

# Internet 分形特征研究

李鹤群, 徐久强, 王进法, 赵海  
(东北大学 计算机科学与工程学院, 辽宁 沈阳 110169)

**摘 要:** 采用  $k$ -核分解算法对 Internet 分形特征进行研究. 对高核区间和低核区间拟合结果的差异进行分析, 提出 Internet 分形特征与其结构的层次性存在关联的猜想. 随后, 从删边数、删边比例和子网分布三个角度对  $k$ -核分解过程进行观察, 确定 Internet 核心层与边缘层的分割边界, 并利用骨架树理论研究 Internet 的层次分形特征. 最后, 从度分布指数、同配系数和结构熵等常见统计角度对 Internet 部分与整体, 以及部分与部分之间的关系进行观察. 观察表明, 在以上统计角度, Internet 边缘子网可以表征网络整体.

**关 键 词:** 复杂网络; Internet; 分形;  $k$ -核分解; 层次分形

中图分类号: TP 393.4      文献标志码: A      文章编号: 1005-3026(2016)12-1691-05

## Research on Fractal Property of Internet

LI He-qun, XU Jiu-qiang, WANG Jin-fa, ZHAO Hai  
(School of Computer Science & Engineering, Northeastern University, Shenyang 110169, China. Corresponding author: LI He-qun, E-mail: lihequn@stumail.neu.edu.cn)

**Abstract:** The  $k$ -core decomposition algorithm was applied to the study of the Internet fractal property. During the study, the difference of fitting results between high-core region and low-core region was analyzed, and a conjecture was proposed that Internet fractal is related to its hierarchical structure. Subsequently, the  $k$ -core decomposition process was observed from three aspects: the number and ratio of deleted edges, and subnets distribution. A partition boundary was found which was useful to distinguish core and periphery, then the Internet fractal hierarchy was studied with the help of the skeleton theory. Finally, relations between part and whole, part and part were observed from degree distribution exponent, assortativity coefficient, and entropy. The observations show that the subnets in the Internet edge can characterize the network from the statistics above.

**Key words:** complex networks; Internet; fractal;  $k$ -core decomposition; fractal hierarchy

Internet 作为信息传播的主要载体, 在政治、经济、教育乃至日常生活等方面发挥着重要作用. 近年来, Internet 的安全性、鲁棒性、可控性和可管性等问题正逐渐受到人们的关注. 而以上问题的解决, 都离不开对其拓扑结构的研究.

当前, 在 Internet 拓扑分析方面已经取得了一些成果<sup>[1-3]</sup>, 这些工作主要集中在网络拓扑的全局层面上. 面对规模呈指数型增长的 Internet, 单纯从全局角度进行研究往往是事倍功半的. 为了迎合大规模拓扑环境下的新需求, 寻找新的研究角度成为当务之急. 而将分形理论应用到

Internet 拓扑研究, 成为一种新的尝试<sup>[4]</sup>.

分形简单来说就是“组成部分以某种方式与整体相似的图形”, 图 1 就是一个分形. 在图 1 中, 不同观测尺度(矩形框)下观测到的部分子集, 其结构都与整体保持一致, 这种部分与整体的一致性被称作自相似性.

如式(1)所示, 在分形理论中, 若覆盖研究对象所需最小盒子数  $N_B$  和盒子的尺度  $l_B$  具有幂律关系, 则说明其具有分形特征. 其中  $d_B$  为分形维数,  $d_B$  越大, 研究对象越复杂, 其内部规律性越不明显.

$$N_B(l_B) \sim l_B^{-d_B}. \tag{1}$$

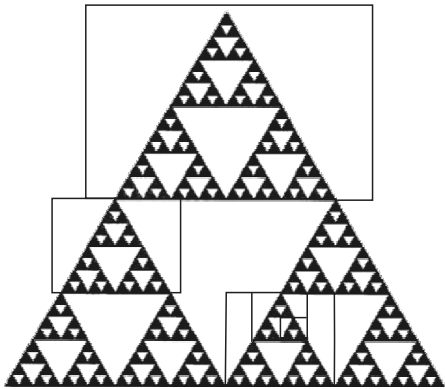


图 1 谢宾斯基三角形  
Fig. 1 Sierpinski triangle

分形理论打破了整体与部分之间的隔膜,找到了由部分过渡到整体的媒介——自相似性,可以有效降低研究对象的复杂性. 2005 年, Song 等<sup>[5]</sup>提出了广义盒子计数法,第一次将分形理论应用到网络对象拓扑分析中. 随后, Song 等<sup>[6]</sup>又提出了燃烧算法, Schneider 等<sup>[7]</sup>提升了算法的效率, Kim 等<sup>[8]</sup>提出了随机序列算法, Gao 等<sup>[9]</sup>提出了球覆盖算法.

以上算法均为普适性方法,研究过程中所有节点的地位都是一致的,并不适宜研究 Internet 这种核心层和边缘层节点存在较大差异的网络. 为此本文采用  $k$ -核分解算法<sup>[4]</sup>对 Internet 的分形特征进行了研究,该方法可以更准确地反映 Internet 内部的结构特征,并有助于研究其结构的规律性. 在研究过程中,本文提出 Internet 具有层次分形特征的猜想,并在验证猜想的同时,从统计角度观察了 Internet 部分与整体、部分与部分之间的关系.

1 基于  $k$ -核分解的分形研究方法

1.1  $k$ -核分解算法

$k$ -核分解算法<sup>[4]</sup>源于求解  $k$ -核,网络的  $k$ -核是指反复去掉度值小于  $k$  的节点后,所余的最大子图.  $k$ -核分解算法选择节点核数  $k_s$  作为度量尺度  $l_B$ ,选择  $k_s$ -核的节点数  $N_s(k_s)$  作为  $N_B$ ,如式(2)所示,若  $k_s$  与  $N_s(k_s)$  服从幂律,则认为研究对象具有分形特征.

$$N_s(k_s) \sim k_s^{-d_s} \tag{2}$$

其中  $d_s$  为分形维数,由式(2)等价有

$$d_s \approx -\frac{\ln N_s(k_s)}{\ln k_s} \tag{3}$$

即可以通过统计不同核数下的节点数,并在双对数坐标系下进行直线拟合,通过计算拟合直线斜

率绝对值的方式获得分形维数,如图 2 和图 3 所示.

$k$ -核分解过程逐步删除了网络中小度值节点,换个角度来看,可以认为所有度值小的节点都被与其连通的大度值节点所吸收,求解  $k$ -核的过程实际上是研究 Internet 更深层结构的过程.

1.2 问题分析

图 2 和图 3 分别统计了  $k_2$  至  $k_{15}$  和  $k_{15}$  至  $k_{26}$  两个区间上的数据,并对统计结果在双对数坐标系下进行了线性拟合. 实验数据采用 CAIDA Archipelago 项目 2014 年 12 月第 5 个探测周期的探测结果,实验数据包含 84.5 万个节点、219.5 万条边.

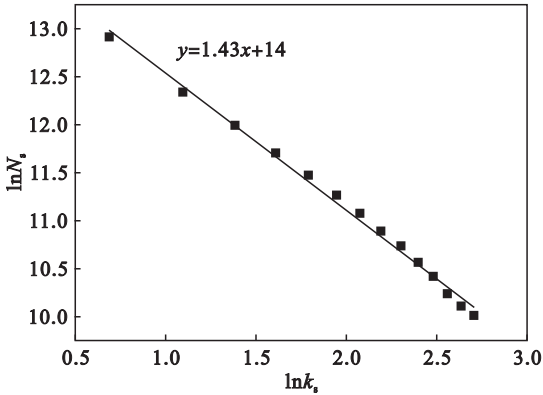


图 2  $k_2$  至  $k_{15}$  线性拟合  
Fig. 2 Linear fitting from  $k_2$  to  $k_{15}$

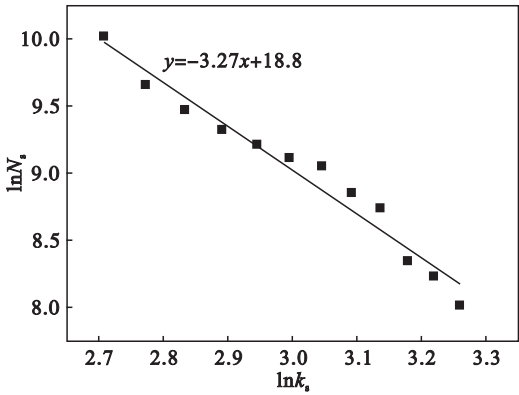


图 3  $k_{15}$  至  $k_{26}$  线性拟合  
Fig. 3 Linear fitting from  $k_{15}$  to  $k_{26}$

观察拟合直线的斜率可以发现,在  $k$ -核分解过程中,随着核数的升高,其前后两个子集对应节点数的下降速率有着显著不同. 对应的分形维数分别为 1.43 和 3.27,后者是前者的 2.28 倍,即整个  $k$ -核分解过程并不具有一致性. 针对这一现象,有学者认为:“任意一个实际存在的复杂系统,都仅存在有限层次的分形特征,当研究层次深入到一定程度时,研究对象将不再表现出分形特征.”以上解释为针对网络对象的普适性结论,结

合 Internet 自身结构特征,以下解释更为合理.

图 4 给出了 Internet 结构示意图<sup>[10]</sup>,由图可知,Internet 具有三级层次结构,分别是核心骨干网(简记为核心层)、边缘 ISP(简记为边缘层)和用户终端. Internet 的边缘层和核心层各自具有统一网络结构,核心层结构类似于全局耦合网络,节点度值较大;边缘层呈类树形结构,节点度值相对较小. 在图 2 对应的  $k$ -核分解过程中,由于核数较低,核心层节点所占比例较小,因此主要反映 Internet 的边缘层特征. 随着核数的升高,整个过程逐渐逼近网络核心,因而图 3 反映的是 Internet 核心层特征. 而图 2 和图 3 的差异恰恰说明了 Internet 核心层和边缘层之间的差异.

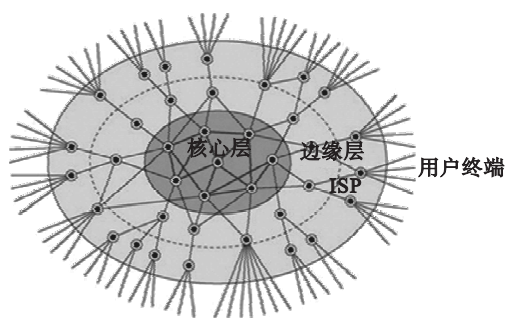


图 4 Internet 结构示意图  
Fig. 4 Diagram of the Internet structure

经过以上实验和分析,本文猜测,Internet 可能具有层次分形特征,其核心层和边缘层具有完全不同的结构,具有不同的分形维数.

## 2 面向分形特征的拓扑分割

为了验证 1.2 节猜想,需要区分 Internet 的核心层和边缘层,再分别进行研究. 鉴于  $k$ -核分解过程恰好是从 Internet 边缘层向核心层逼近的过程,因此可以从该过程入手,寻找合适的  $k_s$  作为分割边界.

### 2.1 基于“核心-边缘”的 Internet 拓扑分割

$k$ -核分解通过删边,由边缘层逼近核心层,因此不妨观察  $k$ -核分解过程中的删边情况. 图 5 中纵坐标为删边数  $N$ ,实验数据同前.

如图 5 所示,随着层次的深入,删边数呈下降趋势. 当  $k_s$  较小时,边缘层节点在网络中占主体地位,该层主要为类树形结构,存在大量小度值节点,因而在图像前半部分有大量的边被删除. 在图像后半部分,从  $k_s = 11$  开始,删边数开始呈现出

不规则趋势. 本文认为这种不规则性恰好是对 Internet 核心层结构性质的反映,其原因在于

Internet 核心层除了平均度较高外,并不具有其他性质. 由此猜测,核心层和边缘层的边界可能在  $k_s = 11$  左右出现.

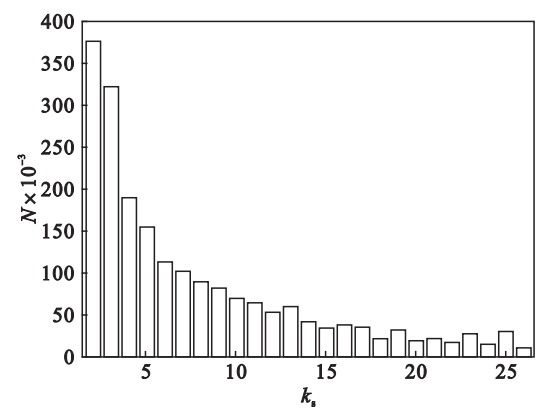


图 5  $k$ -核分解删边统计  
Fig. 5 Statistics of deleted edges during  $k$ -core decomposition

图 6 给出了  $k$ -核分解过程中删边数占当前总体的比例,实验数据同前文. 图 6 可以在  $k_s = 11$  处分为特征鲜明的两个部分. 在前半部分,随着  $k_s$  上升,对应比例呈下降趋势. 即随着层次的深入,需要删除的节点逐渐减少;在后半部分,随着  $k_s$  升高,对应比例呈现出

不规则性,即每次需删除的节点数不具有规律性. 由此可知, $k$ -核分解在前半部分主要是对边缘层进行研究,而后半部分则是观测核心层性质,以上结论都与图 5 的观测结果保持一致.

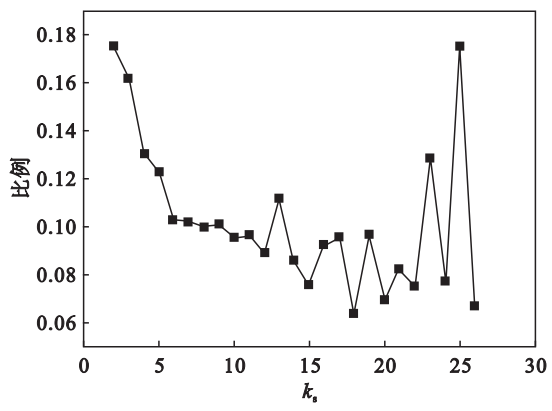


图 6  $k$ -核分解删边比例  
Fig. 6 Ratio of deleted edges during  $k$ -core decomposition

图 7 为网络总体去掉  $k$ -核分解生成的  $k_s$ -核后,生成的子网数. 图中横坐标为核数  $k_s$ ,纵坐标为子网个数,实验数据同前文. 由图 4 可知,Internet 核心层节点相互连接,而边缘层往往仅同其上层和下层节点相连,因此在去掉网络核心后,会生成大量相互独立的边缘子网,子网数与去掉

的  $k_s$  - 核直接相关. 在图 7 中, 当  $k_s$  取较大值时 ( $k_s \in [15, 26]$ ), 子网数相对稳定. 这说明有一部分核心层节点没有被删除, 正是这部分核心层节点保证了剩余部分的连通性, 使得子网数稳定在一个较低值. 由此可知, 分割边界在  $k_s = 15$  前出现. 图中  $k_s = 2$  似乎是一个异常点, 结合  $k$  - 核定义可以发现, Internet 的 2 - 核涵盖了网络中的大部分节点, 因此在该点子网数很少.

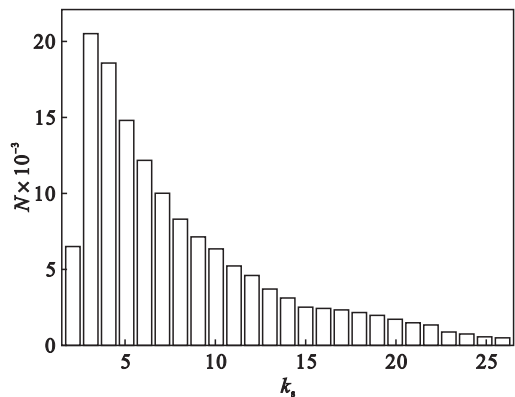


图 7  $k$  - 核分解子网分布

Fig. 7 Subnets distribution during  $k$ -core decomposition

由以上分析可知, Internet 核心层和边缘层的分割边界在  $k_s = 11$  左右, 即  $k_s \in [10, 12]$ , 并在  $k_s = 15$  之前出现. 为了简化计算, 本文选择  $k_s = 11$  作为分割边界, 进行后续实验.

2.2 分割合理性验证

2.1 节根据观察结果用 Internet 的 11 - 核表征 Internet 核心层, 用余下边缘子网表征 Internet 边缘层. 下面研究分割合理性.

在研究过程中, 本文借用了 Kim 等<sup>[11]</sup>有关骨架树的相关研究成果. Kim 等将网络的介数作为对应边的权值, 将原网络转化为一个加权无向图, 并计算其权值和最小的生成树, 求得的生成树即为该网络的骨架树. 研究发现, 一些分形网络的分形维数与其骨架树的分形维数相同, 而另一些分形网络则不具备该性质, 该现象同分形网络结构的差异性有关.

图 8 和图 9 分别对 Internet 核心层、边缘层最大子网和对应骨架树的分形维数进行了计算. 在计算过程中, 由于骨架树结构的特殊性, 本文选择另一种常见算法——随机序列算法<sup>[8]</sup>计算分形维数, 本次实验中采用的实验数据同前文.

由图 8 和图 9 可知, Internet 核心层和边缘层都具有分形特征, 且两者的分形维数具有较大差异, 这说明两者的复杂程度和规律性并不相同. 此外, Internet 核心层与其骨架树的分形维数不具有

一致性, Internet 边缘子网与其骨架树的分形维数具有一致性, Internet 的核心层和边缘层为两种完全不同的结构.

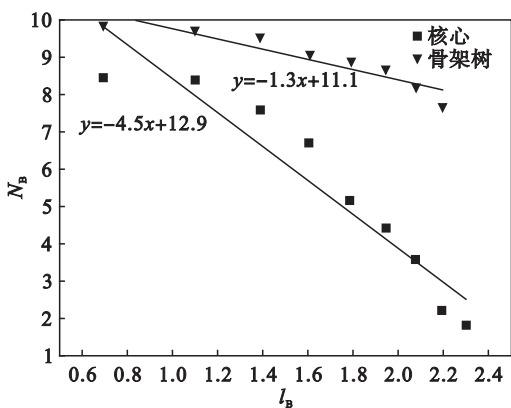


图 8 Internet 核心及其骨架树分形维数

Fig. 8 Fractal dimensions of the Internet core and its skeleton

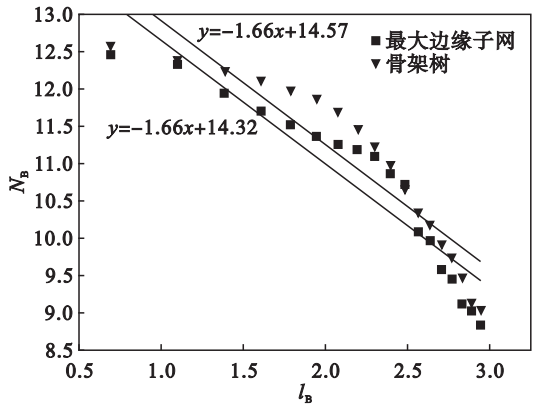


图 9 Internet 边缘子网及其骨架树分形维数

Fig. 9 Fractal dimensions of the Internet periphery and its skeleton

由以上分析可知, 猜想成立, Internet 的核心层和边缘层各自具有分形特征, 但两者的复杂程度、内部的规律性和组织结构均存在较大差异, 即 Internet 具有层次分形特征.

3 基于“核心 - 边缘”分割的分形研究

从度分布指数、同配系数和结构熵三个统计角度对 Internet 的层次分形特征进行了观察. 实验采用 CAIDA Archipelago 项目 2010 年至 2014 年每年 6 月和 12 月第 5 次探测的探测数据.

图 10 为 Internet 整体、核心、最大边缘子网 (子网 A) 和随机子网 (子网 B) 在 10 个时间点上的度分布指数. 其中子网 B 在规模最大的 100 个边缘子网中随机选择. 观察发现, 在度分布指数方面, Internet 核心和边缘子网均可表征整体性质. 此外, Internet 核心和边缘子网可以相互表征.

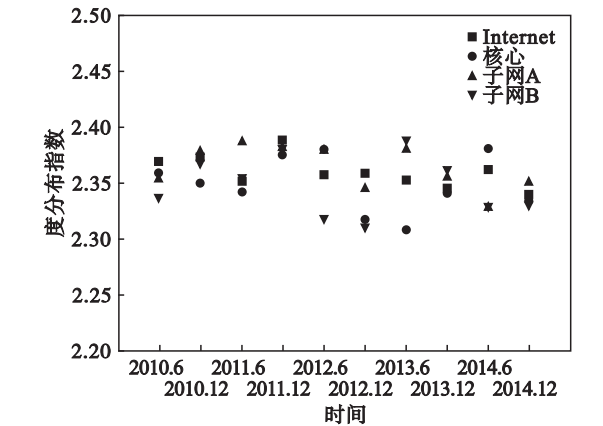


图 10 Internet 及其子网度分布指数

Fig. 10 Degree distribution exponents of the Internet and its subnets

图 11 展示了 Internet 整体、核心、子网 A 和子网 B 的同配系数波动情况。在同配系数方面，边缘子网可以表征整体性质，而核心则与整体存在较大差异。同时，边缘子网可以相互表征。

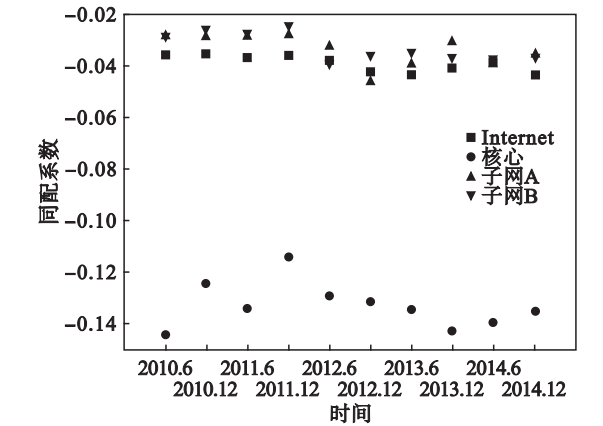


图 11 Internet 及其子网同配系数

Fig. 11 Assortativity coefficient of the Internet and its subnets

图 12 分别展示了 Internet 整体、核心、子网 A 和子网 B 的结构熵。与图 11 类似，在结构熵方

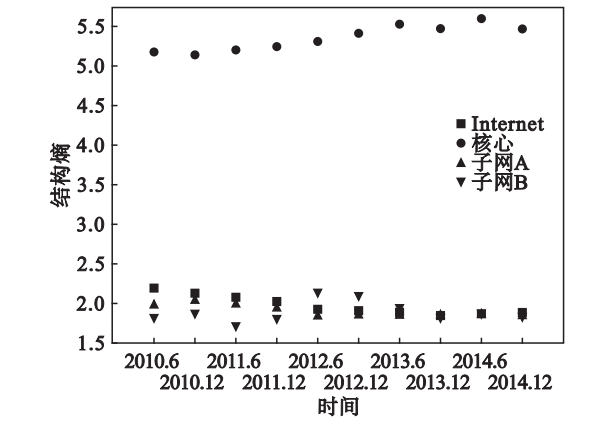


图 12 Internet 及其子网结构熵

Fig. 12 Entropy of the Internet and its subnets

面,Internet 边缘子网可以表征整体性质,而核心则不能表征.此外,边缘子网可以相互表征.

## 4 结 语

研究发现因特网具有层次分形特征. Internet 的核心层与边缘层具有完全不同的组织结构、复杂程度和规律性,对应的分形维数也存在较大差异.值得注意的是,Internet 虽然并非具有单一分形特征,在度分布指数、同配系数和结构熵等方面,其边缘子网仍可表征整体性质.与之对应,网络核心在度分布指数方面与整体具有一致性,在其他统计角度则不具备该性质.此外,在以上三个统计角度,Internet 的边缘子网均可以相互表征.

## 参考文献：

[ 1 ] Jun A,Hai Z,Carley K M,et al. Evolution of IPv6 Internet topology with unusual sudden changes[ J ]. *Chinese Physics B*,2013,22( 7 ):078902.

[ 2 ] Wang X, Loguinov D. Understanding and modeling the Internet topology; economics and evolution perspective[ J ]. *IEEE/ACM Transactions on Networking*, 2010, 18 ( 1 ): 257 – 270.

[ 3 ] Siganos V, Faloutsos M, Faloutsos P, et al. Power laws and the AS-level Internet topology[ J ]. *IEEE/ACM Transactions on Networking*,2003,11(4) :514 – 524.

[ 4 ] 关世杰,赵海. 互联网中路由级和 IP 级拓扑分形特征分析[ J ]. *通信学报*,2013,34(11) :162 – 170.  
( Guan Shi-jie,Zhao Hai. Analysis of fractal characterisitic of internet router-level and IP-level topology [ J ]. *Journal of Communications*,2013,34(11) :162 – 170. )

[ 5 ] Song C, Havlin S, Makes H A. Self-similarity of complex networks[ J ]. *Nature*,2005,433(7023) :392 – 395.

[ 6 ] Song C,Lazaros K G, Havlin H A. How to calculate the fractal dimension of a complex network; the box covering algorithm[ J ]. *Journal of Statistical Mechanics; Theory and Experiment*,2007,3(3) :1742 – 5468.

[ 7 ] Schneider C M,Kesselring T A, Andrade Jr J S, et al. Box-covering algorithm for fractal dimension of complex networks [ J ]. *Physical Review E*,2012,86(1) :016707.

[ 8 ] Kim J S, Goh K I, Kahng B, et al. Fractality and self-similarity in scale-free networks[ J ]. *New Journal of Physics*, 2007,9(6) :177.

[ 9 ] Gao L,Hu Y,Di Z. Accuracy of the boll-covering approach for fractal dimensions of complex networks and a rank-driven algorithm[ J ]. *Physical Review E*,2008,78(4) :046109.

[ 10 ] Newman M E J. *Networks; an introduction* [ M ]. New York: Oxford University Press,2010:23 – 25.

[ 11 ] Kim D H,Noh J D,Jeong H. Scale-free trees; the skeletons of complex networks [ J ]. *Physical Review E*, 2004, 70 ( 4 ): 046126.