]、复杂网络中的混沌现象^[2]以及节点驱动的动力网络^[3]等方面进行了深入的研究. 特别是其对复杂网络从非混沌向混沌状态的过渡问题的研究, 阐述了复杂网络中存在的混沌现象^[4].

不断涌现的关于复杂系统和复杂网络^[5,6]的研究成果使人们逐渐清晰地认识到, Internet 也是一个特殊的复杂网络系统^[7], 其整体特性不能简单地通过对各个子节点特性

叠加而得到, 而必须运用系统论的观点从宏观层次上予以把握, 只有这样才能对 Internet 的发展作出正确的判断. 大量的研究发现, Internet 的拓扑呈现出无尺度特征^[8,9]. 虽然其在域级和路由级可得到两个不同的拓扑^[10], 但已有研究结果显示无论哪一种拓扑度分布都服从幂律. 从域层次上看, 域为节点, 两个域边界路由器间的一个或多个连接作为一条边; 在路由器层, 节点是路由器, 边是对路由器之间的物理连接.

本文利用 Internet 节点的度分布, 计算了其标准结构熵^[11]的变化规律, 并通过协同学原理求出了影响其拓扑结构演化的序参量, 分析了 Internet 拓扑结构的有序演化

机制. 另外, 针对 Internet 的节点平均连接度的混沌特性, 文中计算了其分形维数 (关联维数)、最大 Lyapunov 指数和 Kolmogorov 熵^[12], 同时, 本文利用全球 Internet 监测数据, 研究了 Internet 拓扑结构演化过程中其平均连接度的分形特征.

2 Internet 拓扑结构的演化

2.1 Internet 节点的连接度

定义 1 度 (节点连接度) 设 N 是一个网络, $V(G)$ 是所有顶点的集合, $E(G)$ 是所有边的集合. 顶点 v 的度 d_v 是指与此顶点 v 连接的边的数量, $v \in V$. 即

$$d_v = \sum_{l \in E} \delta_l \quad (1)$$

其中 δ_l 记号取值为 1, 当 l 包含顶点 v , 否则为零, 即

$$\delta_l = \begin{cases} 1, & \text{if } l \text{ include } v \\ 0 & \text{if not} \end{cases} \quad (2)$$

定义 2 Internet 节点的平均连接度 $\bar{x} = \frac{1}{n} \sum_{i=1}^n d_i$

Internet 节点的平均连接度是 Internet 的一个重要参数, 它反映了 Internet 的平均连通性能. 这个值越大, 表明 Internet 的平均连通性能越好. Internet 节点的连接度分布 (简称度分布) 并不是均匀的, 通过计算发现, 当前 Internet 上约有 2% 的节点占有整个 Internet 上总连接度数的 80%, 而其余的节点只占有总连接度数的 20% 左右. 这也说明在 Internet 上节点之间具有明显的差异性, 少量的节点具有极高的连接度, 而大量的节点只有少量的连接.

Internet 节点的连接度呈幂函数分布, 称 Internet 具有 Power-law 或重尾分布 (heavy-tail) 特征, 可表示为 $p(k) \sim k^{-\alpha}$, 其中 k 为节点的连接度, α 是一个常数, $p(k)$ 是连接度为 k 的节点的概率, 即 Internet 的度分布.

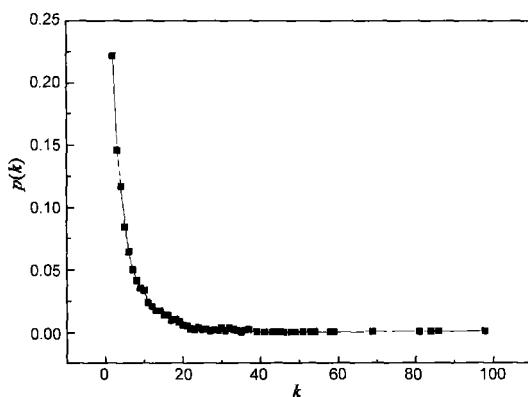


图 1 Internet 节点连接度 k 与 $p(k)$ 关系图

利用 CADA (The Cooperative Association for Internet Data Analysis) skitter 项目提供的全球 Internet 监测数据, 计算了 Internet 上 31408 个节点 (路由器) 的度数及度分布, $p(k)$ 和 k 的关系如图 1 和图 2 所示, 通过回归分析计算得到常数 α 的值是 2.66 相关系数为 0.949

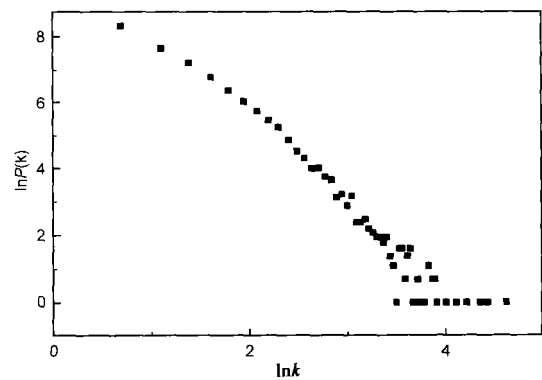


图 2 Internet 节点连接度 k 与 $p(k)$ 对数曲线

2.2 Internet 标准结构熵

熵最初是作为一个热力学概念而引入的, 是系统能量分布均匀性的一种度量, 可以表示物体所处状态是否稳定及系统变化的方向, 系统能量分布越均匀, 其熵越大; 反之, 则熵越小^[13]. 目前熵作为描述复杂系统结构的物理量, 在复杂系统理论中受到越来越多的关注, 成为研究复杂系统的一个重要工具.

定义 3 Internet 网络结构熵 $H = - \sum_{k=1}^{n-1} p(k) \cdot \lg p(k)$, 其中, $p(k)$ 为 Internet 中节点的度分布, n 是 Internet 中路由器的总数, 熵 H 表示 Internet 中度分布信息的不确定程度.

定理 1 当 Internet 结构完全均匀时, 网络结构熵取最大值, 即 $H = \lg_2 n$, n 为网络中节点总数.

证明 根据最大熵原理, 当设网络中节点总数为 n , 因为网络完全均匀, 所以所有的节点的度相同, 令其为 d , 这样, 网络中所有节点的度之和为 $\sum_{i=1}^n d_i = nd$, 那么所有节点的度分布都相同, 为 $1/n$. 将其带入 Internet 结构熵公式, 有

$$H_{\max} = - \sum_{k=1}^{n-1} \lg_2 \left(\frac{1}{n} \right) = -n \left(\frac{1}{n} \right) \lg_2 \left(\frac{1}{n} \right) = \lg_2 n \quad (3)$$

证毕.

定理 2 当 Internet 结构为星形时, 网络结构熵取最小值, 即 $H = \lg_2 n - [(n-1)/n] \lg_2 (n-1)$.

证明 当网络为星形结构时, 网络有一个中心节点, 其度分布为 $1/n$, 其余节点的度分布为 $(n-1)/n$. 根据网络结构熵公式有

$$\begin{aligned} H_{\min} &= - \left(\frac{1}{n} \lg_2 \frac{1}{n} + \frac{n-1}{n} \lg_2 \frac{n-1}{n} \right) \\ &= \lg_2 n - [(n-1)/n] \lg_2 (n-1) \end{aligned} \quad (4)$$

证毕.

为了消除节点数对网络结构熵造成的影响, 取 Internet 的标准结构熵 H_s 为

$$H_s = \frac{H - H_{\min}}{H_{\max} - H_{\min}} \quad (5)$$

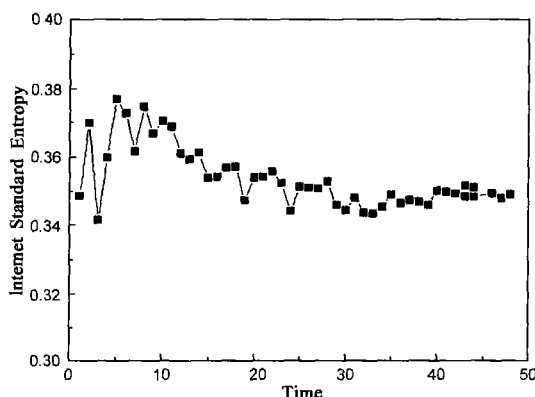
利用 CADA (The Cooperative Association for Internet

Data Analysis) skitter项目中 riesling和 apan_jp这两个网络监测点在 2000 年 4月到 2004 年 5月监测到的 Internet 数据,利用前面的公式,计算了每个月的 Internet网络标准结构熵的平均值,鉴于篇幅的限制,取相对于 2000 年 4月的相对时间,每 10个月取一个标准熵值,如表 1所示.

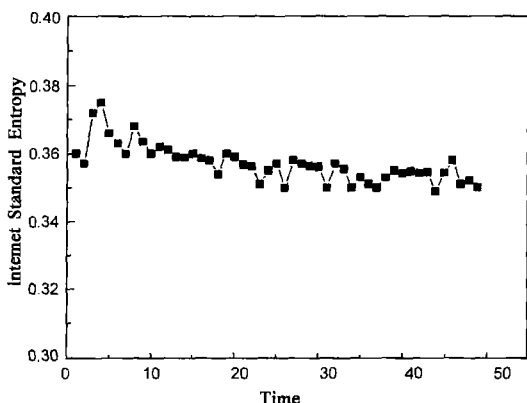
表 1 Internet网络标准结构熵

	1个月	10个月	20个月	30个月	40个月	49个月
riesling	0.3487	0.3706	0.354	0.34448	0.35019	0.34919
apan_jp	0.3604	0.3599	0.359	0.3561	0.3542	0.3507

从表 1可以看出, Internet网络标准结构熵在 2000 年 4月至 2004 年 5月期间有缓慢的降低,尽管降低的幅度不大,但是这种趋势是非常明显的. 为了对这个趋势有更直观的感受,可见图 3 (其中的横坐标 Time为相对于的月份值,纵坐标为 Internet的标准结构熵-Internet Standard Entropy.)



(a) 通过 riesling 节点数据获得的 H_s 与时间关系图



(b) 通过 apan_jp 节点数据获得的 H_s 与时间关系图

图 3 H_s 与时间关系图

由图 3(a)和(b)可知,随着 Internet的演化,其标准结构熵向着熵减少的方向发展,也就是, Internet的演化是向着结构有序的方向进行的,其结构所蕴涵的信息量在缓慢增加, Internet所蕴涵的巨大的信息量也是吸引各国、各地区竞相发展 Internet的主要动力,而恰恰由于人类的广泛关注,进一步推动了 Internet结构的有序化. Internet作为一个巨大的耗散结构系统,其标准结构熵是一种信息熵. 按照热力学第二定律的基本原理, Internet这种结构有序的演

化也正是其不断地与外界进行物质和能量的交换,使 Internet的标准结构熵逐渐减少的结果.

2.3 Internet拓扑演化的序参量

根据 Venkat等学者建立了复杂网络的进化适应模型^[14-16],可构造 Internet演化的效用函数为

$$P = \alpha_1 \phi_{STS} + \alpha_2 \phi_{LTS} \quad (6)$$

其中, P 为 Internet进化效用函数, ϕ_{STS} 为短期生存效用, ϕ_{LTS} 为长期生存效用; $\alpha_1 + \alpha_2 = 1$, $\alpha_1, \alpha_2 \in [0, 1]$, 分别为短期生存效用和长期生存效用的权重. 需要特别指出的是, α_1 和 α_2 的大小关系是动态变化的. Internet的演化过程就是保证这个函数值取得局部最小值.

长期生存效用和短期生存效用是 Internet演化过程中相互作用的两个方面. Internet的短期生存效用是由其效率决定的, Internet的效率越高,其短期生存效用就越小; Internet的长期生存效用由其鲁棒性决定, Internet的鲁棒性越好,其长期生存效用就越小. Internet演化的过程就是两者共同作用,使整个 Internet结构达到动态平衡的过程.

ϕ_{STS} 用 Internet的服务效率来表示,通常可以使用 Internet单位化的平均最短路径来计算.

$$\phi_{STS} = \frac{D}{D_{max}} \quad (7)$$

$$\text{其中 } D = \frac{\sum_{i,j} D_{ij}}{C_N^2}, D_{ij} \text{ 是指定点 } i \text{ 和 } j \text{ 之间的最短路径,}$$

D 为 Internet中节点的平均最短路径长度, $D_{max} = \frac{N+1}{3}$, N

为 Internet的节点总数. 从式 (7) 可以看出, Internet的平均最短路径越小, ϕ_{STS} 值越小, Internet的服务效率越高.

ϕ_{LTS} 用 Internet的鲁棒性来表示,可以用 Internet的平均集聚程度来计算. Internet的集聚程度是其网络集团化的程度,即考察连接在一起的集团各自的紧邻之中有多少是共同的近邻. 对于每一个顶点 d , 找到其近邻集合 N_d , 记 $n = |N_d|$, N_d 中存在的边的数量为

$$L = \sum_{l \in E: x, y \in N_d} \phi_l^x \phi_l^y$$

其中, l 为 Internet中所有边的集合 E 中的一条边. 则顶点 d 的集聚程度为

$$C_d = \frac{L}{C_n^2}$$

于是可得所有节点的平均集聚程度 C . 文中令

$$\phi_{LTS} = 1/C \quad (8)$$

可见,平均集聚程度越高,鲁棒性越好, ϕ_{LTS} 的值越小.

哈肯协同学的基本原理将自然界的有序结构分为两种类型. 一种是晶体 - 有序但无生命的结构,这是稳定的平衡态,是一种死的结构,它主要是靠分子或原子相互作用而引起系统元素具有固定的位置所形成的,一旦处于这种状态,晶体和周围环境就没有了物质和能量的交换,而且系统也没有了任何宏观过程. 而且物理学家认为它只有

在温度下降时才会出现有序态. 但生命过程在温度下降时, 相反会变得缓慢起来, 在温度很低时会完全停止, 很多生物甚至会死亡.

另一种是在远离平衡条件下, 通过与外界环境之间不停顿地进行物质和能量的交换所形成和维持的宏观范围内的时空序, 它是不可逆过程. 远离平衡条件下的有序结构不可能像在平衡条件下的有序结构那样总是有保证, 它有可能失去稳定性, 而涨落放大会使系统达到某种新的有序状态. 普利高津把这种需要能量耗散才能形成和维持的有序结构称为耗散结构. 显然, Internet 是一个复杂的巨型耗散结构系统. 参数 α_1 和 α_2 在本质上就是序参量.

与封闭系统不同, Internet 的演化毫无疑问要受到成本、能量、通讯链路、信息容量以及计算能力等因素的影响. 为了描述 Internet 的演化, 就不能把外力作为给定的常量, 而要把外力作为遵从运动方程的量. 用 F 表示外力 (原因), α 表示演化效用 (效果). 虽然在一般情况下, α 对时间的导数不是 α 的线性函数, 但可利用协同学的支配原理进行线性近似, 所以

$$\alpha' = -l\alpha + F \quad (9)$$

其中, l 为阻尼系数, 式 (9) 描述了 Internet 的有序度, 这就是称 α 为“序参量”的理由, 如果有 $\alpha_1, \alpha_2, \dots, \alpha_n$ 个变量, 可设它们满足如下方程:

$$\begin{aligned} \alpha_1' &= -l_1\alpha_1 + F_1 \\ \alpha_2' &= -l_2\alpha_2 + F_2 \\ &\vdots \\ \alpha_n' &= -l_n\alpha_n + F_n \end{aligned} \quad (10)$$

然后, 按照下述方式编排下标, 把这 n 个变量变成两类性质不同的组, 一组 $i = 1 \dots m$, 指具有小阻尼甚至可以变为不稳定变量 α_i , 即 $l \leq 0$ 而另一组 $j = m + 1 \dots n$, 是指稳定变量 α_j .

用支配原理将稳定变量消去, 把高维问题归结为低维问题, 支配原理表示为

$$\alpha = f_i(\alpha_c) \quad (11)$$

最后对留下的不稳定变量利用上式和协同学结论可得如下微分方程:

$$\alpha_i' = -l_i\alpha_i - \beta\alpha_i^n + F \quad (12)$$

其中 F 表示来自外部或内部的各种涨落力, 如果 F 可以忽略的话, 则上式变为 Bemoulli 方程, β 反映系统的动力学特征, l 为阻尼系数. 当 $l \rightarrow 0$ 时, 可以观察到所谓的临界慢化现象. 设在不考虑外力的条件下, 式 (12) 的解为 α_i , 则

$$\alpha_1 = 1 - \left| \frac{\alpha_i - F}{\alpha_i + F} \right|, \quad \alpha_2 = \left| \frac{\alpha_i - F}{\alpha_i + F} \right| \quad (13)$$

可以看出 α_1 随着 F 的增加而减少, 当 $F \rightarrow 0$ $\alpha_1 = 0$ 当 $F \rightarrow \infty$ 时, $\alpha_1 \rightarrow 1$. 可以看出, 随着外力的不断增加 (外界不断地给 Internet 注入能量), 整个 Internet 网络的演化效用函数由短期生存效应所决定. 而要使短期生存效应最小,

Internet 就必须达到效率最高, 其节点间的平均最短路径就

要最小. 根据式 (6) 可知, Internet 结构为星形时其短期生存效应最小. 根据定理 2 星形网络的标准结构熵最小, 按照热力学第二定律和耗散结构的理论, 要是系统向着有序方向发展, 必须由外界注入能量, 与本模型对应, 必须有外力 F 的不断作用, 也就是说必须有大量的人力、物力和财力等大量注入 Internet 才能使得 Internet 向着高度有序的方向发展; 反之, 随着外力的不断减少, Internet 网络的演化由长期生存效用决定, 要使其最小, Internet 的结构必须非常均匀. 如果 Internet 没有外力的作用的话, 那么可以将其看作封闭系统, 根据热力学第二定律可知, 封闭系统向着熵增的方向发展且不可逆, 而根据前面的定义, Internet 结构规则时其标准结构熵最大. 至此, 分析可知 Internet 网络标准结构熵和序参量所得到的结论是吻合的, 即 Internet 的拓扑正在向着更有序的方向演化.

3 Internet 节点平均连接度的分形

3.1 Internet 节点平均连接度的分形维数

Internet 是一个开放的复杂巨系统, 显然它应该是高维的, 但通常情况下, 人们往往只能得到 Internet 的一维信息, 即一组观测值, 或称一维标量时间序列. 很显然, 如何用这个一维信息准确地刻画高维的 Internet 是非常现实且重要的问题, 因此, 根据有限的数据来重构动力系统就是解决问题的关键. 动力系统重建也叫相空间重构, 即通过一维的时间序列反向构造出原系统的相空间结构. 通过相空间重构, 可以找到隐藏在混沌吸引子中的演化规律, 使现有的数据纳入某种可描述的框架下, 从而为 Internet 时间序列数据的研究提供了一种崭新的方法和思路.

从 CAIDA Skitter 项目获得的数据可得 2000 年 1 月至 2004 年 6 月间 Internet 路由节点平均连接度的一维时间序列如表 2 所示.

表 2 2000 年 1 月 - 2004 年 6 月 Internet 节点平均连接度 (序号为从 2000 年 1 月开始的月份相对值)

序号	平均连接度	序号	平均连接度	序号	平均连接度	序号	平均连接度
1	3 368	2	3 859	3	3 384	4	3 603
5	3 879	6	3 782	7	3 487	8	3 81
9	3 629	10	3 712	11	3 609	12	3 425
13	3 397	14	3 439	15	3 288	16	3 315
17	3 362	18	3 377	19	3 361	20	3 383
21	3 35	22	3 195	23	3 306	24	3 326
25	3 346	26	3 272	27	3 149	28	3 831
29	3 82	30	3 809	31	3 796	32	3 861
33	3 672	34	3 456	35	3 414	36	3 46
37	3 4	38	3 394	39	3 57	40	3 836
41	3 6	42	3 619	43	3 608	44	3 573
45	3 63	46	3 625	47	3 596	48	3 653
49	3 626	50	3 661	51	3 661	52	3 688
53	3 646	54	3 671				

利用重建复杂系统动力学原理,将 Internet 路由节点平均连接度时间序列 $\{x_i\}$, 表示为

$$x_0, x_1, \dots, x_i, \dots, x_n$$

并将其转化为

$$Y_n(m, \tau) = (x_n, x_{n+\tau}, \dots, x_{n+(m-1)\tau}),$$

式中, m 为嵌入维数, τ 为时间延迟, 在 m 维相空间 R^m 中得到 m 维向量 Y_n , 定义时间序列的关联积分为

$$C_m(r) = \frac{2}{N'(N'-1)} V \quad (14)$$

$N' = N - (m-1)\tau$, 表示向量序列的有效长度, 即重构之后相空间矢量的长度; V 是两相点之间距离小于 r 的个数, 可以表示为

$$V = \sum_{i=1}^{N'} \sum_{j=i+1}^{N'} H(r - \|Y_i - Y_j\|) \quad (15)$$

式中, $H(X)$ 为 Heaviside 函数, 即当 $X < 0$ 时, $H(X) = 0$; 当 $X \geq 0$ 时, $H(X) = 1$.

根据重建复杂系统动力学原理, 当 r 充分小时, 关联积分与 r^D 成比例关系, 这里的指数 D 就是混沌吸引子的关联维数. 对关联积分和 r^D 分别取对数, 则 $\ln C_m(r)$ 对 $\ln r$ 的斜率就是 D . 因此只要作出 $\ln C_m(r)$ 和 $\ln r$ 双对数图, 并对其进行线性回归, 所得到的线性方程的斜率就是关联维数, 即

$$\ln C_m(r) = D \ln r + \text{const}(\text{常数})$$

利用分形维的 $G-P$ 算法, 运用 matlab 编程并绘制 $\ln C_m(r) - \ln r$ 曲线, 如图 4 所示. 图 4 为 $\tau = 3, m = 2, 3, 4, 5, 6, 7, 8$ 时的 $\ln C_m(r) - \ln r$ 关系图, 由图可见, 图 4 中存在一个广阔的线性区, 当 $m = 6$ 时, 通过对 $\ln C_m(r)$ 和 $\ln r$ 数据的线性回归, 得 $D = 2.93$, 这就是 Internet 的分形维数. 由此可知, 只需 3 个变量便可以用决定论的方法来建立 Internet 路由级节点拓扑的数学模型.

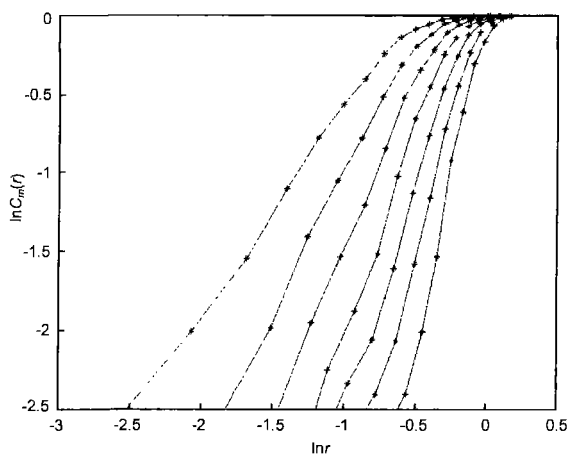


图 4 Internet 平均连接度 $\ln C_m(r) - \ln r$ 对数图

根据关联维数与运动特性的关系可知, 对于随机运动, D 值将随着 m 值线性增加, 即 $D = m$, 该值并不会达到一饱和值, 而会趋于无限大. 当系统为规则的周期运动时

D 为整数值, 当 D 为分数时即表示系统为混沌. 根据上文的计算可知, Internet 的平均连接度为分数维, 故其必是一混沌系统.

3.2 最大 Lyapunov 指数

混沌系统的一个重要的特性就是对于初始条件的敏感依赖. Internet 对于初始条件的敏感依赖的程度可用 Lyapunov 指数^[17]来度量. 它表示重构相空间中相邻两点随时间演化的发散程度. 一个负的 Lyapunov 指数度量收缩——Internet 在受到扰动后需要多长时间才能恢复自己; 一个正的 Lyapunov 指数度量空间中的伸展, 也就是度量邻近的点相互之间发散得有多快. Lyapunov 指数是一个谱系 $\lambda_i (i = 1, 2, 3, \dots, m)$, 其中 $\lambda_1 > \lambda_2 > \dots > \lambda_m$, 若谱系中最大的 $\lambda_1 > 0$, 则系统必是混沌的.

计算最大 Lyapunov 指数的算法如下:

(1) 采用延迟向量法对 Internet 节点平均连接度时间序列进行相空间重构.

(2) 选取两个相距至少一个平均轨道周期的点. 在一个固定的时间间隔之后, 度量两点之间的距离, 用此距离与开始时两点的距离的比值代表轨道的发散程度.

(3) 如果距离变得太大, 超出了所允许的范围, 则搜索一个与原来的那个定向角度相似的替代点, 新的这对点的定向应该尽可能地与原来那对点保持一致, 即这对新点与原来那对点的夹角尽可能的小.

(4) 再求出经过下一个时间间隔后, 得到的两点距离与前边两点距离的比值. 如此继续下去, 穷尽时间序列, 就得到了最大的 Lyapunov 指数.

下面以 2000 年 1 月至 2004 年 6 月的 Internet 节点月平均连接度的平均值为样本, 计算 Internet 节点平均连接度的最大 Lyapunov 指数, 结果如表 3 所示.

由表 3 可知, 最大 Lyapunov 指数随着嵌入维数的增加, 逐渐稳定在 0.34, 故 Internet 节点的平均连接度的时间序列的最大 Lyapunov 指数为 0.34, 并且该最大 Lyapunov 指数大于零, 说明 Internet 节点平均连接度的时间序列为混沌时间序列.

表 3 最大 Lyapunov 指数

嵌入维数	3	4	5	6	7	8	9	10
最大 Lyapunov	0.7622	0.5482	0.4112	0.3611	0.3443	0.3403	0.34	0.34

3.3 Kolmogorov 熵

可以用柯尔莫戈诺夫熵 (Kolmogorov) 描述 Internet 的混沌特性及其程度. 柯尔莫戈诺夫熵又称测度熵, 它给出了轨道在单位时间内产生的平均信息量的一个上限. 当测度熵 K 等于零时, 系统是有序的, 表示可能完全准确预测; K 为无穷大时, 系统是随机的, 表示可能统计平均值预测; K 介于零和无穷大之间, 即当 K 为一确定实数时, 表示系统是混沌系统, 并且系统的混沌程度随 K 值的增大而增大, 这时可以作短期预测.

但测度熵难于计算, 本文用二阶 Renyi 熵 K_2 来代替. 一般情况下, K_2 与测度熵具有类似的性质, 可以将 K_2 作为 K 的一个恰当估计, 而且 K_2 并且易于从时间序列计算. 根据 Grassberger 和 Procaccia 提出的算法, 对离散时间序列, 固定延迟时间 τ , K_2 可表示成与前面定义的关联积分的极限, 即

$$K_2 = - \lim_{\tau \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{M \rightarrow \infty} \frac{1}{M} \log_2 C_m(\tau) \quad (16)$$

对以上的极限求解进行一定的处理, 对离散时间序列, 固定延迟时间 τ 实际计算中 m 不可能取无穷, 一般能保证数据计算结果比较稳定即可, τ 取值在 0.001 ~ 0.05 之间, 这也是与求解关联维数时相一致的.

因 $C(\tau) \propto \tau^D$, 故对其进行一定的数学处理得

$$\log_2 C_m(\tau) = D \log_2 \tau - K_2 m \tau \quad (17)$$

在嵌入维数分别为 m 和 $m+n$ 的情况下,

$$\log_2 C_m(\tau) = D \log_2 \tau - K_2 m \tau \quad (18)$$

$\log_2 C_{m+n}(\tau) = D \log_2 \tau - K_2 (m+n) \tau \quad (19)$

式 (18) 和式 (19) 中, 分别用 $C_m(\tau)$ 和 $C_{m+n}(\tau)$ 表示嵌入维数为 m 和 $m+n$ 下的关联积分.

将上述计算分形维时的线性回归方程改写为

$$y_m = ax_m + b_m \quad (20)$$

式中 m 为嵌入维数, $a = D$ (即 a 等于分形维数), 在嵌入维数分别为 m 和 $m+n$ 的情况下, a 值假定相差不大, 将两式相减得

$$K_2 = \frac{\Delta b}{n \tau} \quad (21)$$

式中 $\Delta b_m = b_{m+n} - b_m$, 这就是测度熵的计算表达式.

对 Internet 节点平均连接度时间序列, 取时滞为 3, 嵌入维数 2~11, $n=1$ 得到的图形如图 5 所示. 从图 5 可知, Internet 节点平均连接度的 Renyi 熵 K_2 稳定在 0.2~0.25 之间, 且 $K_2 > 0$

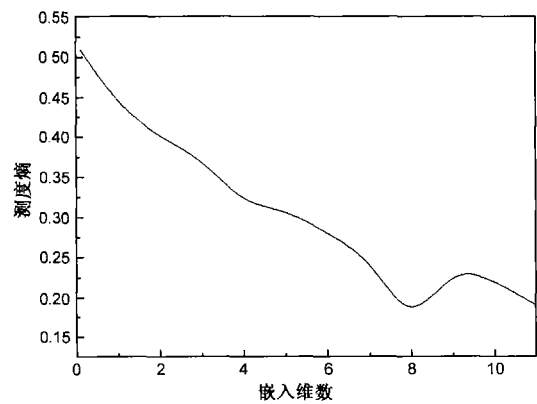


图 5 Internet 节点连接度测度熵计算曲线

根据式 (16), Kolmogorov 熵实际就是信息的平均损失率. 由于信息损失率是描述系统不确定性的程度, 当信息的损失率越大, 表示系统不确定程度就越大. 所以对于周期运动而言, 由于其运动是规则的, 因此它是完全可以预

测的, 信息是不随时间发生变化, 故 $K=0$. 对于随机运动而言, 由于其状态是完全不可预测的, 故 $K \rightarrow \infty$. 当 $0 < K < \infty$ 时, 则表示为一混沌运动, K 越大表示混沌程度越大, 系统更复杂. 根据上面的计算可知, 可以对 Internet 的平均连接度做短期的预测.

3.4 非线性混沌预测

根据上面的分析可知, Internet 节点的平均连接度时间序列具有混沌的特征, 根据最大 Lyapunov 指数的计算结果, 得出对 Internet 节点平均连接度进行精确预测的最大时间长度. 因此, 下文将采用非线性混沌预测方法对 Internet 节点的平均连接度进行短期预测.

Internet 节点平均连接度时间序列为 $\{x_i\}$, 其中 x_i 是第 i 个月的 Internet 节点平均连接度, $i=1, 2, \dots, n$, 对该时间序列作预处理, 即

$$S_i = \frac{x_{i+1}}{x_i} \quad (i=1, 2, \dots, n-1) \quad (22)$$

处理结果如表 4 所示.

对时间序列 $\{S_i\}$, 进行非线性回归, 不失一般性, 可以设非线性自回归模型为

$$S_{i+1} = aS_i^2 - bS_i + c \quad (a \neq 0)$$

将 $S_i = dX_i + f$ 代入上式, 得

$$X_{i+1} = AX_i^2 - BX_i + C \quad (23)$$

其中

$$A = ad, \quad B = b - 2af, \quad C = \frac{af^2 - (b+1)f + c}{d}$$

如果 $C=0$ 即

$$af^2 - (b+1)f + c = 0$$

若要上式成立, 则 $\Delta = (b+1)^2 - 4ac \geq 0$ 而 $f = \frac{(b+1) + \sqrt{\Delta}}{2a}$, 此时式 (23) 可改写为

$$X_{i+1} = AX_i^2 - BX_i$$

表 4 预处理后的 Internet 节点平均连接度

序号	预处理后的连接度	序号	预处理后的连接度	序号	预处理后的连接度	序号	预处理后的连接度
1	1.1146	2	0.8769	3	1.0647	4	1.0766
5	0.975	6	0.922	7	1.093	8	0.9525
9	1.0229	10	0.9723	11	0.949	12	0.9918
13	1.0124	14	0.956	15	1.0082	16	1.0142
17	1.0045	18	0.995	19	1.006	20	0.9911
21	0.9537	22	1.0347	23	1.006	24	1.006
25	0.977	26	0.962	27	1.2166	28	0.9971
29	0.9971	30	0.9966	31	1.017	32	0.951
33	0.9411	34	0.9878	35	1.0135	36	0.9826
37	0.9982	38	1.0519	39	1.0745	40	0.9385
41	1.0053	42	0.997	43	0.9903	44	1.016
45	0.9986	46	0.992	47	1.0159	48	0.9926
49	1.01	50	1	51	1.0074	52	0.9886
53	1.0065						

令 $X_i = \frac{B}{A} Y_i$, $K = -B$, 则

$$Y_{i+1} = KY_i (1 - Y_i) \quad (24)$$

式(24)是著名的逻辑斯特方程. 令 $f(Y_i, K) = KY_i (1 - Y_i)$, 则一次映射的不动点 Y_F 满足方程

$$f(Y_F, K) - Y_F = KY_F (1 - Y_F) - Y_F = 0 \quad (25)$$

根据稳定性和分岔理论, 在 $0 \leq K \leq 1$ 时, 有一个稳定不动点 $Y_F = 0$ 当 $1 < K \leq 3$ 时, $Y_F = 0$ 不稳定, 而 $Y_F = 1 - \frac{1}{K}$ 稳定; 当 $3 < K < 1 + \sqrt{6}$ 时, $Y_F = 0$ 和 $Y_F = 1 - \frac{1}{K}$ 都不稳定, 其二次迭代解稳定; 当 $K = 1 + \sqrt{6}$ 时, 分岔出一对周期 4 解, 而后再分岔出周期 8 解、16 解等等.

对表 2 中数据进行非线性回归, 条件是使得拟合曲线与表中的数据误差最小, 可得拟合曲线方程为,

$$S_{i+1} = 1.0658S_i^2 - 2.5124S_i + 2.445 \quad (26)$$

则

$$\Delta = (b+1)^2 - 4ac = 1.9134 > 0$$

$$f = \frac{(b+1) + \sqrt{\Delta}}{2a} = 2.2967$$

$$B = b - 2af = -2.3832$$

$$K = -B = 2.3832$$

因为 $1 < K < 3$ 则时间序列 $\{Y_i\}$ 收敛于 $\frac{K-1}{K} = 0.5804$

时间序列 $\{x_i\}$ 收敛于 $\frac{B}{A} \cdot \frac{K-1}{K}$, 时间序列 $\{S_i\}$ 收敛于

$$d \cdot \frac{B}{A} \cdot \frac{K-1}{K} + f = \frac{B}{a} \cdot \frac{K-1}{K} + f = 0.9989$$

利用上面的分析和计算结果, 对 2004 年 7 月-12 月 Internet 节点月平均连接度值进行预测, 预测结果如表 5 所示, 实测值如表 6 所示.

由表 5 和表 6 可知, 2004 年 7 月和 8 月 Internet 节点平均连接度的预测值的误差为 0.3% 和 0.33%. 时间越长, 预测的误差越大, 11 月和 12 月的误差分别为 0.84% 和 1.4%. 这虽然说明预测存在着累计误差, 但也说明该动力系统只能作短期预测.

表 5 2004 年 7 月 - 2004 年 12 月 Internet 节点平均连接度预测值

序号	预测的平均连接度	序号	预测的平均连接度	序号	预测的平均连接度	序号	预测的平均连接度
55	3.667	56	3.663	57	3.659	58	3.655
59	3.651	60	3.647				

表 6 2004 年 7 月 - 2004 年 12 月 Internet 节点平均连接度测量值

序号	平均连接度	序号	平均连接度	序号	平均连接度	序号	平均连接度
55	3.657	56	3.651	57	3.67	58	3.674
59	3.684	60	3.701				

另外, 分别采用最大 Lyapunov 指数法和局域法分别对 Internet 节点的平均连接度进行预测, 预测结果如图 6 所示. 分析图 6 可知, 在短时间内, 最大 Lyapunov 指数法和本文所采用的非线性混沌预测法所获得的预测值都非常接近

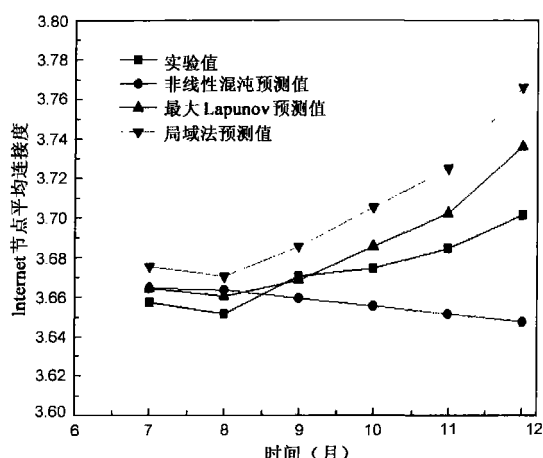


图 6 预测值和实测值关系图

近于实测值, 而局域法预测的误差相对较大. 随着时间的增长, 这三种方法预测值的精度都有下降的趋势, 最大 Lyapunov 指数法预测精度下降相对较慢, 而非线性混沌预测法仍好于局域法. 但是, 局域法比较简单, 计算量较少, 容易实现, 而最大 Lyapunov 指数法虽然预测的精度较高, 但算法比较复杂. 本文所采用的非线性混沌预测法则既有很高的短期预测精度, 又容易实现, 而且充分考虑到了节点平均连接度的混沌特性, 因此, 它是进行短期预测的比较好的方法.

4 结论

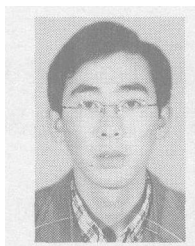
目前对于复杂网络拓扑演化的研究已经取得了众多的研究成果, 但这些主要侧重于对网络规模增长时, 节点之间的动力行为的研究, 如节点之间的同步行为、节点的偏好依附属性等^[18-21], 而本文从 Internet 拓扑的宏观角度, 通过计算其 Internet 标准结构熵和影响其演化的序参量, 发现 Internet 拓扑结构的演化是从低度有序向高度有序的方向进行的. 另外香港城市大学的陈关荣教授及其合作者对于复杂网络从非混沌到混沌转化问题的研究, 虽然与本文的研究方法和思路不同, 但研究的结论却是吻合的. 即本文通过计算 Internet 节点平均连接度的分形维数、最大 Lyapunov 指数和 Kolmogorov 熵, 发现 Internet 节点的平均连接度的混沌特征和不可长期预测性. 最后本文利用 Internet 节点平均连接度时间序列, 使用非线性混沌预测方法, 对 Internet 节点的平均连接度进行了预测, 并与常见的混沌预测方法进行了比较, 结果表明, 短期的预测结果效果很好, 长期的预测效果较差.

参考文献:

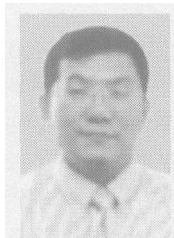
- [1] Lu J, Chen G R. A time-varying complex dynamical network model and its controlled synchronization criterion[J]. IEEE Trans on Auto Contr 2005 50(6): 841-846
- [2] Li X, Chen G R, Ko K T. Transition to chaos in complex

- dynamical networks[J]. *Physical A*, 2004 338(7): 367–378
- [3] Fan Z P, Chen G R. Evolving networks driven by node dynamics[J]. *Int J Modern Physics* 2004 14(7): 2540–2546
- [4] 张琪昌, 王洪礼等. 分岔与混沌理论及应用 [M]. 天津: 天津大学出版社, 2005. 256–280
- [5] 于景元. 钱学森关于开放的复杂巨系统的研究 [J]. *系统工程理论与实践*, 1992 (12) 5: 8–12
- [6] 钱学森, 于景元, 戴汝为. 一个科学新领域——开放的复杂巨系统及其方法论 [J]. *自然杂志*, 1990 (1): 3–10
- [7] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. *Reviews of modern physics* 2002 74: 47–97.
- [8] Jeong H, Tombor B, Albert R, et al The large-scale organization of metabolic networks[J]. *Nature* 2000 407: 651–653
- [9] Valverde S, Ferrer-Cancho R, Sole R V, Ricard V. Scale-free networks from optimal design [J]. *Europhys Lett* 2002 60(4): 512–517
- [10] Faloutsos M, Faloutsos P, Faloutsos C. On power law relationships of the Internet topology [J]. *Am Sigcomm Computer Communication Review*, 1999 29(4): 251–262
- [11] 谭跃进, 吴俊. 网络结构熵及其在非标度网络中的应用 [J]. *系统工程理论与实践*, 2004 24(6): 1–3
- [12] 张军, 林汉同, 熊国庆等. 分形理论及其应用 [M]. 成都: 四川大学出版社, 1989. 42–46
- [13] Cover T M, Thomas J A. Elements of Information Theory [M]. New York: John Wiley & Sons 1991. 1–22
- [14] Valverde S, Cancho R F, Sole R V. Scale-free networks from optimal design[J]. *Euro phys Lett* 2002 60: 512–517
- [15] Ferrer R, Sole R V. Optimization in complex networks [J]. *Lecture Notes in Physics* 2003 625: 114–126
- [16] Venkatasubramanian V, Katare S, Parker P R, et al Spontaneous emergence of complex optimal networks through evolutionary adaptation[J]. *Computers and Chemical Engineering*, 2004, 28(9): 1789–1798
- [17] 吕金虎, 陆君安, 陈士安. 混沌时间序列分析及其应用 [M]. 湖北: 武汉大学出版社, 2002. 114–135
- [18] Eisenberg E, Levanon E Y. Preferential attachment in the protein network evolution[J]. *Physical Review Letters* 2003 91(13): 138701–138704
- [19] Dorogovtsev S N, Mendes J. Evolution of networks[J]. *Adv Phys* 2002 1079–1187
- [20] Venkatasubramanian V, Katare S, Parker P R, et al Spontaneous emergence of complex optimal networks through evolutionary adaptation[J]. *Computers and Chemical Engineering* 2004, 28(9): 1789–1798
- [21] Barabasi A L, Jeong H, Neda Z. Evolution of the social network of scientific collaborations[J]. *Physical A*, 2002 311: 590–614

作者简介:



张文波 男, 1973年出生于黑龙江伊春, 沈阳理工大学副教授, 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, 嵌入式系统. E-mail: zwk@neuxa.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, 数据融合.

孙佩刚 男, 1976年出生于安徽明光, 东北大学信息科学与工程学院博士研究生, 沈阳炮兵学院教师, 主要研究方向为嵌入式系统、复杂网络.

徐野 男, 1976年出生于辽宁沈阳, 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系统, 人工智能.

张昕 男, 1979年出生于辽宁沈阳, 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系统, 系统仿真.