

互联网(IPv4/IPv6)宏观拓扑结构生命特征

刘 晓^{1,2} 赵 海¹ 李少锋¹ 王进法¹ 李鹤群¹

¹(东北大学信息科学与工程学院 沈阳 110819)

²(杜伦大学生物与生物医学学院 英国 杜伦 DH1 3LE)

(xiao.liu3@durham.ac.uk)

Vital Signs of IPv4/IPv6 Macroscopic Internet Topologies

Liu Xiao^{1,2}, Zhao Hai¹, Li Shaofeng¹, Wang Jinfa¹, and Li Hequn¹

¹(College of Information Science & Engineering, Northeastern University, Shenyang, 110819)

²(School of Biological & Biomedical Sciences, Durham University, Durham, UK, DH1 3LE)

Abstract The Internet is considered under the comparative process of biological evolution, which can predict the future of the Internet. The IPv4 and IPv6 whole networks data from 2010 to 2014 authorized by CAIDA was adopted in this paper. On the basis of the previous researches, this study combines the Eigen's definition of 'vital signs' with the Internet topology characters; the metabolism of the Internet is measured by the standard Internet structure entropy, and the process of the Internet self-replication is represented by the various degree distributions of K -core networks, and the errors happened in the process of Internet self-replication is characterized by the changes in the evolution of the average clustering coefficient and average path length. The vital signs are observed in both IPv4 and IPv6 Internet topologies, and IPv6 which has prominent vital signs behaves more vitality. The detection of the vital signs from Internet topology is a mark of introducing biological principles to researches on Internet topology successfully. For the further studies which are researched on or predict the dynamic of the Internet, the discovery in this study calls for attention to treat the Internet as a creature. It enriches the research methods and brings more space of development for the Internet researches. At the meantime, some advice for research direction about redesigning and rebuilding Internet is put forward in this study.

Key words macroscopic Internet topology; IPv4; IPv6; vital signs; system evolution and analysis

摘 要 从生物进化的角度看待互联网演化,可以操控并引导网络演化方向,预测未来互联网结构与功能的发展趋势. 择取 CAIDA 互联网全球研究机构 2010—2014 年互联网 IPv4 及 IPv6 全网探测数据. 将 Eigen 对“生命特征”的定义与互联网宏观拓扑特征量的定义融合;以标准网络结构熵表征网络代谢行为;以 1-25 核网络度分布频度变化描述网络自复制特性;从网络平均聚集系数及网络平均路径长度的演化情况观测互联网自复制过程中出现的误差. 对比分析互联网 3 种生命特征,发现 IPv4 与 IPv6 互联网均具有生命特征,且 IPv6 的生命特征较 IPv4 更明显,网络拓扑更具生命力. 从互联网宏观拓扑中提取生命特征,是将生物学思想成功引入互联网领域的标志,这不仅对未来互联网研究提供更丰富的研究方法和更大的研究空间,更为有效地实现网络功能、预测并指导互联网发展提供理论支持.

收稿日期:2015-12-20;修回日期:2016-02-03

基金项目:国家科技支撑计划基金项目(2012BAH82F04);辽宁省科学技术计划基金项目(2015401039)

This work was supported by the National Key Technology Research and Development Program of China (2012BAH82F04) and Liaoning Science and Technology Project (2015401039).

关键词 互联网宏观拓扑; IPv4; IPv6; 生命特征; 系统演化分析

中图法分类号 TP393.01

互联网作为人类文明及社会信息化的标志,其规模增长非常迅速.互联网从诞生到现在,网络宏观拓扑结构每时每刻都在发生变化.随着现代人类需求的提高,互联网需要提供质量更好的服务,其演化速度比以往任何时候更快,网络拓扑结构正向着更高的复杂度演化^[1-3].近年来,以传统的机械手段分析研究互联网已经成为一个成熟且成果卓著的研究领域^[4-6].但随着网络结构复杂度的提高,人类想要实现最有效的挖掘、利用互联网资源对互联网进行精确测量及控制,变得昂贵且不现实.

在此背景下,研究人员尝试放弃对互联网的精确控制,开始探索新的研究思路及方向.以生物科学领域的理论与规则指导互联网研究,是一个大胆却潜力无限的尝试.从20世纪90年代,计算机科学家Langton就曾多次在其著作中提到“人工生命”,他指出,“生命是一个过程,它的组成材料并不重要,重要的是它做了什么^[7-8]”.一旦互联网被证明是一个生命过程,将对互联网宏观拓扑的研究及发展方向提供更便捷、有效的研究方法和更大的研究空间.近年来,虽然有学者尝试从互联网宏观拓扑中提取生命逻辑^[9-11],却始终缺乏系统性研究结论,相关研究成果并未引起广泛的关注与认可.2005年, Song等人提出广义的盒子计数法,第一次试图将分形理论应用到网络拓扑结构研究当中^[12].然而通过分形证明互联网生命过程中的自复制及自相似特性的方法在2012年被Schneider等人证明是不合理的^[13].在前人研究的基础之上,本文融合生物学思想,引入Eigen的超循环理论,从机械的互联网宏观拓扑特征量出发,着眼于大尺度时间跨度内宏观拓扑框架下的互联网.本文尝试从互联网宏观拓扑中提取生命特征,将互联网视作具有自适应、自我进化能力的宏大智慧生命体,对IPv4及IPv6互联网的生命特征及进化行为进行初步探究.

本文将按照如下5种结构进行叙述:1)明确数据来源及相关定义;2)根据互联网标准网络结构熵进行网络拓扑结构的代谢特征分析;3)观察分析经过K-核解析之后互联网1-25核网络2010—2014年间网络度分布频度的相似性,探讨互联网的自复制行为;4)通过分析互联网的网络聚集系数及平均路径长度这2个代表网络结构特征及数据传输性能的特征参数,分析互联网自复制行为过程中产生的误差;5)对观察到的现象进行讨论并总结.

1 实验基础:数据来源与相关定义

本文基于CAIDA(The Cooperative Association for Internet Data Analysis)_Ark探测项目,选取2010—2014年共60个月IPv4与IPv6互联网宏观拓扑数据,以月(month)为探测单元(time unit),探究这2种协议构架下互联网拓扑所呈现的生物属性. Barabási等人的研究^[14]证明抽样数据替代全数据的研究具有可靠性,即测量较少的探测源点便足以获取互联网拓扑本质特性.分析之前,首先给出定义如下:

定义1. Eigen判据.德国科学家Eigen在20世纪70年代在生物学领域的研究中提出超循环理论. Eigen的超循环理论是关于非平衡态系统的自组织现象的理论.该理论指出,不同尺度的生命活动中都存在循环现象^[15-16].在之后对于生命属性的研究中,生物学领域内的研究者将Eigen的超循环理论总结为代谢行为、自复制行为及自复制过程中产生的误差,并普遍以其作为发现和判定生命属性的典型判据^[17-18].

1) 代谢行为.它是整个系统的运行与信息传递过程,是一种拓扑的融合能力,该过程中网络结构熵值会发生变化.

2) 自复制行为.它是系统保持原有信息结构的内在能力,保证了在系统运行与信息代谢过程中结构的稳定及信息不会完全遗失.

3) 自复制过程中产生的误差.它是指系统运行与信息传输过程中的“噪声”,下一时刻与上一时刻的“不同”即为误差.它使数据在自复制过程中无法达到绝对精确,本文将其理解为系统优化的动力.

定义2. 网络结构熵.其宏观意义描述了复杂网络的拓扑结构状态的有序程度,系统结构越有序,熵就越小,反之则越大^[19];而从微观层面,它是系统组成单元混乱程度及有序程度的度量,亦可衡量系统微观状态下内部动态发展的可能性,熵越小,系统内部各组成单元之间秩序性越强,系统越具发展动力,生命力越强,反之熵越大^[20-21].

$$H = - \sum_{k=1}^N p(k) \lg p(k), \quad (1)$$

其中, $p(k)$ 为网络节点的度分布, k 是节点度数, H 为网络结构熵.由式(1)可知,网络结构熵是由网络

节点的度分布决定的, 根据网络结构熵的极值性, 可得如下定理^[20]:

定理 1. 当网络拓扑结构呈完全均匀状态时, 网络被认为是最混乱且无序的. 该极端情况下的网络拓扑, 由于结构内部能量分布均匀而极不稳定, 随时面临系统的崩溃和拓扑的解构, 系统随时会进入稳定平衡态, 网络内部通信等“活动”停止, 此时网络结构熵最大, 即 $H_{\max} = \lg N$.

定理 2. 当网络呈星形结构时, 结构最不均匀, 是一种极端的集权式网络, 然而该网络呈现的秩序性最为明显, 此时网络由于极强的秩序性使得网络拓扑内部各部分间有序自组织运行, 因此, 星形网络拓扑在不受外界干扰或控制的情况下秩序性最强、最具有生命力. 此时的网络结构熵达到极小值, 即

$$H_{\min} = \lg N - [N/(N-1)] \lg(N-1).$$

定义 3. 标准网络结构熵. 对网络结构熵进行归一化处理后得到标准网络结构熵^[20], 如此处理以消除数据样本对统计结果的影响.

$$H_s = \frac{H - H_{\min}}{H_{\max} - H_{\min}}. \quad (2)$$

归一化处理使标准网络结构熵与所采用的统计数据样本无关, 该操作消除了网络中节点数量对网络结构熵值的影响, 因此可以通过标准网络结构熵的演化规律, 从网络拓扑内部探究其演化潜力和代谢行为^[22].

定义 4. 度分布频度. 度分布是较为常用的一个网络特征评价参数, 指节点度值的频率分布. 互联网的拓扑特征存在于节点与边的关系之中, 幂律分布描述的正是互联网连接边的异质性特征. 对于幂律分布有 $p(k) \sim k^{-\alpha}$, 其中 k 为节点度值, $p(k)$ 是度值为 k 的节点出现的概率, α 为幂指数^[23-24].

定义 5. K -核. 核的解析可以用来描述度分布所不能描述的网络特征, 揭示源于系统特殊结构的结构性质和层次性质^[25-26]. 通过递归地移走图中所有度值小于 k 的节点, 直到剩余图中所有节点的度值都大于或等于 k 的方法来获得.

定义 6. K -核网络. 由集合 $C \subseteq V$ 推导出的子图 $H = (C, E|C)$, 当且仅当对集合 C 中的任意节点 v 的度值大于或等于 k , 即 $\forall v \in C: \text{degree}_H(v) \geq k$, 具有该性质的最大子图就叫做 K -核网络^[27-28]. 核中的节点数称为该 K -核网络的大小.

定义 7. 聚集系数 (clustering coefficient). 描述网络中节点的集团化程度, 即连接在一起的各个集团节点的邻居之中有多少是共同邻居. 网络中单个

节点的聚集系数 C_i 指该节点其邻居节点之间边的实际连接数与最大连接数目之比^[24]:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \quad (3)$$

其中, E_i 是节点 i 其邻居节点之间连边数目, k_i 是节点 i 的邻居节点数.

定义 8. 平均聚集系数. 整个网络的聚集系数称为平均聚集系数 C , 指在网络中与同一节点连接的 2 节点之间也存在连边的平均概率, 即所有节点的聚集系数 C_i 的平均值, 是描述网络宏观拓扑特征的重要特征量.

$$C = \frac{1}{N} \sum_{i=1}^N C_i. \quad (4)$$

定义 9. 平均路径长度 (average path length). 网络的平均路径长度 L (或平均最短路径) 是网络中任意 2 个节点之间最短路径长度的平均值, 它是网络的全局特征, 是衡量网络转发通信能力的重要指标. 用平均路径长度来衡量网络的通讯和传输性能, 较短的平均路径长度将使互联网具备更快的信息传输速度 (网络平均路径长度越小越好)^[24].

$$L = \frac{1}{N(N-1)} \sum_{i,j} d_{ij}, \quad (5)$$

其中, d_{ij} 为这 2 个节点 i 和 j 之间的最短路径, 是网络中任意 2 个节点 i 和 j 之间的所有通路中连通这 2 个节点的最少边数.

2 互联网生命特征的代谢行为分析

根据定义 1, Eigen 对生命系统的初步判断证据即为具有代谢行为. 对互联网而言, 在时间和空间的有限结合中, 其宏观拓扑结构的重复性破坏了系统的平衡, 却要求平稳地融合并生成新的宏观拓扑结构. 融合只对其中间状态起作用, 中间状态从高熵简单结构生成, 又将分裂为低熵复杂结构 (即更多的连边)^[29]. 以此为目的的拓扑融合, 我们将其称之为互联网的“代谢行为”. 根据定义 2 对标准网络结构熵的描述, 系统的稳定与否正是通过该特征参数体现的. 系统运行得越稳定, 结构熵值越低; 系统越面临崩溃或停滞, 熵值越大. 在此基础上, 对于动态系统, 熵增越小, 系统结构拓扑在时间、空间及功能上的有序度越好, 系统越具有代谢动力^[30].

本节以此为理论基础, 从互联网宏观拓扑结构出发, 以月为探测单元, 分析比较 2010—2014 年 IPv4, IPv6 互联网标准网络结构熵的演化情况, 2 种网络拓扑代谢行为演化分别如图 1(a)(b) 所示:

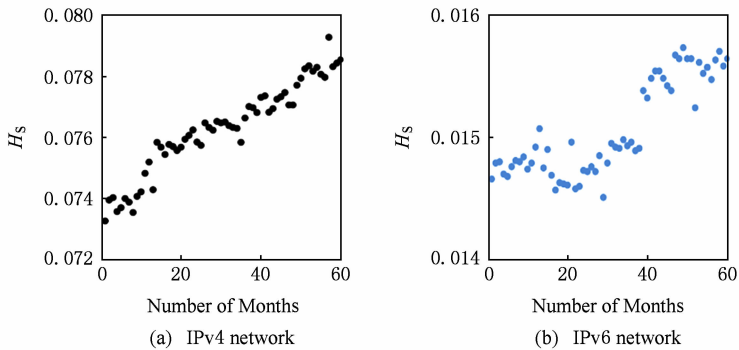


Fig. 1 Evolution of Internet structure entropy of the Internet.
图 1 互联网标准网络结构熵演化图

由图 1 可知,IPv4 与 IPv6 互联网标准网络结构熵均在时序上不断增大,但增长幅度并不相同.图 1(a)中 IPv4 结构熵从 0.074 波动上升至 0.079 附近,上升约 0.005;图 1(b)IPv6 结构熵较为稳定,探测初期基本维持在 0.0148 左右,在 2013 年下半年小幅阶跃至 0.0158 左右,上升数值为 0.001.可见在 60 个月的探测期间,IPv4 网络的熵增大于 IPv6.将 IPv4 与 IPv6 互联网标准网络结构熵演化情况进行对比,不同协议下互联网的标准网络结构熵存在差异,比较结果为 IPv4>IPv6.说明 IPv6 网络宏观拓扑较 IPv4 具有更强的拓扑演化秩序性,网络拓扑具有更强的发展潜力和更稳定的运行性能,即系统代谢功能更强,且经过相同的发展时间之后仍是如此,这与 IPv6 区域组网这一协议特点有很大关系.网络结构熵作为一个描述系统状态的函数,熵值变化趋势可直接反映系统状态演化方向.根据定理 1,标准网络结构熵最大值为 1,故相对而言,即便是标准网络结构熵值较大的 IPv4 互联网络,其量级也仅为 10^{-2} ,远非均匀网络,因此,互联网仍是宏观有序的拓扑结构,IPv4 互联网仍处于平稳正常的发展状态.未来从 IPv4 向 IPv6 或其他下一代网络的转型与演变过程当中,IPv4 仍具有代谢动力,系统不会出现崩溃或失稳等不良运行迹象.从系统论的角度来看,观察到的熵增趋势只是网络拓扑演化自组织代谢过程中的一次循环,IPv6 熵值演化平稳未见明显涨势,也可看作此过程中一次自组织相变之后暂时的稳态.可以说,互联网拓扑结构秩序的发展与演化是在一次次的系统自组织代谢过程中完成的.

3 互联网生命特征的自复制行为分析

根据定义 1 可知,生命系统在复杂的“生存”过

程中,为了保证代谢(融合)过程中不会导致信息的完全丧失,融合过程中的简单复制单元需要有一种内在的能力,指导其简单结构自身组合并且保持系统迄今积累起来的信息.这种内在的自催化能力,我们称之为“自复制能力”,它能够以最快的速度形成新的网络拓扑,并保证功能的稳定和可用信息的完整.同时,根据定义 5、定义 6,通过 K-核解构过程逐步删除网络中小度值节点,网络的核数表示节点在图中的深度,描述了网络拓扑的层次特征^[31],故求解 K-核的过程实际上是研究互联网的深层解构过程^[25-26].透过 K-核解构后的网络度分布频度可以观测到初始网络拓扑状态下度分布频度所不能描述的网络内部构造,揭示源于系统的特殊结构的拓扑属性和层次特征.

本节通过求解 IPv4 与 IPv6 互联网 1-25 核网络,逐层解构互联网初始拓扑结构,并分别逐层对比各核网络(1-核、5-核、9-核、13-核、17-核、21-核、25-核与 29-核)在 2010—2014 年的度分布频度,分析各层结构下 IPv4 与 IPv6 网络拓扑及其网络拓扑内部在探测期间的自复制行为.

图 2 为互联网 IPv4 各层网络(1-核、5-核、9-核、12-核、15-核、18-核、21-核与 25-核)在 2010—2014 年的度分布频度图.对数坐标下互联网 IPv4 同一 K-核网络,在 2010—2014 年的节点度分布频度的演化过程并未发生重大变构,且每层几乎如此,K-核网络度分布频度及网络拓扑的标度在每年均大致相似或基本相同.如图 2 可知,互联网 IPv4 网络 1-15 核之间的网络拓扑度分布频度呈更明显的线性分布,度分布频度重叠得更整齐,5 年间变化十分微小.说明宏观下的 IPv4 互联网拓扑由内核及外层逐层实现更精确的自复制行为,越涉及到外层网络,网络拓扑的自复制行为越会保留其原有的拓扑信息结构.

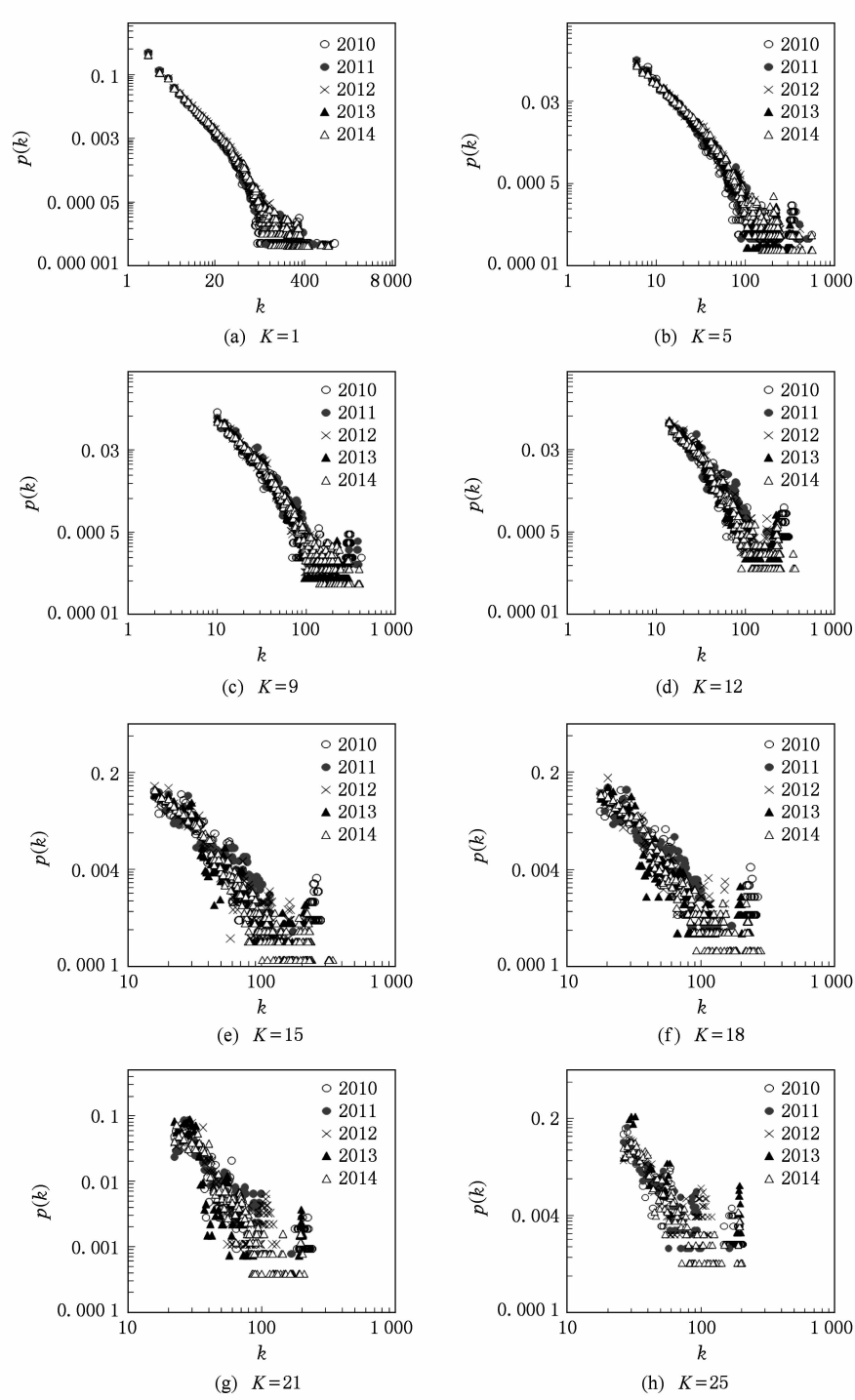


Fig. 2 Degree distributions of 1-25 core networks of IPv4 Internet topology form 2010 to 2014.

图2 2010—2014 年互联网 IPv4 拓扑 1-25 核网络度分布频度对比

如图 3 所示,互联网 IPv6 的 K -核网络(1-核、5-核、9-核、12-核、15-核、18-核、21-核与 25-核)度分布频度在 2010—2014 年间具有明显的相似性,说明每层 K -网络随时间的演化,网络度分布频度是大致不变的,网络拓扑的标度是相似的. 不过由于 IPv6 仍然是处于疾速发展中的新型协议互联网,其并不像 IPv4 那样具有成熟的拓扑及深度的内核,尤其在

探测初期的 2010 年,解构至 17 核,全网只剩 6 个节点;不过在探测后期,2012—2014 年内,全网均可以解构至 25 核,且解构后的网络度分布频度亦具有明显的相似性. 由于度分布描述了网络中节点与边的关系,那么每层 K -核网络在 2010—2014 年的演化过程中网络的宏观拓扑特征是相似的,保留了相同的节点与边的关系,即保留了网络原有的信息与结构.

通过图 3 的对比与演化分析可以得出,IPv6 网络拓扑正在逐渐向着更复杂、具有更深内核的方向演

化,随着拓扑的演化,系统内可见明显的自复制演化行为.

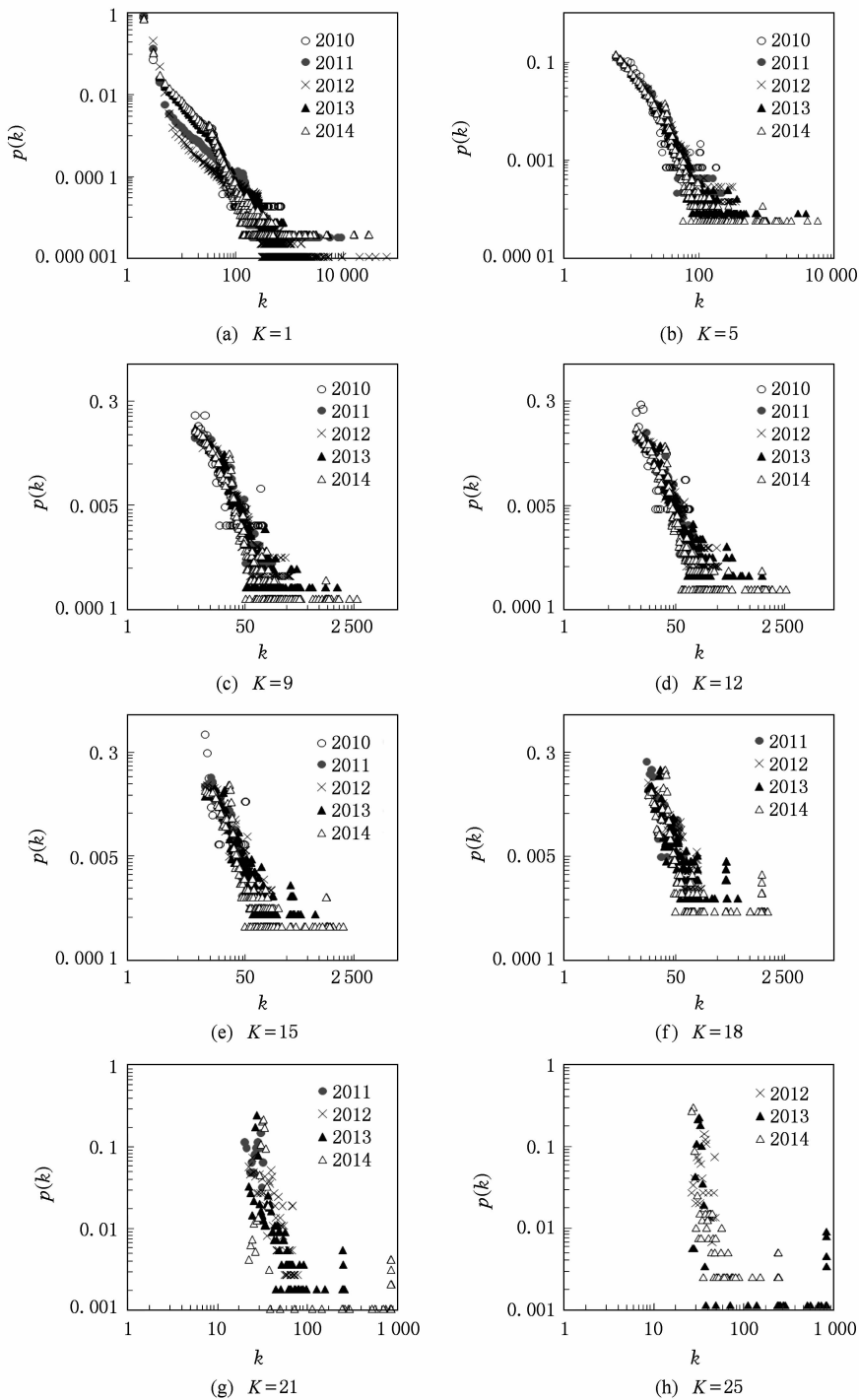


Fig. 3 Degree distributions of 1-25 core networks of IPv6 Internet topology form 2010 to 2014.

图3 2010—2014 年互联网 IPv6 拓扑 1-25 核网络度分布频度对比

综上,通过本节的分析,IPv4 与 IPv6 互联网 K-核网络在时序上度分布频度基本不变,虽然对比中发现每年之间的数据会有细微变化,但如此微小的扰动并不影响每一层网络在演化过程中都保留原有的拓扑结构及标度. 这表明互联网 IPv4 与 IPv6 网

络在运行与发展的过程中都在一定程度上保留了原有的信息结构,即具有自复制能力. 这在根本上保证了互联网拓扑能在最短时间内形成,并能够保证系统正常运行,且保证原有可用的结构功能稳定不变.

4 互联网自复制过程中产生的误差

由定义 1 可知,生命系统在“生存”的过程中,为了保持其“进化”的动力,自复制过程中会表现出一定的不精确性. 互联网的噪声,即数据的突变性,限制了处于一定状态的数据自复制过程的精确度. 本文将这种数据的突变称之为“自复制误差”. 值得注意的是,在生物学领域中,遗传过程中产生的误差是进化的基础,互联网中的复制误差则是形成新结构的主要结构信息来源. 定义 7~9 中的网络平均聚集系数和平均路径长度是网络拓扑结构的 2 个常规特征参数,表征网络拓扑特性和通信能力.

本节通过分析 IPv4 与 IPv6 互联网网络聚集系数与平均路径长度的演化情况,同时从网络拓扑(网络聚集系数)与网络性能(平均路径长度)的角度出发,探索并分析互联网在 2010—2014 年间网络自复制演化行为过程当中出现的不精确的复制“误差”.

4.1 互联网网络聚集系数演化

如定义 7,8 所描述,平均聚类系数是网络中节点聚集程度的度量^[23,32],可以有效地体现网络的聚集特性,是解析、描述网络拓扑结构形态的常规特征参数^[33]. 网络的聚集系数越大,其网络拓扑表现为邻居节点之间的连边越多. 依据式(3)(4)对 IPv4 与 IPv6 互联网 2010—2014 年数据进行统计分析,得到平均聚集系数时序演化情况,如图 4(a)(b)所示.

图 4(a)中 IPv4 及图 4(b)中 IPv6 互联网网络平均聚集系数在探测期间以小涨落的演化形式不断变化、逐渐上升,说明 2 种网络拓扑结构及拓朴性能并不是稳定不变的,而是随时间变化. 通过在拓朴内部建立更多的连边,调整网络的连通性,造成网络拓朴复制演化过程中的微小变化,使参数呈现波动与震荡. 通过图 4(a)(b)可以看出,2 种网络的聚集系数增量亦不相同,比较结果为 $\Delta\text{IPv6} > \Delta\text{IPv4}$,说明 IPv6 网络拓朴变化更活跃,变化更明显;同时,对比 2 种网络平均聚集系数,结果为 $\text{IPv6} > \text{IPv4}$,表明 IPv6 互联网中邻居节点之间连通性较 IPv4 更高,网络中节点与其邻居节点更倾向于构造一种密不可分的关系,彼此之间的连边数量更多. 从网络性能的角度评价,这样的网络鲁棒性好、局部连通性强,节点获得资源的机会更多,在一定程度上可以使网络尽可能地忽视掉由网络中个别节点失效所引起的一系列影响网络性能与传输效率的问题.

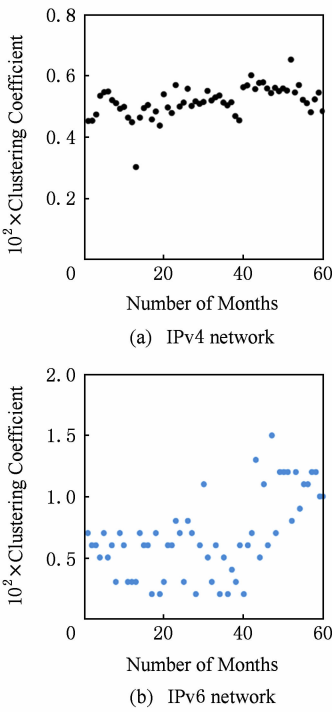


Fig. 4 Evolution of clustering coefficient of the Internet.

图 4 网络聚集系数时序演化图

4.2 互联网平均路径长度演化

Albert 与 Ash 等人分别在其研究中指出,平均路径长度是衡量网络性能直接且重要的拓扑特征量,网络平均路径长度越短,网络数据传输效率越高,网络中个别节点失效等微小扰动对网络通信造成致命打击的可能性越小,网络的性能越好^[32-33]. 根据式(5)对 2010—2014 年 IPv4 与 IPv6 互联网网络平均路径长度进行统计,其结果演化如图 5(a)(b)所示.

图 5(a)(b)中 IPv4 与 IPv6 网络的平均路径长度均始终处于动态变化中. 图 5(a)中 IPv4 网络的平均路径长度变化基本保持在 14.5~16.0 之间,并呈缓慢下降趋势,整体演化趋势较为平稳;图 5(b)中 IPv6 网络平均路径长度则处于相对活跃的波动演化状态,在前 40 个探测单元中 IPv6 网络平均路径长度维持在 12 的水平上,阶跃至 14 后开始逐渐下降至 13 左右. 可见,通过这一特征参数描述的 2 种协议下互联网网络拓朴及性能在演化过程中并不是稳定不变的,而是通过不精确的复制或是有“误差”的数据传输使参数呈现动态的震荡. 通过对比可以得出结论,IPv4 与 IPv6 网络平均路径长度比较结果为 $\text{IPv4} > \text{IPv6}$,IPv6 网络数据转发能力及转发效率始终优于 IPv4.

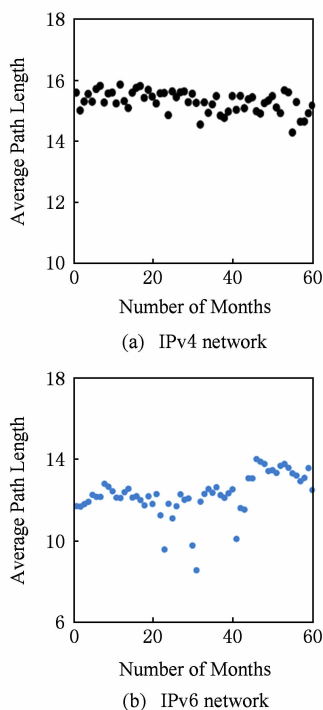


Fig. 5 Evolution of average path length of the Internet.

图5 网络平均路径长度演化图

通过本节的对比,在选取的2010—2014年探测期间内,IPv6网络较IPv4在网络拓扑性能参数上呈更活跃的演化状态,且具有更优良的结构与性能.说明IPv6网络更善于在互联网发展与演化的过程做出改变,并更易于在改变中调整自身拓扑结构,使拓扑结构与系统功能处于“最优”状态,提供质量更高的服务.这也充分验证了IPv6路由算法不断地更新路由表,构造不同的逻辑拓扑结构的协议优势.同时通过本节的对比分析,不难得出这样的结论:IPv4与IPv6互联网网络拓扑结构不是静止固定的,网络特征参数也不是稳定不变的,互联网拓扑结构会通过增减连边或节点等行为改造自身拓扑结构,在精确的自复制过程中创造出不精确的“误差”,以调整网络拓扑及性能,这一过程便会通过网络特征参数直观表现出来.同时,在本节的分析中观测到,探测期间网络拓扑与性能是在不断优化的,本文将该现象理解为由于在代谢与复制过程中出现了“误差”,通过系统的不断调适而导致具有生命特性的系统呈现出“优化”或“进化”,但对于这一猜想仍需进一步的研究证实.

5 结 论

本文借鉴Eigen超循环理论对生命特征的判定

依据,将其应用到互联网宏观拓扑数据中,创新性地实现了以生物学思想启发互联网宏观拓扑研究.通过对2010—2014年IPv4与IPv6互联网全网拓扑数据统计分析,并分别以标准网络结构熵、K-核网络度分布幂指数及网络聚集系数和平均路径长度分别量化3个判据:代谢行为、自复制及自复制过程中的误差,得出如下结论:

1) 探测期间,IPv4与IPv6互联网标准网络结构熵不断增大且增长幅度不同,比较结果为 $IPv4 > IPv6$. IPv6网络宏观拓扑较IPv4具有更强的拓扑演化秩序性,IPv6的网络拓扑具有更强的发展潜力和更强的系统代谢功能.但这只是相对而言,由于量级之微,2种协议下互联网均为有序代谢的拓扑结构,均具有良好的代谢功能.宏观上,互联网拓扑结构是一个由混沌态向稳态发展的过程,期间的过程通过一次次小涨落不断循环完成,每一次稳态中秩序的产生都是通过系统自组织的代谢行为推动的.

2) IPv4与IPv6互联网1-25核网络在时序上(2010—2014年)度分布频度基本不变.互联网IPv4与IPv6网络均具有拓扑自复制能力,K-核网络拓扑不会随时间变化,而是会在系统发展与演化的过程中保留原有的信息结构.正是这种保留原有信息结构稳定不变的自复制能力,在根本上保证了互联网的新生拓扑结构能在最短时间内形成,并能够保证系统正常运行.

3) IPv4及IPv6互联网网络平均聚集系数与平均路径长度均在探测期间以小涨落的形式不断变化.拓扑内部建立更多、更复杂的连边,调整网络的连通性,形成网络拓扑复制演化过程中的微小变化.对比发现,探测期间IPv6网络较IPv4更善于在自复制过程中进行调整,通过调整而演化出的网络拓扑具有更优良的结构与性能.同生物领域对遗传过程中产生的误差一样,互联网发展演化过程中表现出的不精确性,是互联网形成新结构的主要信息来源,是系统“进化”的动力.

IPv4与IPv6互联网均具有生命特征,且由于IPv6网络采取了改进网络通信协议、简化数据包路由策略同时改进区域组网和建设并简化首次加入网络时的配置过程等策略,这使得IPv6不仅解决了IPv4地址分配殆尽、可供扩展的空间逐渐萎缩这个问题,也让IPv6网络在进化过程当中拓扑结构与运行性能更佳,生命特征更明显、更具发展潜力.从复杂人工系统(互联网宏观拓扑)中提取生命特征,是将2个学科联通、融合的桥梁,这使借鉴生物学思想指导互联网宏观拓扑研究成为可能,为未来借鉴生

物理学思想指导互联网研究提供依据. 以此为理论基础, 将互联网理解为具有自组织、自适应过程的宏大智慧生命体, 可以将互联网的演化过程视作生物进化; 在对互联网的演化分析过程中参考 Darwin 的生物进化理论; 于不同尺度下观测到的网络拓扑具有与生态系统、物种、种群、个体相类似的区别与联系; 借鉴生态学中对生态系统的调整与干预手段. 从该角度出发不仅有助于深入对互联网宏观拓扑的理解, 也可以实现对互联网未来结构与功能的预测, 操控并引导未来互联网的进化方向, 更有效地实现网络功能.

参 考 文 献

- [1] Albert R, Jeong H, Barabási A L. Internet: Diameter of the world-wide Web [J]. *Nature*, 1999, 401(6749): 130-131
- [2] Yook S H, Jeong H, Barabási A L. Modeling the Internet's large-scale topology [J]. *Proceedings of the National Academy of Sciences*, 2002, 99(21): 13382-13386
- [3] Oliveira R V, Zhang B, Zhang L. Observing the evolution of Internet as topology [C] //Proc of ACM SIGCOMM Computer Communication Review. New York: ACM, 2007, 37(4): 313-324
- [4] Fay D, Haddadi H, Thomason A, et al. Weighted spectral distribution for Internet topology analysis: Theory and applications [J]. *IEEE/ACM Trans on Networking*, 2010, 18(1): 164-176
- [5] Ahn Y Y, Bagrow J P, Lehmann S. Link communities reveal multiscale complexity in networks [J]. *Nature*, 2010, 466(7307): 761-764
- [6] Zhang G Q, Wang D, Li G J. Enhancing the transmission efficiency by edge deletion in scale-free networks [J]. *Physical Review E*, 2007, 76(1): 017101-1-017101-4
- [7] Langton C G. *Artificial Life: An Overview* [M]. Cambridge, MA: MIT Press, 1997: 179, 319
- [8] Gabbay D M, Thagard P, Woods J. *Philosophy of Biology* [M]. Amsterdam: Elsevier, 2007: 585-601
- [9] Zhang Wenbo. Research on the life characteristic of Internet macroscopic topology [D]. Shenyang: Northeastern University, 2006 (in Chinese)
(张文波. Internet 宏观拓扑结构的生命特征研究[D]. 沈阳: 东北大学, 2006)
- [10] Levy A Y, Weld D S. Intelligent Internet systems [J]. *Artificial Intelligence*, 2000, 118(1): 1-14
- [11] Brent E. *The Sage Handbook of Online Research Methods* [M]. London: SAGE, 2008: 452-468
- [12] Song C, Havlin S, Makes H A. Self-similarity of complex networks [J]. *Nature*, 2005, 433(7023): 392-395
- [13] Schneider C M, Kesselring T A, Andrade Jr J S, et al. Optimal box-covering algorithm for fractal dimension of complex networks [J]. *arXiv preprint arXiv:1204.3010*, 2012
- [14] Barabási A L, Albert R. Emergence of scaling in random networks [J]. *Science*, 1999, 286(5439): 509-512
- [15] Eigen M. Selforganization of matter and the evolution of biological macromolecules [J]. *The Science of Nature*, 1971, 58(10): 465-523
- [16] Eigen M, Schuster P. *The Hypercycle: A Principle of Natural Self-Organization* [M]. Berlin: Springer, 2012: 7-8, 24-58
- [17] Smith J M. Hypercycles and the origin of life [J]. *Nature*, 1979, 280(5722): 445-446
- [18] Lee D H, Severin K, Yokobayashi Y, et al. Emergence of symbiosis in peptide self-replication through a hypercyclic network [J]. *Nature*, 1997, 390(6660): 591-594
- [19] Lewis T G. *Network Science: Theory and Applications* [M]. Hoboken, NJ: John Wiley & Sons, 2011: 124-126, 136-137
- [20] Tan Yuejin, Wu Jun. Network structure entropy and its application to scale-free networks [J]. *Systems Engineering-Theory & Practice*, 2004, 24(6): 1-3 (in Chinese)
(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)
- [21] Luo Peng, Li Yongli, Wu Chong. Complex networks evolution research using the network structure entropy [J]. *Complex Systems and Complexity Science*, 2013, 10(4): 62-68 (in Chinese)
(罗鹏, 李永立, 吴冲. 利用网络结构熵研究复杂网络的演化规律[J]. *复杂系统与复杂性科学*, 2013, 10(4): 62-68)
- [22] Zhang G, Quoitin B, Zhou S. Phase changes in the evolution of the IPv4 and IPv6 AS-Level Internet topologies [J]. *Computer Communications*, 2011, 34(5): 649-657
- [23] Wang Xiaofan, Li Xiang, Chen Guanrong. *The Theory and Application of Complex Network* [M]. Beijing: Tsinghua University Press, 2006: 10-53 (in Chinese)
(汪小帆, 李翔, 陈关荣. *复杂网络理论及其应用*[M]. 北京: 清华大学出版社, 2006: 10-53)
- [24] Guo Shize, Lu Zheming. *The Basic Theory of Complex Network* [M]. Beijing: Science Press, 2012: 41-72 (in Chinese)
(郭世泽, 陆哲明. *复杂网络基础理论*[M]. 北京: 科学出版社, 2012: 41-72)
- [25] Alvarez-Hamelin J I, Dall'Asta L, Barrat A, et al. Large scale networks fingerprinting and visualization using the k -core decomposition [C] //Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2005: 41-50
- [26] Bollobás B. *Random Graphs* [M]. Berlin: Springer, 1998: 447-457
- [27] Pleschiutchnigg F P, Parschat L, Rahmfeld W, et al. Secondary cooling systems for slab casters: From air mist to dry casting [J]. *Iron Steel Engineering*, 1987, 64(1): 51-55
- [28] Wang Q, Liu T, Gao A, et al. A novel method for in situ formation of bulk layered composites with compositional gradients by magnetic field gradient [J]. *Scripta Materialia*, 2007, 56(12): 1087-1090

[29] Schneider E D, Kay J J. Life as a manifestation of the second law of thermodynamics [J]. Mathematical and Computer Modeling, 1994, 19(6): 25-48

[30] Kauffman S. At Home in the Universe: The Search for the Laws of Self-Organization and Complexity [M]. Oxford, UK: Oxford University Press, 1995

[31] Zhang H H, Zhao H, Cai W, et al. A qualitative method for analysis the structure of software systems based on k -core [J]. Special Issue on Software Engineering and Complex Networks of Dynamics of Continuous, Discrete and Impulsive Systems Series B, 2007, 14(S6): 18-24

[32] Ash J, Newth D. Optimizing complex networks for resilience against cascading failure [J]. Physica A: Statistical Mechanics and Its Applications, 2007, 380(2007): 673-683

[33] Albert R, Barabási A L. Statistical mechanics of complex networks [J]. Reviews of Modern Physics, 2002, 74(1): 47-94



Liu Xiao, born in 1986. PhD candidate. Her main research interests include complex networks theory, evolution of the Internet topology and the complex ecosystem.



Zhao Hai, born in 1959. Professor and PhD supervisor. His main research interests include pervasive computing, network science, embedded operating system and data fusion (zhaoh@mail.neu.edu.cn).



Li Shaofeng, born in 1992. Master candidate. His main research interests include complex network, the complexity of Internet and energy Internet (shaofengli@hnu.edu.cn).



Wang Jinfa, born in 1988. PhD candidate. His current research interests include complex networks, Internet topology analysis and data science (jinfa.wang@qq.com).



Li Hequn, born in 1989. PhD candidate. His main research interests include Internet measurement and complex network (lihequn@outlook.com).