

doi : 10. 3969/j. issn. 1005 - 3026. 2016. 09. 009

林木虫害大数据的网络科学分析方法

刘 晓^{1,2},赵 海¹,冯 颖^{1,3},何 璇¹

(1. 东北大学 计算机科学与工程学院,辽宁 沈阳 110819 ;2. 杜伦大学 生物与生物医学学院,杜伦 DH1 3LE ;
3. 辽宁林业职业技术学院,辽宁 沈阳 110101)

摘 要 : 择取国家森防总站 2009—2013 年辽宁省林木虫害大数据 根据林木虫害时空复杂性 提出一种基于时空影响域的虫害关系网络构造方法. 以昆虫生活习性确定时间影响范围 以虫害危害等级确定空间影响范围 并将松毛虫作为研究对象. 结果表明 松毛虫虫害关系网络为无标度网络 服从幂律分布 松毛虫传播扩散快 虫害易聚集发生 网络拓扑具有鲁棒性. 该建网方法能够反映真实世界 是解读林木虫害大数据的有效方法. 期待通过本文对松毛虫虫害关系网络的复杂网络理论分析 能对实际林木虫害防治工作中防治策略的制定及防控力度的估计提供理论指导.

关 键 词 : 网络科学 ;虫害关系网络 ;时空影响域 ;辽宁林木虫害 ;松毛虫

中图分类号 : TP 393 文献标志码 : A 文章编号 : 1005 - 3026(2016)09 - 1254 - 05

Analysis Method of Network Science Based on Forestry Pest Big Data

LIU Xiao^{1,2} , ZHAO Hai¹ , FENG Ying^{1,3} , HE Xuan¹

(1. School of Computer Science & Engineering , Northeastern University , Shenyang 110819 , China ;2. School of Biological & Biomedical Sciences , Durham University , Durham DH1 3LE , UK ;3. Liaoning Forestry Vocation-Technical College , Shenyang 110101 , China. Corresponding author : LIU Xiao , E-mail : xiao_6xiao@126. com)

Abstract : Big data of Liaoning forestry pests was adopted , which published by General Station of Forest Pest Management , State Forestry Administration. A construction method based on space-time influence domain for insects network was proposed according to the insects occurring complexity in space and time. Taking pine caterpillar (*Dendrolimus spp.*) as a research sample , the expansion of the time window is determined according to the insects ' lifestyles and habits , and the range of influence effect are determined according to the damage of pest. The results show that the proposed pine caterpillar network following the power-law distribution is scale-free , *Dendrolimus* diffuse fast , the pine caterpillars are more likely to emerge clustered , and the topology of the pine caterpillar network is robust. This analysis which using the complex network method is scientific , and the real-world phenomenon can be reflected in the network construction method. Such a analysis of network science on pine caterpillar network is intended to provide a guidance regarding to forestry pest control strategies.

Key words : network science ; insects network ; space-time influence domain ; Liaoning forestry pests ; pine caterpillar (*Dendrolimus spp.*)

森林生态系统是一个较接近自然状态的复杂生态系统,在该系统中,昆虫等各要素间关系紧密,各种昆虫波幅相对稳定,因此对森林生态系统较易实行生态调控^[1]. 不同时期,生态学家们对生态格局和生态系统各要素间互作过程的认知不同^[2],对生态系统及生物网络的研究,目前只是从生态学与生物学等不同角度进行^[3]. 森林生态系统中任何一部分的演化又表现出复杂的拓扑动

力学行为,而传统分析虫害发生关系及森林生态系统的方法,未从全局角度对其进行分析.对于从系统上层构建宏观生物网络拓扑、网络特性、演化趋势与演化动力等方面的研究,国内外仍较少见.

复杂网络理论^[4]是认识复杂系统的关键,从宏观角度研究系统结构和功能,系统内在联系及拓扑动力机制.基于生态学原理,从复杂网络的角度探讨生态系统及昆虫间的关系,研究其拓扑演化动力原理,发挥生态系统中自然因素调控作用,对维持生态系统稳定,保护林木、种群生态,以及生态系统都具有重要意义^[5].

1 昆虫网络构造方法

1.1 时空影响域

定义 1 时空影响域:某地发生虫害,该地对周围的影响范围及其持续影响的时间窗口,定义为时空影响域.其中,空间影响半径,指虫害发生后,相同生态位昆虫间由于竞争产生种群取食压力,当地昆虫种群会向周围空间范围扩散,对周围林地造成影响的范围;时间窗口,指当虫害在某一时间发生后,害虫种群在当地及其空间影响范围内造成影响的时间长度.

需明确的是,虫害发生愈严重,生态压力愈大,其向周围空间扩散的能力愈强,即在空间上产生影响范围就愈大;另外,时间窗口长度为昆虫危害属性,不受虫害发生严重程度影响(对于同类害虫,在时间上造成持续影响的时间窗口长度相等).

首先根据调查结果,确定林木虫害危害程度及面积,通过式(1)统计林木虫害危害等级指数 M ^[6]:

$$M = \sum_{i=1}^3 K_i \times \lg(15 \times S_i). \quad (1)$$

式中: K_i 为林木虫害发生的第*i*种危害程度指标,共有 3 种($i=1, 2, 3$); S_i 为第*i*种危害程度下林木虫害发生面积.该 S_i 值的采集以国家林业局颁布的《森林病虫害预测预报管理办法》中对林木病虫害危害程度等级分级标准为依据,将虫害对森林造成的损失程度划分为轻、中、重三级(分别对应*i*=1, 2, 3).

林木虫害等级指数 M (即虫害发生严重程度)与空间影响半径 L 的关系如下:

$$\lg L = aM + b. \quad (2)$$

式中: L 为虫害影响区域半径; M 为林木虫害危害等级指数; a 与 b 为常量.

1.2 基于时空影响域的昆虫网络模型

定义 2 定义昆虫网络为有向图 $G=(V, E)$,其中 V 为图 G 的节点集 $|V|=n$,每个节点代表林木病虫害防治调查中实际发生的虫害; E 指网络中有向边构成的集合 $|E|=e$,对于任意两节点 $i \in V$ 和 $j \in V$,设 $R_i(T_i, S_i)$ 为节点*i*的时空影响域,那么当 $j \in R_i(T_i, S_i)$,即*j*既在*i*的空间影响范围内又在其时间窗口中时,则存在一条由*i*到*j*的边(i, j).

每次虫害发生都会对时空造成影响,发生一次虫害后该区域会形成以自身为中心的圆形影响区域 S ,而虫害对周围产生持续影响的时间窗口大小为 T .由此,每次害虫发生均会产生一个时空影响区域 $R(T, S)$,在该影响区域内发生的虫害均视为受本次虫害影响所致,而在该影响区域外发生的虫害则与本次虫害无直接关系.本研究中,以虫害发生地经纬坐标定位节点.

如图 1 所示, t_A 时刻虫害*A*发生形成影响域 $R_A(T_A, S_A)$, t_D 时刻虫害*D*发生,由于 $t_D > t_A + T_A$ (*D*处在虫害*A*影响时间窗口外),即*A*地的虫害在*D*发生之前种群就已消失,无论*D*是否在*A*的空间影响半径 S_A 内,均被认作虫害*A*对虫害*D*无直接影响,即无连边;另一种情况, t_B 时刻*B*发生虫害,之后在 t_C 时刻*C*也发生虫害,由于*C*既在*B*的空间影响半径内又在其产生的影响时间窗口中,即*C*在*B*的时空影响域内,*B*对*C*有直接影响,即存在 1 条由*B*指向*C*的有向边.

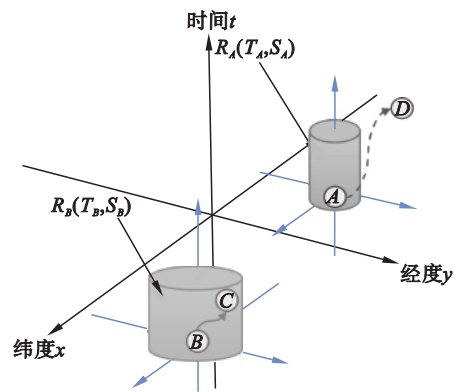


图 1 基于时空影响域构建网络示意图
Fig. 1 Schematic diagram of constructing a network based on space-time influence domain

以图 2 为例构建昆虫网络,根据式(1)(2)统计虫害发生后形成的空间影响半径如图 2a 中圆形区域所示,圆形区域面积因虫害发生轻重程度不同而异,重叠部分说明地区间虫害存在影响;图 2b 中虫害发生时间顺序为*A*到*I*,假设该类虫害时间窗口为 3 个时间单位,如图中条带覆盖长

度所示,即在 0 时刻发生的虫害所造成的影响只能持续到时刻 2. 那么根据定义 1,示例网络拓扑如图 2c. ρ 时刻 A 处发生虫害在时间上形成的影响会持续到时刻 2,但其影响区域并未覆盖至时刻 2 发生的虫害 B ,因此无连边; B, C 在相同地点,可看作同一节点,且 C 发生在 B 产生的时间影响窗口内,因此产生一条自连边;网络中其他节点及连边关系以此类推.

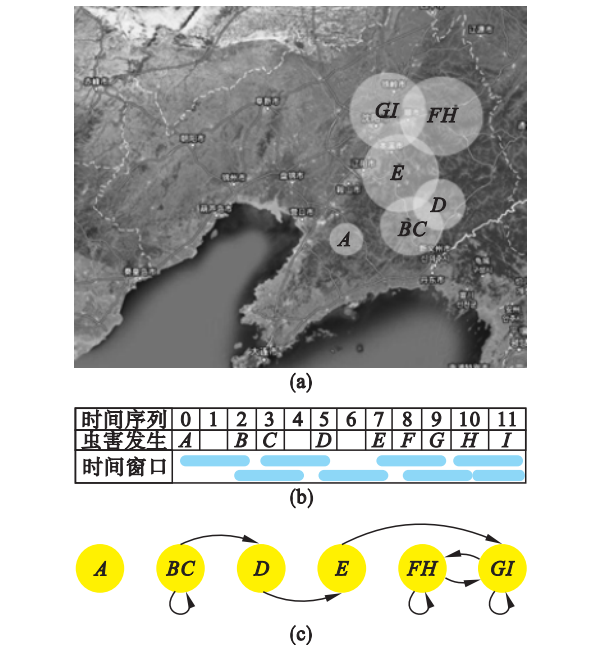


图 2 基于时空影响域构建昆虫网络
Fig. 2 Schematic diagram of constructing the insect network based on space-time influence domain
(a)—空间影响半径示意图;(b)—时间影响窗口示意图;
(c)—基于时空影响域的昆虫网络拓扑.

2 数据选取

本文数据采自 2009—2013 年辽宁省林区,辽宁省地处中国东北地区南部,位于东经 $118^{\circ}53'$ 至 $125^{\circ}46'$ 北纬 $38^{\circ}43'$ 至 $43^{\circ}36'$ 之间,东西约 550 km,南北约 550 km. 全省森林面积约 418 500 km^2 ,共设检测点 73 个,拟监测林地面积 155 411.13 km^2 ,实际监测面积 153 565.33 km^2 . 数据由国家森防总站统一采集,监测记录初始数据按照国家《森林病虫害防治统计报表》及统计指标说明规定进行整理.

根据调查结果,辽宁省内发生较重的针叶林叶部害虫为落叶松毛虫(*Dendrolimus superans*)、油松毛虫(*Dendrolimus tabulaeformis*)、赤松毛虫(*Dendrolimus spectabilis*),均属枯叶蛾科(*Lasiocampidae*)昆虫,为我国北方常见的林木害

虫. 其发生量大,生活习性与生态位类似,本文以此类害虫为研究对象,建立辽宁林地松毛虫虫害关系网. 同时,对时间影响窗口及空间影响半径做如下说明:设虫害造成影响持续时间为 12 个月,1 年当中无论虫害发生在哪个月份,所形成的时间窗口均为 12 个月;另外,由于冬季所有种群消亡,松毛虫只会以越冬虫卵形式越冬并在次年发生,故当时间窗口跨年时,次年只有松毛虫首次发生的月份(3、4 月)会受前一年影响,之后则不会受前一年的影响^[7];根据松毛虫生活习性,最长迁飞距离不超过 37 km ^[8],故设空间影响域半径 L 上限为 37 km.

在该昆虫网络以网络科学手段分析之前,首先给出如下复杂网络特征量定义.

定义 3 对于幂律分布有 $p(k) \sim k^{-\alpha}$,其中 $p(k)$ 是度值为 k 的节点出现的概率, α 为幂指数^[9]. 度分布的幂律现象愈明显^[10],网络愈呈现无标度特性,网络的鲁棒性就愈强^[11].

定义 4 平均路径长度:网络的平均路径长度 L (或平均最短路径)是网络中任意节点对之间最短路径长度的平均值,它是衡量网络转发通信能力的重要参数. 较短的平均路径长度使网络具备更快的传输速度^[12].

$$L = \frac{1}{N(N-1)} \sum_{i,j} d_{ij}, \quad (3)$$

式中 d_{ij} 为节点 i 和 j 间的最短路径.

定义 5 聚集系数,描述网络节点集团化程度. 网络中单个节点的聚集系数 C_i 是该节点与其邻居节点之间边的连接数与所有邻居间存在最大连边数目之比^[13]:

$$C_i = \frac{2E_i}{k_i(k_i-1)}. \quad (4)$$

其中 E_i 是节点 i 与其邻居节点之间边的数量, k_i 是节点 i 邻居节点的数量.

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两节点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. 网络聚类系数越大,其邻居节点间连通性越强.

$$C = \frac{1}{N} \sum_{i=1}^N C_i. \quad (5)$$

3 辽宁林木虫害网络

根据国家森防总站提供的 2009—2013 年辽宁林木病虫害发生防治统计数据,建立辽宁省枯

叶蛾科昆虫(落叶松毛虫、油松毛虫和赤松毛虫)虫害网络 根据定义 3~5 统计松毛虫网络基本特征量,以复杂网络常规研究手段,对其进行如下分析讨论.

根据时间、空间影响域模型建立松毛虫网络,网络中共 214 个节点,1 121 条有向边.节点颜色根据节点间关系紧密程度及节点间集团化程度的不同进行区分,网络拓扑如图 3 所示.2 个节点(以正方形框出)较其他节点度值更高,与其他节点连边更多.根据节点信息,将两次虫害定位于盖州市-赤松毛虫虫害和凌海市-赤松毛虫虫害,且据此可以判断这两处林地松毛虫虫害发生较为严重,松毛虫在这两处与周围林地间活动频繁,对外界造成明显影响.在实际森防工作中,可以对此类重要节点进行重点防治,且可以尝试把防治工作向“重要节点与外部节点的连边的阻断和干扰上”偏转.由此可见,松毛虫虫害关系网络从网络拓扑中分析虫害关系的结构模式与内在属性,是一种解读林木虫害大数据的有效方法.

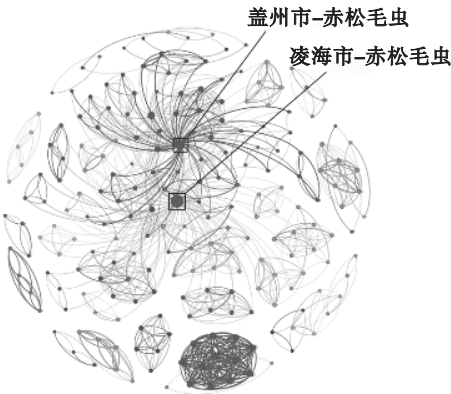


图 3 辽宁省林木松毛虫网络拓扑图

Fig. 3 Topology of *Dendrolimus* network of Liaoning forest

3.1 无标度特征

网络拓扑特征存在于节点与边的关系之中,幂律分布归纳的正是复杂网络边的分布特征.根据定义 3 统计该松毛虫网络节点度分布频数,双对数坐标下网络节点度分布频度如图 4 所示.图中直线表示以幂律对其进行拟合的结果,幂指数 $\alpha = 1.88$,拟合优度 $R^2 = 0.93$,说明节点度分布符合幂律分布,网络为无标度网络.幂指数 $\alpha = 1.88$ 说明网络中存在一定数量的大度值“中心节点”,图 3 中标出的两个节点正是此类节点,这类节点发生虫害重,对周围林地造成较广泛影响;由松毛虫网络的无标度特征,可知网络中除了一定数量度值很高的“中心节点”,小度值节点也大量存在,网络拓扑具有一定鲁棒性.

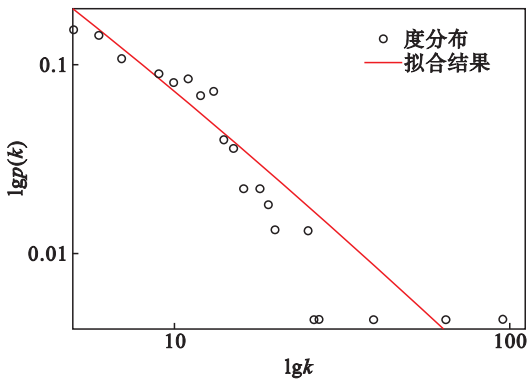


图 4 松毛虫网络节点度分布频度图

Fig. 4 Degree distribution of *Dendrolimus* network

可见,松毛虫网络为无标度网络,网络具有鲁棒性,节点受到外界干扰时仍能保持其拓扑结构或性能完好无损或几乎不变.那么,当某地发生松毛虫虫害,只对发生地的松毛虫虫害采取防治手段,对松毛虫网络拓扑的影响作用很小,防治很难达到理想效果.从本文提出的方法所构建的昆虫网络中观察到幂律特性,说明以该手段提取的网络拓扑可以作为真实世界自然现象与事件的真实反映,是将昆虫网络进行复杂网络分析的可实现的科学手段.

3.2 平均路径长度

网络的平均路径长度 L 除了衡量网络通讯和传输能力,也是衡量网络鲁棒性的重要拓扑特征量. L 越短,信息传输速度越快,网络中微小扰动所造成致命打击的可能性越小,即网络拓扑鲁棒性越好^[11,14].根据式(3),对 2009—2013 年辽宁省林木松毛虫网络平均路径长度进行统计分析.

松毛虫网络平均路径长度较短, $L = 1.751$,网络中各节点间虫害传输速度较快,说明松毛虫在网络中传播效率高,某地发生松毛虫虫害后,在形成的空间影响半径范围内传播快;同时,由于松毛虫网络平均路径长度较短,说明该网络鲁棒性高,虫害发生之后,在外界作用下,无论对节点的控制或对连边的阻碍作用,都对网络拓扑无法形成打击或影响.由此可见,林木松毛虫害具抗打击性,在林木松毛虫的防治工作中应该注意到,由于松毛虫在林地扩散速度较快,针对孤立节点的防治措施无法达到全面防治的目的.

3.3 平均聚集系数

由定义 5 可知,网络平均聚集系数 C 既是网络中节点的宏观聚集程度,也可衡量网络拓扑鲁棒性^[14-15].网络的聚集系数越大,局部连通性越强,节点间互通获得资源的机会增多,这在一定程

度上可以使网络忽视掉由个别节点失效而引起的一系列影响网络拓扑与性能的问题. 已有研究表明, 真实世界中的复杂网络, 节点往往呈现出非常明显的高聚集特性^[16]. 根据式(4)和式(5)计算并分析 2009—2013 年辽宁省林木松毛虫网络的网络平均聚集系数.

辽宁省林木松毛虫网络平均聚集系数 $C = 0.713$, 聚类系数较高, 说明松毛虫网络节点之间局部连通性较好, 害虫发生地会对周围造成较广泛的影响, 之后周围林地之间也会交互影响. 网络中节点与其邻居节点间的相互影响关系倾向于构造一种密不可分的连接关系, 即松毛虫虫害的发生会呈现一定的聚集现象, 同时, 网络由于拓扑结构紧密, 所以对外部微扰具有一定的自愈能力, 即鲁棒性. 这不仅可以指导森防策略的制定, 也进一步对森防措施的力度提出了要求.

4 结 论

本文基于时间与空间影响模型, 提出一种基于时空影响域的松毛虫虫害关系网络构造方法, 根据松毛虫生活习性确定时间影响范围, 根据虫害危害等级确定空间影响半径. 通过对 2009—2013 年辽宁省林木松毛虫虫害数据的统计分析, 建立松毛虫网络, 同时以复杂网络研究手段分析松毛虫网络特征量. 松毛虫网络具有幂律特性, 为无标度网络, 松毛虫虫害在网络中传播迅速, 向周围林地传播速度快, 松毛虫网络局部连通性高, 虫害较易聚集发生, 松毛虫网络具有鲁棒性, 单一的防治手段无法有效控制虫害网络拓扑.

综上, 在实际的松毛虫防治工作中, 应该注意松毛虫的发生及扩散具抗打击性, 不能采取孤立的“点”式治理措施, 而应通过建立网络拓扑寻找较为重要的虫害发生节点, 采取“网”式治理手段. 同时, 在对松毛虫虫害“节点”进行防治的同时, 也应该注意对重要节点间“边”的阻断和干预. 今后的研究可以此为基础, 通过“重构”以及“解构”昆虫网络拓扑, 从拓扑动力学角度出发探索虫害发生扩散行为, 观测分析森林生态系统下物种间的互作行为, 并尝试干预其间的相互作用, 使森林生态系统具有更好的自我调节能力和稳定性, 从而使其最大、最充分地持续发挥其经济、生态效益和作用.

参考文献:

[1] 孔建, 王海燕, 赵白鸽, 等. 苹果园主要害虫生态调控体系

的研究[J]. 生态学报, 2001, 21(5): 789–794.
(Kong Jian, Wang Hai-yan, Zhao Bai-ge, et al. Study on ecological regulation system of the pest control in apple orchard[J]. *Acta Ecologica Sinica*, 2001, 21(5): 789–794.)

[2] McIntosh R P. Concept and terminology of homogeneity and heterogeneity in ecology[C]// Kolasa J, Pickett S T A. *Ecological Heterogeneity*. Ecological Studies 86. Berlin: Springer-Verlag, 1991: 24–46.

[3] May R M. Models for single populations[C]// *Theoretical Ecology*. Oxford: Blackwell Scientific Publications, 1976: 5–29.

[4] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. *Reviews of Modern Physics*, 2002, 74(1): 47–97.

[5] 戈峰, 李典谟. 可持续农业中的害虫管理问题[J]. 昆虫知识, 1997, 34(1): 39–45.
(Ge Feng, Li Dian-mo. Pest management problem in sustainable agriculture[J]. *Entomological Knowledge*, 1997, 34(1): 39–45.)

[6] 陈凤学, 赵杰, 耿海东. 森林病虫害灾害等级研究[J]. 中国森林病虫, 2004, 23(3): 4–7.
(Chen Feng-xue, Zhao Jie, Geng Hai-dong. Classification of disaster grades of forest pests[J]. *Forest Pest and Disease*, 2004, 23(3): 4–7.)

[7] 侯陶谦. 中国松毛虫[M]. 北京: 科学出版社, 1987: 135–162.
(Hou Tao-qian. The pine caterpillars in China[M]. Beijing: Science Press, 1987: 135–162.)

[8] 车锡冰, 夏乃斌, 宋长义, 等. 油松毛虫成虫迁飞特性的研究[J]. 北京林业大学学报, 1996, 18(3): 61–65.
(Che Xi-bing, Xia Nai-bin, Song Chang-yi, et al. Study on the flight activity of dendrolimus tabulae formis adults[J]. *Journal of Beijing Forestry University*, 1996, 18(3): 61–65.)

[9] Newman M. Networks: an introduction[M]. Oxford: Oxford University Press, 2010: 480–481.

[10] 王林, 戴冠中. 复杂网络的度分布研究[J]. 西北工业大学学报, 2006, 24(4): 405–409.
(Wang Lin, Dai Guan-zhong. On degree distribution of complex network[J]. *Journal of Northwestern Polytechnical University*, 2006, 24(4): 405–409.)

[11] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. *Reviews of Modern Physics*, 2002, 74(1): 47–97.

[12] Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations[C]// *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*. New York: ACM, 2005: 177–187.

[13] Newman M. Networks: an introduction[M]. Oxford: Oxford University Press, 2010: 307–310.

[14] Ash J, Newth D. Optimizing complex networks for resilience against cascading failure[J]. *Physica A: Statistical Mechanics and Its Applications*, 2007, 380(7): 673–683.

[15] Zhang J, Zhao H, Xu J Q, et al. Characterizing and modeling the Internet router-level topology—the hierarchical features and HIR model[J]. *Computer Communication*, 2010, 33(16): 2001–2011.

[16] Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks[J]. *Nature*, 1998, 393(6684): 440–442.