

互联网中路由级和 IP 级拓扑分形特征分析

关世杰^{1,2}, 赵海¹

(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110819; 2. 沈阳理工大学应用技术学院, 辽宁 抚顺 113122)

摘 要: 首先使用粗粒度重整化方法对互联网 IPv4 路由级拓扑和 IPv6 IP 级拓扑进行处理, 得出度分布幂指数不随粗粒度化测量尺度变化的特性; 其次, 利用分形理论中分布函数的方法计算互联网在不同维度下的分形维数, 探索互联网度分布的分形特征; 最后通过对 CAIDA 数据源的数据分析, 研究了 IP 级数据在时间演化上的分形特征, 并得出相应的结论。研究结论能够为互联网结构特性研究提供一定的理论基础。

关键词: 复杂网络; 分形; 度; 重整化算法; k -core 分解

中图分类号: TP393.4

文献标识码: B

文章编号: 1000-436X(2013)11-0162-09

Analysis of fractal characteristic of internet router-level and IP-level topology

GUAN Shi-jie^{1,2}, ZHAO Hai¹

(1. School of Information Science & Engineering, Northeastern University, Shenyang 110819, China;

2. Polytechnic School of Shenyang Ligong University, Fushun 113122, China)

Abstract: The preliminary works have found that fractal theory has greater leverage in the analysis area of the Internet architecture. So firstly the method of the renormalization of coarse-grain was used to deal with the IPV4 and IPV6 router level topology of the internet, and it was found that the characteristics of power index of degree distribution didn't change with measurement scales of coarse grain. Secondly, the it was found that method of the fractal distribution function was used to calculate the fractal dimension of different dimensions of internet and explore the fractal characteristics of degree distribution of internet. Lastly, through the analysis of the CAIDA data source, the fractal characteristics of IP-level data were studied on during the time evolution, and then the appropriate conclusions were got. The research conclusions of this article can give the theoretical basis to the study on structural characteristics of the internet.

Key words: complex network; fractal; degree; renormalization algorithm; k -core decomposition

1 引言

Internet 作为一个典型的复杂网络实例, 对其宏观拓扑结构特征的分析是目前研究的主要内容和热点问题, 受到了学术界的广泛关注^[1,2]。近年来人们在该领域的研究取得了长足的进展^[3-5], 但主要集中在拓扑的全局层面上。对 internet 宏观拓扑结构进行粗粒度化处理及分形特征分析, 可以帮助人们更好地理解互联网部分与整体以及部分与部分之间的关系, 对人们通过少量信息认识研究对象的本质有重要的意义。20 世纪 60 年代,

复杂网络的研究才刚刚兴起, 当时只把随机图理论作为研究方向, 复杂网络的小世界特征、无标度特征的发现, 引发了研究复杂网络的热潮。到 21 世纪初期, 复杂网络的研究领域已经渗透到工程学科、生命学科和数理学科等多个不同的科学研究领域, 研究方法也由图论发展到物理学、概率与数理统计学和分形理论。分形一词是 1975 年提出的, 分形是一个零碎或者粗糙的集合形状, 把它分成若干部分, 而且这些部分在一定程度上是和整体相近的形状, 这个性质也称为自相似。由于分形理论具有丰富的研究成果和成熟的分析

收稿日期: 2013-02-17; 修回日期: 2013-06-25

基金项目: 国家自然科学基金资助项目(60973022)

Foundation Item: The National Natural Science Foundation of China(60973022)

工具,而且复杂网络呈现出了自相似的分形特征,使复杂网络中对分形理论的研究成为科学家们感兴趣的热点问题。

德国著名物理学家普朗克在 20 世纪初就认为:“科学是内在的整体,它被分解为单独的整体不是取决于事物本身,而是取决于人类认识能力的局限性。实际上存在从物理学到化学、通过生物学和人类学到社会学的连续的链条,这是任何一处都不能打断的链条”。许多研究者们从系统的整体出发,研究复杂网络。复杂网络的特征作为复杂网络一个重要的研究方向,经历了一个从简单到复杂的过程,经历了从最初的随机特性到之后的无标度特性、小世界特征、分形特征逐渐深化的过程。SONG 等人通过分形理论来描述复杂网络的分形自相似特性,他提出了复杂网络新的研究角度^[6],而且在不断的深入研究中,他提出由于度值高的节点之间的互斥性是复杂网络产生分形特征的原因,之后又提出了盒维数的计算方法^[7]。2004 年 KIM 等人通过研究被称为骨架(skeleton)的模型结构建立复杂网络的分形网络模型^[8],并提出分形网络特征形成的原因,他们认为分形网络最初是由分形骨架和节点之间最短的连接组成^[9-11]。以上研究成果的提出使复杂网络的分形特征的研究获得了飞速发展^[12,13]。

本文的研究工作正是以分形理论为基础,结合 CAIDA ARK 项目的海量实测数据,计算了 internet IPv4 路由级拓扑、IPv6 IP 级拓扑的分形维数,并通过度分布幂指数在不同重整化网络 and 不同标度尺度下的不变性给予了证明。本文采用核入口方法对网络进行粗粒度重整化处理,并在此基础上分析网络部分与部分、部分与整体之间所存在的自相似特征。

2 核入口粗粒度重整化方法

粗粒度化方法是把网络的核数作为粗粒度化

网络的基础,相关定义如下。

定义 1 度:节点 i 的度是指与其连接的其他节点的数目。

定义 2 核数:拓扑图的 k -核是指反复去掉图中度值小于 k 的节点及与之相连的边后剩余的子图,若节点核数为 k ,则该节点存在于 k -核中,而不存在于 $(k+1)$ -核中。拓扑图中节点核数的最大值为该拓扑图的核数。

定义 3 KC 网络:由集合 $C \subseteq V$ 推导出的子图 $H = (C, E|C)$,当且仅当对 C 中的任意节点 v ,其度值均大于或等于 KC ,即 $\forall v \in C: degree_H(v) \geq KC$,具有该性质的最大子图就叫做 KC 网络。核中的节点数称为该 KC 网络的大小。

定义 4 粗粒度重整化:将原始网络中具有某些共同性质的节点集合构成一个具有较少节点的新的网络。该过程可以持续进行,直至网络为空。

依据 KC 网络定义,图 G 的 KC 网络可以重复移去度值比 KC 小的节点和与其连接的边,当图中余下的所有节点的度值全部大于或等于 KC 时结束。因此可以通过核入口重整粗粒度化方法对原始网络层层分离,直到网络不为空的最大的 KC 值,揭示了复杂网络的层次结构特征。该重整化的过程如图 1 所示。

图 1 说明了核入口粗粒度重整化的整个过程。通过不断去掉度值小于 KC 的节点和与该节点相连的边。余下的子图都是度值 $\geq KC$ 的节点和与这些节点连接的边。图 1(a)是去掉度值小于 1 的节点和与该节点连接的边后余下的子图。可知原始网络是 $KC=1$ 的子图。图 1(b)为连续去掉度值小于 2 的节点和与该节点相连的边,图中为去掉节点 1、2、11 和与这 3 个节点相连的边(图中用虚线表示的连边)。剩余子图为 $KC=2$ 的网络。图 1(c)为连续去

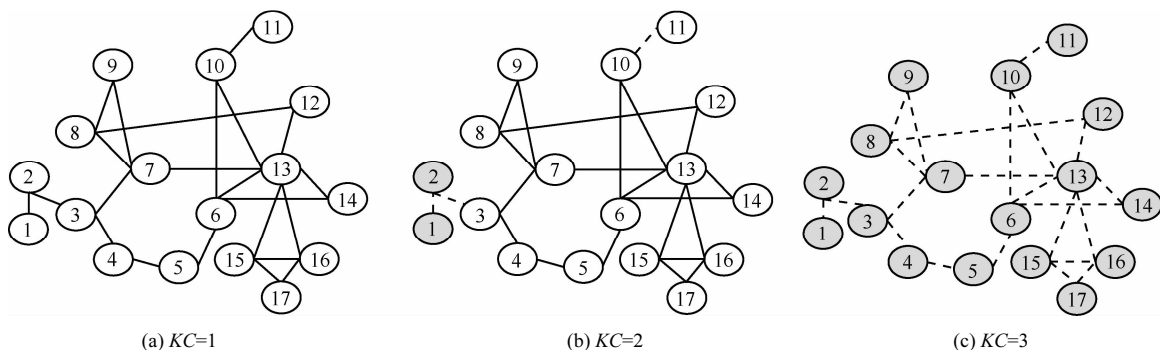


图 1 核入口粗粒度重整化网络过程

的发展迅速，核层次随着数据量的增多而不断加深。

4 路由级拓扑分形特征分析

4.1 KC 网络的标度不变性

定义 5 度分布：是指图中所有节点的度的分布情况，可用分布函数 $p(k)$ 描述。 $p(k)$ 表示的是一个随机节点的度恰好是 k 的概率。

定义 6 标度不变性：即无特征长度。具有分形特征的系统在分形有效范围内的任何局部区域，不论将其放大或缩小，他的形态、复杂程度、不规

则性等各种特性均不会发生变化。这是具有自相似特性的物体（系统）必定满足的性质。

最常用的复杂网络的度量特征是度分布，由于发现 internet 拓扑幂律度分布，促使对 internet 拓扑特性的研究迅速兴起，而以前的 internet 模型不能体现出这个特性。由于 IPv6 的使用刚刚起步，其路由级数据相对较少，如果作为分析样本不足以说明问题，所以本文选取了 2011 年 2 月的 CAIDA 里 15 个探测节点探测到的 IPv4 路由级数据，并对数据进行统计分析，对 KC 取不同的值，复杂网络的

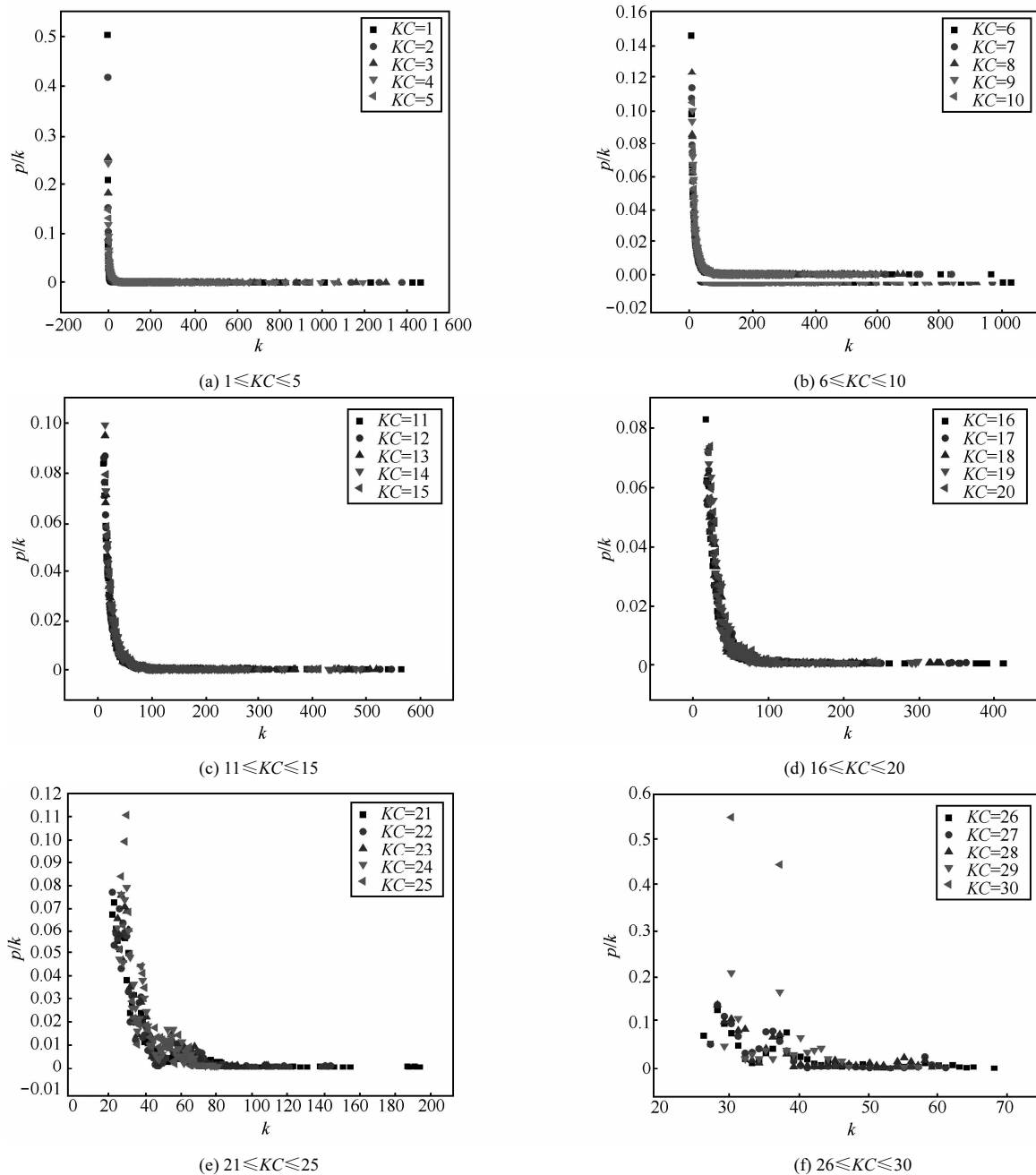


图 3 KC 网络度分布

度分布如图 3 所示。

图 3 是 30 个 IPv4 路由级数据粗粒度网络的度分布图。可以得出低度值节点数量最多，随着度值不断增大，该度值节点数量迅速减小，而高度值节点的数目最少。图 3(a)~图 3(d)在 1~20 核中，这一区域幂律特性特别明显。图 3(e)和图 3(f)代表 21~25, 26~30 网络的度分布。在 21~25 核区间内还存在幂律性，但是中间度值节点有微小震荡，这说明中高度值节点在粗粒度化的过程中出

现的概率增大。在 26~30 核区间范围内，基本上是高度值节点，并且节点分布比较均匀，度分布不满足幂律特性。

对图 3 IPv4 路由级的度分布取双对数，如图 4 所示。

判断网络或者集合是否具有分形自相似特征，主要判断度分布的幂指数变化。如果网络是具有分形特征，那么无论观察尺度如何变化，度分布的幂指数始终是固定不变或者变化幅度不大，从而证明

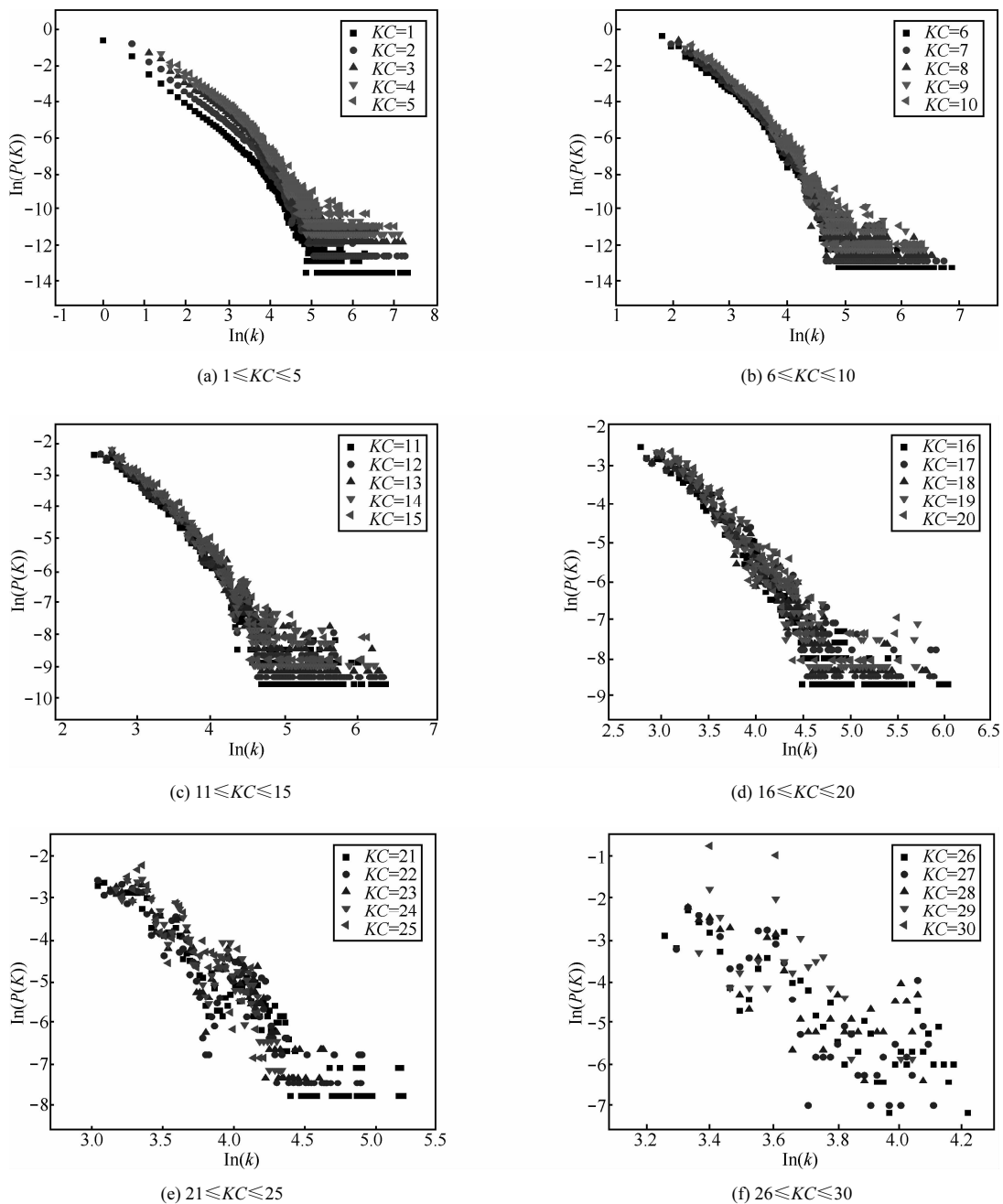


图 4 路由级拓扑各层节点度分布拟合

其具有标度不变性。

图 4 描述了 2011 年 2 月 IPv4 路由级数据 30 个粗粒度网络的度分布对数拟合情况。在 1~20 核的范围内 (图 4(a)~图 4(d))，度分布图形基本重合，拟合之后的幂指数值约为 2.68。在 21~25 核的区间内 (图 4(e)) 线性拟合的效果不理想，和 1~20 核的区间的幂律性不同。在 26~30 核中 (图 4(f))，幂律特性不存在。这说明在低、中核区间内的节点度分布没有受到观察尺度 KC 值的影响，且有明显的幂律特性，而且幂指数基本保持不变。而在高核区幂律特性完全消失，是因为节点特别稀少且分布差别较小。IPv4 路由级拓扑整体与粗粒度化网络度分布都满足幂律特性，这说明 IPv4 路由级拓扑网络具有分形自相似特征。

4.2 分形维数方法选择

下面结合网络本身具有的较深核层次性，用分布函数计算分形维数。

节点 i 的核数记为 k_i ，对应的核数分布为 $p(k_i)$ ， KC 网络的 KC 分布记为

$$P(kc) = \sum_{k_i=kc}^{\max(kc)} p(k_i) \quad (1)$$

大于 KC 的 KC 网络存在的分布 $P(KC)$ ，记为

$$P(KC) = P(kc > KC) = \sum_{kc=KC+1}^{\max(KC)} P(kc) \quad (2)$$

对观察尺度 KC 进行 $KC = \lambda KC$ 的变换时，则对任意的 $\lambda > 0$ ，都有

$$P(KC) \propto P(\lambda KC) \quad (3)$$

KC 网络的分布函数 $P(KC)$ 满足下面的幂型

$$P(KC) \propto KC^{-D} \quad (4)$$

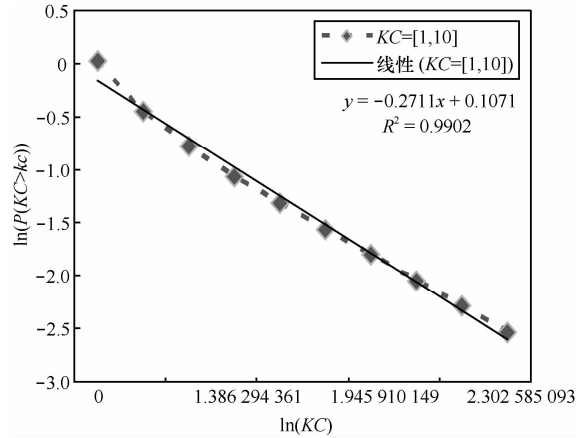
其中，幂指数 D 是 KC 网络的分形维数。并且在分形有效范围内，无论 KC 为何值， KC 网络分形维数 D 值始终不变。

4.3 分形维数

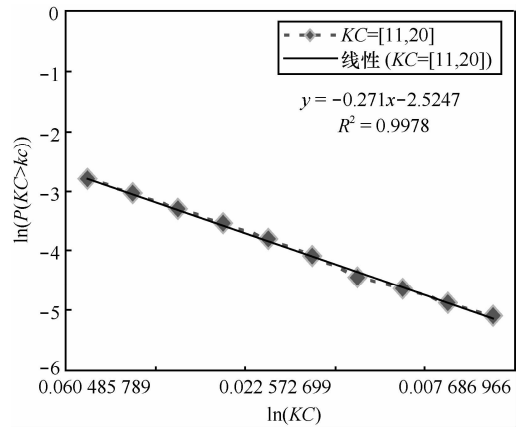
路由级数据的粗粒度化结果，根据式(4)求得的分形维数如图 5 所示。

图 5 将 IPv4 路由级数据的 30 个粗粒度化网络分为 3 个阶段分别计算分形维数。其中 KC 取 1~10 与 11~20 的网络的分形维数相同，约为 0.271。但是当 KC 值为 21~30 时的网络，分形维数约为 1，此值为整数，证明此网络不具有分形特征。不同学科中所研究的分形与自然界存在的分形，它们并不具

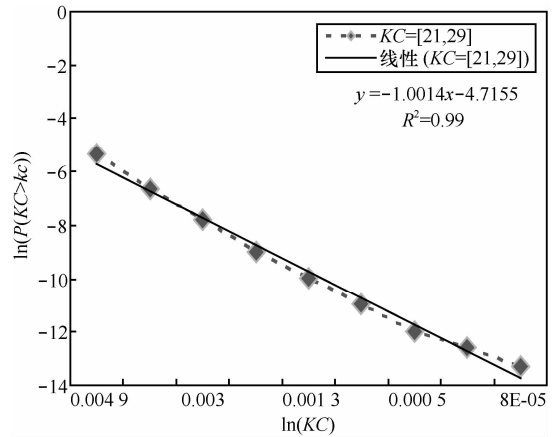
有无穷的嵌套结构，而是具有有限的嵌套层次。而 KC 取值较高的粗粒度网络就超过了网络嵌套范围。



(a) $1 \leq KC \leq 10$



(b) $11 \leq KC \leq 20$



(c) $21 \leq KC \leq 30$

图 5 不同核尺度下的分形维数

5 IP 级拓扑分形特征及时间演化分析

5.1 KC 网络标度不变性

IPv4 经过多年的应用已经非常成熟，但是 IPv4

的数据量较大，计算比较困难，所以 IPv4 的 IP 级拓扑数据并不适合用于统计或计算。而 IPv6 仍处于起步阶段，为下一代互联网的 IP 级拓扑结构的研究提供了合适的机会，故选择 2010 年 6 月的 IPv6 IP 级拓扑数据进行分析，结果如图 6 所示。

其中图 6(a)为 KC 值为 1~5 的网络，图 6(b)为 KC 值为 6~10 的网络，图 6(c)和图 6(d)分别是图 6(a)

和图 6(b)取双对数结果。 KC 值较低时度分布幂律特性明显，幂指数值约为 2.07， KC 取值较高时度分布幂律性不明显。这与 IPv4 路由级拓扑在不同粗粒化网络中表现的幂律特性相一致。

5.2 度相关性

下面对 IPv6 进行度相关性分析，主要分析累积邻节点平均度比例的变化情况，结果如图 7 所示。

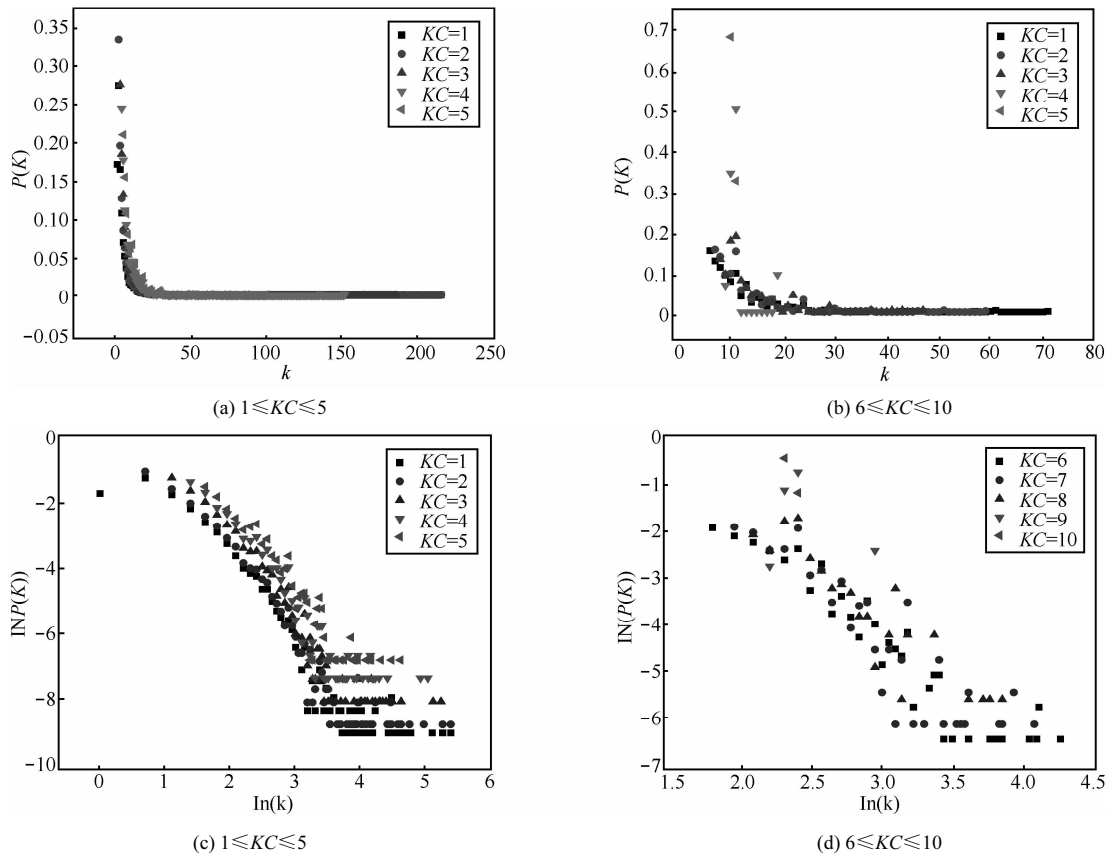


图 6 IPv6 IP 级拓扑度分布特性

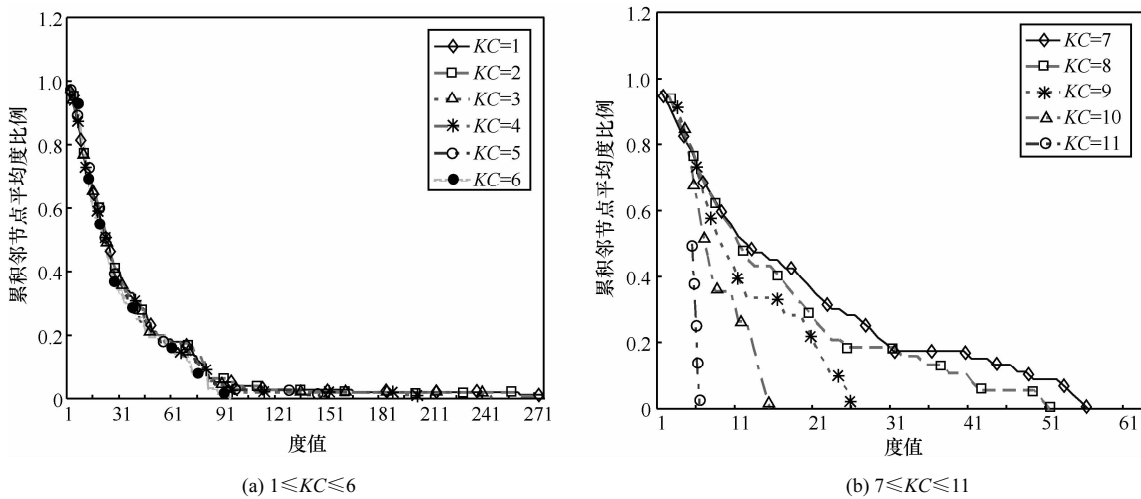


图 7 IPv6 IP 级粗粒化网络累积邻节点平均度比例变化

图 7 为 2010 年 4 月粗粒度化网络的累积邻节点平均度比例变化趋势。图 7(a)，图 7(b)分别为 1~6，7~11 核的累积邻节点平均度比例变化。图 7(a)曲线走势基本保持重合；图 7(b)高核区域节点度值分布比较平均，没有出现重尾现象，这与 IPv4 路由级数据所表现的特征相一致。可以得出 KC 取值较低时具有分形特征，而在网络的核心部分不存在这一个特征。

5.3 分形维数

由于 IPv4 路由级拓扑的数据量比较大，所以没有在时间序列上分析。下面对 IPv6 IP 级数在时间尺度上进行分析，分析结果如图 8 所示。

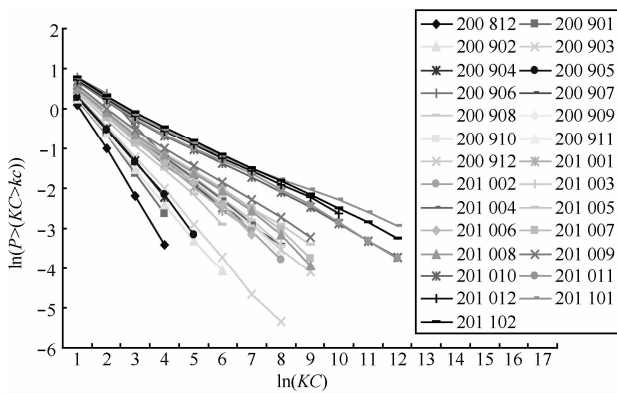


图 8 IPv6 IP 级数据时间序列上分形维数变化

下面对 2008 年 12 月到 2011 年 2 月共 27 个月的高度值节点间的连接情况进行分析。该时间段上度值最大的 0.5%和 1%的节点集的富人俱乐部特性的分析结果如图 9 所示。

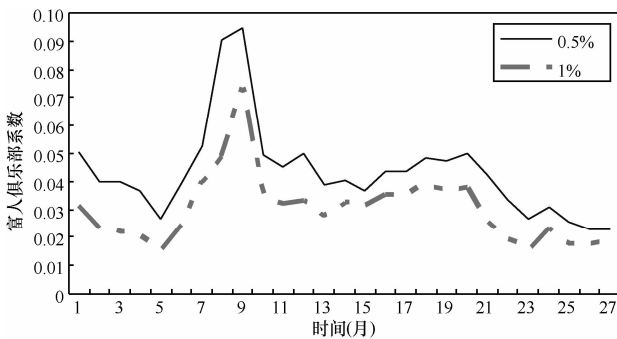


图 9 IPv6 IP 级数据时间序列上的高度值节点之间连接性变化

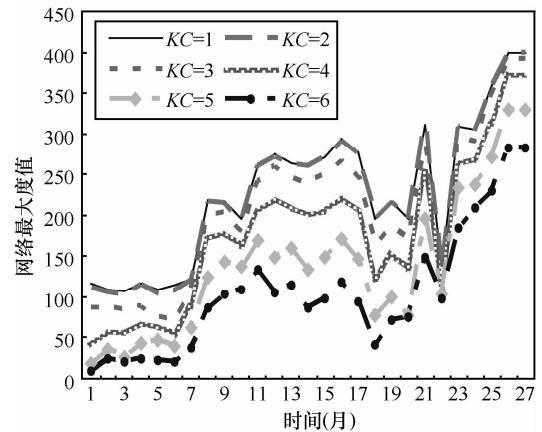
图 9 显示了 IPv6 富人俱乐部的特性随着时间变化的演化趋势。在网络中部分度值很高的节点，称为“富”节点，用富人俱乐部连接特性描述这些节点间连接倾向。如果该值越大，则表明“富”节点之间连接比较紧密；反之，则表明“富”节点间联

系比较松散。2 个节点集变化的趋势基本一致，虽然有一定幅度的震荡，但是总体趋势是下降的。该值的变化趋势和分形维数的变化趋势是一致的，这说明高度值节点之间的互斥性增强导致分形维数的降低，使分形特征更加明显。

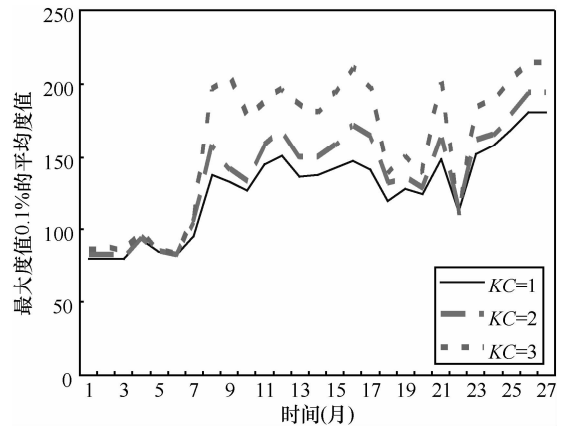
5.4 粗粒度化网络时间演化分析

下面将从粗粒度化网络的宏观拓扑特征进行分析，如图 10 所示。

图 10(a)显示了在时间序列上前 6 个粗粒化网络中网络最大度值的变化情况。图 10(b)显示了前 3 个粗粒度化网络的度值最大的前 1%节点的平均度值。在两幅图中， $KC=1$ 的网络和 $KC=2$ 的网络最大度值基本相同，说明最大度值节点和叶子节点连接的可能性比较小。而 2009 年 8 月与 2010 年 8 月的不同粗粒度化网络的值非常接近，说明该时间点上高度值节点和度值等于 KC 节点的连接极少，异配性较弱。其余时间点上的值在不同粗粒度化中差别增大，异配性较强。



(a) $1 \leq KC \leq 6$



(b) $1 \leq KC \leq 3$

图 10 网络最大度值及高度值节点平均度变化

根据以上分析,需要对时间序列上的同配系数的变化进行分析,验证是否一致,如图 11 所示。

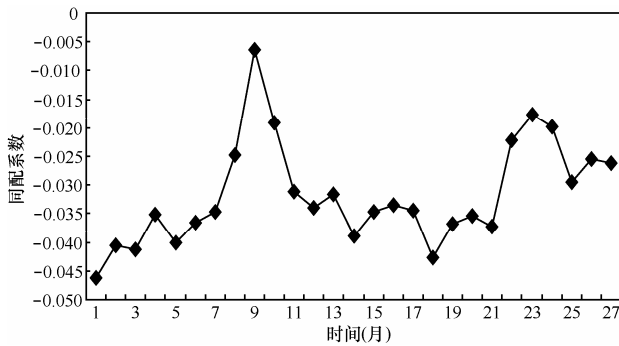


图 11 网络同配系数变化

图 11 显示了 IPv6 IP 级拓扑在 2008 年 12 月到 2011 年 2 月上同配系数的变化。该值在 2009 年 8 月和 2010 年 8 月达到最高值,说明了它的异配性很差,这正是造成了图 10 两幅图中该时间点上各粗粒度化网络之间差别较小的原因。和推论得到的结论是一致的。

6 结束语

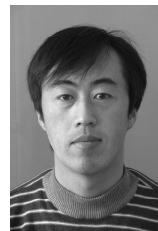
本文基于核的概念分别对互联网中 IPv4 路由级拓扑和 IPv6 IP 级拓扑进行粗粒度重整化处理,得出度分布幂指数不随粗粒度化测量尺度变化的特性,利用分形理论中分布函数的方法计算互联网在不同维度下的分形维数,由此得出: KC 取值不高的网络具有分形自相似特征, KC 较高时该特征不存在。针对 IPv6 的 IP 级数据分析,从图 9 中可以看出,2008 年 12 月~2011 年 2 月时间段上,度值最大的 0.5% 和 1% 的节点集的富人俱乐部系数均在震荡下降,这表明高度值节点之间的连接随着时间维度的发展在不断下降,即互斥性不断增强。而该规律与图 8 中的分形维度变化规律一致,这表明高度值节点之间的互斥性增强是分形维数降低的一个关键因素,互斥性的增强使分形特征更加明显,即分形维数在时间序列上不断减小,以上结论对进一步研究互联网结构特征具有一定的参考价值。

参考文献:

- [1] 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.
- [2] ALBERT R, JEONG H, BARABASI A L. Error and attack tolerance of complex networks[J]. Nature, 2000, 406(6794): 378-382.

- [3] 郭虹, 杨白薇, 兰巨龙等. AS 级 Internet 拓扑层次性分析与建模[J]. 通信学报, 2011, 32(9):182-190.
GUO H, YANG B W, LAN J L, *et al.* Hierarchy analysis and modeling on the Internet AS-level topology[J]. Journal on Communications, 2011, 32(9): 182-190.
- [4] 杨波, 赵海, 张君等. AS 级 Internet 拓扑突变性分析[J]. 东北大学学报(自然科学版), 2011, 32(3):376-380.
YANG B, ZHAO H, ZHANG J, *et al.* Analyzing catastrophe tendency of AS-level Internet topology[J]. Journal of Northeastern University(Natural Science), 2011, 32(3):376-380.
- [5] 张君, 赵海, 杨波等. AS 级 Internet 网络拓扑的中心性测量. 东北大学学报(自然科学版), 2012, 33(3): 340-343.
ZHANG J, ZHAO H, YANG B, *et al.* Centrality measurement of Internet network topology at AS-level[J]. Journal of Northeastern University(Natural Science), 2012, 33(3): 340-343.
- [6] SONG C, HAVLIN S, MAKSE H A. Self-similarity of complex networks[J]. Nature, 2005, 433(7024): 392-395.
- [7] SONG C, GALLOS L K, HAVLIN S, *et al.* How to calculate the fractal dimension of a complex network: the box covering algorithm[J]. Journal of Statistical Mechanics: Theory and Experiment, 2007, 2007(03):P03006.
- [8] KIM D H, NOH J D, JEONG H. Scale-free trees: the skeletons of complex networks[J]. Physical Review E, 2004, 70(4):046126.
- [9] GOH K I, SALVI G, KAHNG B, *et al.* Skeleton and fractal scaling in complex networks[J]. Physical Review Letters, 2006, 96(1): 018701.
- [10] KIM J S, GOH K I, SALVI G, *et al.* Fractality in complex networks: critical and supercritical skeletons[J]. Physical Review E, 2007, 75(1): 016110.
- [11] YOON S H, RADICCHI F, MEYER-ORTMANNS H. Self-similar scale-free network and disassortativity[J]. Physical Review E, 2005, 72(4):045105.
- [12] BACKES A R, BRUNO O M. Shape classification using complex network and multi-scale fractal dimension[J]. Pattern Recognition Letters, 2010, 31(1): 44-51.
- [13] SCHNEIDER C M, KESSELRING T A, ANDRADE J J S, *et al.* Box-covering algorithm for fractal dimension of complex networks[J]. Physical Review E, 2012, 86(1):016707.

作者简介:



关世杰(1977-), 男, 辽宁沈阳人, 东北大学博士生, 主要研究方向为复杂网络。



赵海(1959-), 男, 辽宁沈阳人, 东北大学教授、博士生导师, 主要研究方向为复杂网络、数据融合。