

### 基于加权 $K$ -阶传播数的节点重要性

黄丽亚 汤平川 霍宥良 郑义 成谢锋

#### Node importance based on the weighted $K$ -order propagation number algorithm

Huang Li-Ya Tang Ping-Chuan Huo You-Liang Zheng Yi Cheng Xie-Feng

引用信息 Citation: *Acta Physica Sinica*, 68, 128901 (2019) DOI: 10.7498/aps.68.20190087

在线阅读 View online: <https://doi.org/10.7498/aps.68.20190087>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

---

### 您可能感兴趣的其他文章

#### Articles you may be interested in

基于复杂网络动力学模型的无向加权网络节点重要性评估

Evaluation methods of node importance in undirected weighted networks based on complex network dynamics models

物理学报. 2018, 67(9): 098901 <https://doi.org/10.7498/aps.67.20172295>

基于多重影响力矩阵的有向加权网络节点重要性评估方法

Evaluation method of node importance in directed-weighted complex network based on multiple influence matrix

物理学报. 2017, 66(5): 050201 <https://doi.org/10.7498/aps.66.050201>

基于层间相似性的时序网络节点重要性研究

Node importance identification for temporal network based on inter-layer similarity

物理学报. 2018, 67(4): 048901 <https://doi.org/10.7498/aps.67.20172255>

基于领域相似度的复杂网络节点重要度评估算法

Node importance measurement based on neighborhood similarity in complex network

物理学报. 2017, 66(3): 038902 <https://doi.org/10.7498/aps.66.038902>

加权网络中基于冗余边过滤的 $k$ -核分解排序算法

A ranking approach based on  $k$ -shell decomposition method by filtering out redundant link in weighted networks

物理学报. 2017, 66(18): 188902 <https://doi.org/10.7498/aps.66.188902>

节点中心性对复杂网络传播模式的影响分析

Analysis of the effect of node centrality on diffusion mode in complex networks

物理学报. 2017, 66(12): 120201 <https://doi.org/10.7498/aps.66.120201>

# 基于加权 $K$ -阶传播数的节点重要性\*

黄丽亚<sup>1)2)†</sup> 汤平川<sup>1)</sup> 霍宥良<sup>1)</sup> 郑义<sup>1)</sup> 成谢锋<sup>1)2)</sup>

1) (南京邮电大学, 电子与光学工程学院, 微电子学院, 南京 210023)

2) (射频集成与微组装技术国家地方联合工程实验室, 南京 210003)

(2019年1月16日收到; 2019年4月5日收到修改稿)

节点重要性对于分析网络结构具有重要意义. 为了充分刻画网络全局和局部特性, 本研究基于网络拓扑结构对疾病传播过程进行了抽象, 分别设置各个节点为传染源, 在经历传播时长  $K$  后, 将网络中已感染节点的数量定义为  $K$ -阶传播数, 最终基于不同  $K$  值下的  $K$ -阶传播数得到节点重要性结果. 对 Watts-Strogatz 小世界网络和海豚网络的仿真实验表明, 加权  $K$ -阶传播数法对节点重要性的评估较其他方法更为合理, 能够细致地刻画小世界网络中长途连接对信息传输的影响, 提高海豚网络中对社区交流起关键作用的节点的重视程度. 本文利用蓄意攻击策略对美国西部电网、芝加哥公路网络、网络科学家合著网络以及小鼠神经纤维束网络进行了研究, 即依照节点重要性由高到低的排序依次攻击网络. 结果显示, 相较于其他方法, 基于加权  $K$ -阶传播数法仅需移除少量重要节点便可实现对网络结构的充分破坏.

**关键词:** 复杂网络, 节点重要性, 传播模型**PACS:** 89.70.Cf, 89.75.Fb**DOI:** 10.7498/aps.68.20190087

## 1 引言

复杂网络是现实系统的抽象表达. 网络节点依靠边相互联系, 通常存在着重要性差异. 节点重要性是分析设计网络结构、提升系统鲁棒性等研究的重要基础<sup>[1-3]</sup>. 目前, 已有诸多研究各自从不同的角度提出了节点重要性评价指标.

度中心性法<sup>[4]</sup>认为节点拥有的邻居数越多, 其直接影响力越强. 对食物链网络<sup>[5]</sup>、P2P 网络<sup>[6]</sup>、电子邮件网络<sup>[7]</sup>以及蛋白质网络<sup>[8]</sup>的研究指出, 当度值较大的节点被移除后, 网络结构将变得更加脆弱. 此外, 度中心性法计算简便, 其时间复杂度为  $O(N)$ , 适用于网络规模较大的情况. 然而, 度中心性法未考虑非邻居节点的影响, 利用的信息较为有限, 不能充分地反映网络的全局特性和桥连接节点的重要性. 任卓明等<sup>[9]</sup>在度中心性法的基础上, 将

邻居节点相互连接的紧密程度, 即局部聚类系数也纳入了评价体系 (下文简称 Ren 法), 结果虽优于度中心性法, 但其反映网络全局特性的能力仍然有限. 此外, Ren 法利用趋同化函数将度与聚类系数处理后直接加和, 并未设置比例系数, 即该方法认为二者同等重要, 其合理性有待商榷. 为了更充分地利用网络结构信息, Chen 等<sup>[10]</sup>提出了一种基于多级邻居信息指标的半局部中心测度的方法 (下文简称 Chen 法), 首先确定节点的一级重要性为最近邻与次近邻节点的个数之和, 再计算节点的二级重要性为所有邻居节点的一级重要性之和, 最后定义节点的三级重要性为所有邻居节点的二级重要性之和, 并作为最终的重要性评价指标. 但为了保证较低的算法复杂度, Chen 法仅将分析范围扩展到了次近邻节点, 因而对网络全局特性的挖掘也并非十分充分. 介数中心性<sup>[11]</sup>指网络中所有最短路径通过某节点的占比, 是节点对网络传播信息的影

\* 国家自然科学基金 (批准号: 71671093, 61271334) 资助的课题.

† 通信作者. E-mail: huangly@njupt.edu.cn

响或对节点预期负载的度量. 紧密度<sup>[12,13]</sup>指由某节点出发, 到达其余节点的最短路径的倒数, 是将信息从给定节点传播到网络中其他可达节点的时间的度量. 介数中心性与紧密度虽提高了对桥节点的重视程度, 但需要计算任意一对节点之间的最短路径, 其时间复杂度为 $O(N^3)$ , 不适用于大型网络, 且对随机网络的解释也不够充分. 特征向量法<sup>[14,15]</sup>将网络邻接矩阵最大特征值对应特征向量的元素作为各个节点的重要性指标, 本质上是将单个节点的拓扑性质线性叠加, 结果较为片面. Katz<sup>[16]</sup>认为节点的重要性正比于网络邻接矩阵  $A$  的幂级数  $\sum_{k=0}^{\infty} \alpha^k A^k$  与单位阵  $E$  的差的各列元素之和, 其中  $\alpha$  为权重衰减因子. Katz 指标虽充分利用了网络的全局特性, 但  $\alpha$  却无法定量计算, 只能根据不同的网络进行人为设置, 且该方法还认为节点影响力随路径的增加呈指数形式衰减, 较为主观. 此外, 现实世界中的网络均是有限的, 而 Katz 指标为了得到收敛形式, 使路径长度取值无穷大, 其结果包含了大量的冗余信息. 为了解决 Katz 指标存在的问题, Zhang 等<sup>[17]</sup>以网络节点为变量, 将其余所有节点对当前节点的影响力进行加和, 并假设节点的影响力随路径的增加呈高斯形式衰减. 该方法在一定程度上解决了 Katz 指标的信息冗余等问题, 但节点影响力的衰减形式仍较为主观.  $K$ -核分解法<sup>[18,19]</sup>试图递归地移去网络中度中心性的值小于等于  $K$  的节点, 其时间复杂度为  $O(N)$ , 相比于度中心性、介数等指标更能反映诸如演员网络、社交网络等实际网络的节点重要性. 但  $K$ -核分解法对节点的排序并不细致, 常常赋予大量节点相同的重要度, 不适合于树状网络和无标度网络的分析. PageRank 算法<sup>[20]</sup>认为节点的重要性与随机浏览者访问的频率成正比, 被广泛地应用在了网页排名等领域, 但该算法对含孤立节点或社团结构的网络会出现重要性排序不唯一等问题. 为了解决该弊端, Lü 等<sup>[21]</sup>提出了 LeaderRank 算法, 在原始网络的基础上, 增加了一个与所有节点双向连接的 Ground 节点. 这一操作使网络变为强连通的, 其结果比 PageRank 算法更加准确, 但 LeaderRank 算法不适用于无向网络. Zhong 等<sup>[22]</sup>利用网络节点被移除前后流行阈值的差来表征节点的全局重要性, 并利用度中心性来表征节点的局部重要性, 最后将归一化后的全局和局部重要性结果进行加权求和 (下文简称 CI 法), 并利用美国航空网络等

九个真实网络证明了 CI 法的有效性. 然而 CI 法中全局和局部重要性的加权系数对最终结果的影响较大. 虽然文献<sup>[22]</sup>经仿真分析归纳指出当加权系数近似为 0.5 时得到的节点重要性结果较好, 但缺乏更为深入的理论证明. 此外, CI 法为何选择度中心性而非其他局部重要性指标有待进一步说明.

以上方法均存在各自的不足, 本研究将提出一种新的节点重要性评价方法, 即加权  $K$ -阶传播数法, 并试图利用 WS (Watts-Strogatz) 小世界网络、海豚网络、美国西部电网、芝加哥公路网络、网络科学家合著网络以及小鼠神经纤维束网络进行仿真分析, 以证明该方法的有效性.

## 2 加权 $K$ -阶传播数模型

SI (susceptible-infective), SIS (susceptible-infective-susceptible) 和 SIR (susceptible-infective-removed) 等模型<sup>[23]</sup>被广泛地应用在疾病以及信息传播等领域, 最初均是对某地区疾病传播过程的抽象. 其中, 个体能否恢复健康和是否具有免疫力是造成上述模型差异的重要因素. SI 和 SIS 模型假设个体不具备免疫力, 将人群分为易感染者和已感染者两类. 此外, SI 模型认为个体一旦被传染便永久处于已感染状态, 而 SIS 模型则认为个体被传染后将以一定概率恢复为易感状态, 并会被再次传染. SIR 模型假设已感染者可被治愈且将具有终身性的免疫力, 将人群分为易感染者、已感染者和免疫移出者三类. 然而, 以上模型均假设疾病传播为随机接触, 并未考虑个体间的拓扑关系. 受上述模型启发, 本文将结合复杂网络对已感染者无法恢复健康且不具有免疫力的最为简单的疾病传播过程进行抽象, 最终得到加权  $K$ -阶传播数法来评价节点的重要性, 描述如下.

首先给定无向网络图  $G(V, E)$ , 其中,  $V = \{v_1, v_2, \dots, v_n\}$  为节点集, 共  $n$  个节点, 代表个体;  $E$  为边集,  $e_{ij}$  表示节点  $v_i$  与  $v_j$  之间的边. 这里假设疾病传播过程中该网络的结构不会发生变化, 且已感染者只能传染给与其直接接触的易感染者. 现假设某节点  $v_i$  为已感染者, 与其相邻的易感染者集合记为  $\Gamma(v_i)$ . 对节点  $v_j \in \Gamma(v_i)$  而言,  $v_i$  将以一定概率  $0 \leq p_{ij} \leq 1$  向  $v_j$  传播疾病. 同时,  $v_i$  向  $v_j$  传播疾病需要耗费一定时间  $t_{ij}$ , 通常受到边  $e_{ij}$  的影响. 若除  $v_i$  外,  $v_j$  同时受到其他相邻的已感染者的传播,

还需进行综合考量. 以上描述考虑到了节点间疾病传播的概率以及耗时等因素, 但若网络边是无权的, 且不考虑节点以及边的意义, 则可进一步抽象, 作出以下假设:

- 1) 已感染者会向其相邻的所有易感染者传播疾病;
- 2) 已感染者向其相邻的易感染者传播疾病的耗时相等, 并设为 1;
- 3) 易感染者一旦受到其任一相邻的已感染者的传播, 便转化为已感染者.

在衡量节点的重要性时, 较为常用的方法是分别将各个节点设置为传染源进行疾病传播, 并以网络中所有节点被转化为已感染者的总耗时作为节点重要性的评价指标, 总耗时越少, 则证明节点越重要. 然而, 当网络非连通时, 从不同节点出发最终能够传播到的节点总数未必相同. 为了保证一致性, 另一种节点重要性衡量方法仍是将各个节点设置为传染源, 但比较的是在经历相同的传播时长  $K$  后网络中已感染者的数量, 数量越大则证明该节点越重要. 基于假设 2 可将传播时长  $K$  取值离散, 并规定为非负整数, 这与 SI 等模型中时间取值连续的假设略有不同. 其中, 当  $K = 0$  时可认为只有传染源节点已被感染, 但尚未开始传播.

此外, 基于假设 1 和 3 可以得到以  $v_i$  为传染源, 传播了时长  $K$  后网络中已感染者的数量为

$$N_{v_i}^K = \left\| \left( \sum_{k=0}^K \mathbf{A}^k \right) \cdot \mathbf{e}_i \right\|_0 = \left\| \sum_{k=0}^K \mathbf{A}^k \cdot \mathbf{e}_i \right\|_0, \quad (1)$$

这里, 将  $N_{v_i}^K$  命名为  $K$ -阶传播数;  $\mathbf{A}$  为网络的邻接矩阵;  $\|\cdot\|_0$  表示向量的  $L_0$  范数;  $\mathbf{e}_i$  表示第  $i$  个分量为 1, 其余分量为 0 的  $n$  维向量. 事实上, (1) 式表示的是矩阵多项式  $\mathbf{A}^0 + \mathbf{A}^1 + \dots + \mathbf{A}^K = \sum_{k=0}^K \mathbf{A}^k$  中第  $i$  行或列向量 (无向图的邻接矩阵  $\mathbf{A}$  为对称阵, 其多项式亦对称) 的  $L_0$  范数, 即非零元素的个数. (1) 式与文献 [24] 中定义的  $K$ -阶邻居数相同, 同样可以衡量  $K$  步内可达的节点总数. 此外, 文献 [24] 指出当  $K$  大于网络最大连通部分的直径  $d$  时, 任意节点的  $N_{v_i}^K$  均不再随  $K$  值变化, 因此有  $K \in \{0, 1, \dots, d\}$ .

传播时长  $K$  的取值是影响节点重要性评价结果的关键. 文献 [24] 利用  $N_{v_i}^K$  并基于信息熵定义了  $K$ -阶结构熵  $H^K$ , 衡量了网络的异构性:

$$H^K = - \sum_{i=1}^n \frac{N_{v_i}^K}{\sum_{j=1}^n N_{v_j}^K} \log \left( \frac{N_{v_i}^K}{\sum_{j=1}^n N_{v_j}^K} \right), \quad K \in \{0, 1, \dots, d\}. \quad (2)$$

结构熵  $H^K$  取值越小, 网络的异构性越强, 并以结构熵序列为依据研究了 WS 小世界、BA 无标度等网络的异构性. 若从疾病传播的角度出发, 可认为  $H^K$  取值越大, 以各个节点  $\{v_1, v_2, \dots, v_n\}$  分别作为传染源的网络  $K$ -阶传播数  $\{N_{v_1}^K, N_{v_2}^K, \dots, N_{v_n}^K\}$  之间的差异越小, 即可认为各节点重要性的差异越小, 反之差异越大. 若仅仅以某单一传播时长下网络中已感染者的数量来衡量节点的重要性, 可能会遗漏其他传播时长下的有用信息. 因此, 本文将对  $K$  取从 0 至  $d$  间的所有时刻进行综合考察, 定义节点  $v_i$  的重要性  $Q_{v_i}$  为

$$Q_{v_i} = \sum_{K=0}^d c^K \cdot S_{v_i}^K, \quad (3)$$

其中

$$S_{v_i}^K = \frac{N_{v_i}^K - \min(N^K)}{\max(N^K) - \min(N^K)}, \quad (4)$$

这里  $N^K = \{N_{v_1}^K, N_{v_2}^K, \dots, N_{v_n}^K\}$ ,  $S_{v_i}^K$  是  $N_{v_i}^K$  归一化后的结果,  $N_{v_i}^K$  通常会随着  $K$  的增大而增大, 为了避免较大的  $N_{v_i}^K$  掩盖取值较小时的信息, 本文将  $N^K$  映射至  $[0, 1]$  区间, 仅考虑节点重要性的相对顺序. 此外, 权重系数  $c^K$  为

$$c^K = 1 - \frac{H^K - \min(H)}{\max(H) - \min(H)}, \quad (5)$$

其中  $H = \{H^0, H^1, \dots, H^d\}$ ,  $H^K$  越小, 权重系数  $c^K$  越大. 本文对 (3) 式中节点重要性差异较大的时刻更为关注并进行放大, 相对忽略差异较小的时刻. 最终可利用  $Q = \{Q_{v_1}, Q_{v_2}, \dots, Q_{v_n}\}$  来衡量各个节点的重要性.

### 3 典型网络的节点重要性

为了验证加权  $K$ -阶传播数法在节点重要性评估方面的优势, 本文将利用 WS 小世界网络、海豚网络、美国西部电网、芝加哥公路网络、网络科学家合著网络以及小鼠神经纤维束网络进行仿真分析.

#### 3.1 WS 小世界网络

小世界网络是指同时具有较短特征路径长度



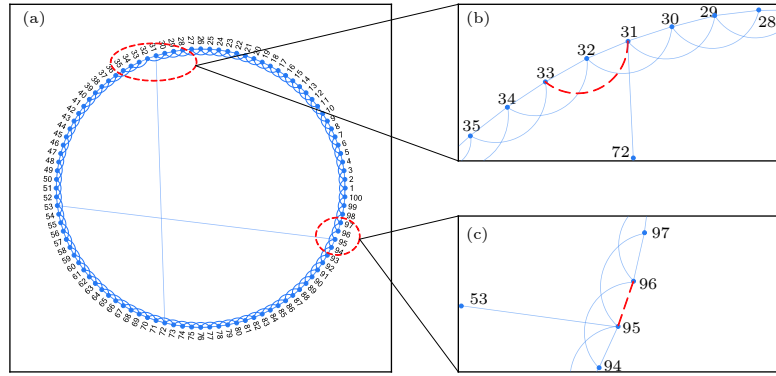


图 1 随机生成的 100 节点 WS 小世界网络 (a) 网络结构; (b) 边 31-33 被改连至边 31-72; (c) 边 95-96 被改连至边 95-53

Fig. 1. A random WS small-world network with 100 nodes: (a) Network structure; (b) the edge 31-33 is rescheduled to the edge 31-72; (c) the edge 95-96 is rescheduled to the edge 95-53.

以及较大平均聚类系数的一种网络类型. Watts 和 Strogatz<sup>[25]</sup> 最早提出了一种构造小世界网络的方法, 即将最近邻耦合网络中的边依概率进行随机重连, 通常将这种网络称为 WS 小世界网络. 图 1(a) 基于 100 节点最近邻耦合网络随机生成了一个 WS 小世界网络, 该最近邻耦合网络中每个节点与其邻近的左右两侧各两个节点相连. 在随机重连后, 边 31-33 和 95-96 分别被改连为 31-72 和 95-53, 见图 1(b) 和图 1(c), 其余边的位置不变.

表 1 列出了该 WS 小世界网络与同规模随机网络的平均聚类系数、特征路径长度等参数. 可见, 该网络的特征路径长度为随机网络的 2.49 倍, 但其平均聚类系数为随机网络的近 20 倍, 这是由于边 31-72 和 95-53 的长程连接提高了网络的连通性.

表 1 WS 小世界网络以及同规模随机网络的网络参数

Table 1. Network features of the WS small-world network and random networks.

网络类型	节点数	边数	平均聚类系数	特征路径长度
WS小世界网络	100	200	0.486	8.4891
随机网络	100	200	0.0254	3.4158

现利用加权  $K$ -阶传播数法对该网络节点的重要性进行分析. 图 2(a) 为网络的  $K$ -阶结构熵  $H^K$ , 图 2(b) 为权重系数  $c^K$ , 图 2(c) 为归一化的节点重要性. 此外, 图 2(c) 中的小图按色度对重要性进行了标注.

由前文所述, 节点 95 和 53, 31 和 72 之间的长程连接提高了网络的连通性, 若移除上述任一节点, 网络的特征路径长度将大为增加, 因此加权  $K$ -阶传播数法认为以上四个节点的重要性最高是较

为合理的. 由于该网络是基于最近邻耦合网络构造的, 而最近邻耦合网络中每个节点与其邻近的左右两侧各两个节点相连. 因此, 与 95, 53, 31, 72 距离相近的左右各两个节点的重要性应当相似, 且距离 95, 53, 31, 72 越远, 节点的重要性越低. 例如, 移除 91, 92, 99 或 98 中的任一节点对网络结构造成的影响是相似的; 此外, 同时移除 99 和 98 相较于同时移除 100 和 1, 会有更多的节点难以通过 95 进行长程通信, 因此 99 和 98 的重要性高于 100 和 1 是较为合理的. 值得注意的是, 图 2(c) 中 97 的重要性显著高于 96, 这是由于边 96-95 被断开, 若依赖 96 进行长程通信需要借助边 95-94, 而 97 与 95 直接相连, 因此通过 97 进行长程通信更为便捷. 此外, 由于边 33-31 被断开, 33, 34, 35 等节点向 30, 29, 28 等节点进行短程通信, 或经由 31 向 72 等节点进行长程通信, 均需要依赖节点 32, 因此其重要性应当较高. 最后, 35, 36, 37, 38 等节点进行长程通信时需要依赖 33 或 34; 对 35, 37 等奇数节点而言, 通过 33 或 34 进行通信的效果相同, 但对 36, 38 等偶数节点而言, 通过 34 进行通信的效率更高, 可见 34 的重要性大于 33, 这也可以解释图 2(c) 中 36 的重要性略高于 35 等现象.

为了进一步证明加权  $K$ -阶传播数法的有效性, 图 3 绘制了度中心性、Ren 法、Chen 法、介数中心性、特征向量法、Katz 指标 (权重衰减因子取 1.5 倍邻接矩阵最大特征值的倒数)、PageRank 算法 (阻尼系数取 0.85) 以及 CI 法 (权重系数取 0.5) 得出的节点重要性结果. 由于  $K$ -核分解法得到的所有节点的重要性相同, 因此其结果未在图 3 给出.

由图 3 可知, 以上方法认为 33 或 96 的重要性

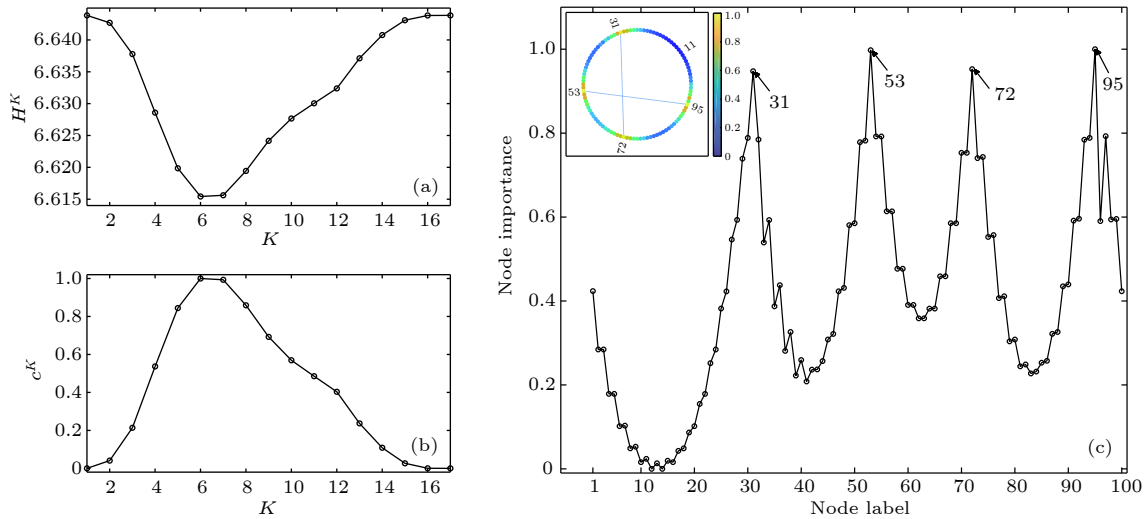


图 2 基于加权  $K$ -阶传播数法的 WS 小世界网络节点重要性结果 (a)  $K$ -阶结构熵; (b) 权重系数; (c) 节点重要性

Fig. 2. Node importance of the WS small-world network based on the weighted  $K$ -order propagation number algorithm: (a) The  $K$ -order structure entropy; (b) the weight coefficient; (c) the importance of nodes.

最低, 而加权  $K$ -阶传播数法却认为 33 和 96 的重要性较高. 由前文分析可知, 虽然在 33 节点移除后, 网络的长程通信不会受到显著影响, 但 27, 28, 29 等节点与 34, 35, 36 等节点间的短程通信, 或 34, 35, 36 等节点依靠 31 进行长程通信只能依赖边 32-34 进行; 同样, 当 96 被移除后, 边 95-97 将参与所有通信, 可见 33 和 96 的存在降低了其余重要节点或边的负载, 起到了分流的作用, 因此认为 33 或 96 在所有节点中重要性最低是值得商榷的; 度中心性、特征向量法、Katz 指标、PageRank 算法和 CI 法认为 72 和 53 的重要性远高于 31 和 95, 但长程通信需要同时依赖以上 4 个节点进行, 因此节点 31 和 72 或 95 和 53 的重要性应当是相似的; 度中心性、Ren 法、Chen 法、Katz 指标、PageRank 算法和 CI 法, 尤其是  $K$ -核分解法对节点的排序并不细致, 存在节点重要性相同的情况; 此外, 介数中心性认为 10, 11, 12 等节点的重要性高于 52, 54, 71, 73 等节点, 49, 51, 55, 57, 74, 76 等节点的重要性高于 50, 52, 54, 56, 73, 75 等节点, PageRank 算法认为 61, 62, 63 等节点的重要性高于 52, 54, 71, 73 等节点, 均缺乏较为合理的解释. 可见, 加权  $K$ -阶传播数法对网络通信的刻画更为细致, 节点重要性的评估优于以上传统方法.

### 3.2 海豚网络

为了进一步验证加权  $K$ -阶传播数法的有效性, 本文以海豚网络为例进行研究. Lusseau 等 [26,27] 对新西兰道特富尔峡湾 (Doubtful Sound) 62 只宽吻

海豚进行研究并构建了海豚社会网络. 基于加权  $K$ -阶传播数法的海豚网络节点重要性结果如图 4 所示, 图 4(a) 为网络的  $K$ -阶结构熵  $H^K$ , 图 4(b) 为权重系数  $c^K$ , 图 4(c) 为归一化后的节点重要性. 网络中节点代表海豚, 边表示海豚间的相关接触.

表 2 列出了基于加权  $K$ -阶传播数法、Ren 法、Chen 法、介数中心性、特征向量法、Katz 指标 (权重衰减因子为 1.5 倍邻接矩阵最大特征值的倒数)、PageRank 算法 (阻尼系数 0.85)、CI 法 (权重系数取 0.5) 得出的前 15 个重要节点的排序结果. 由于度中心性和  $K$ -核分解法存在节点重要性相等的情况, 其排序结果未在表 2 列出. 其中, 基于度中心性法得到的节点重要性排序结果 (前 20 个重要节点) 为  $15 > 38 = 46 > 34 = 52 > 18 = 21 = 30 = 58 > 2 = 14 = 39 = 41 > 10 = 16 = 19 = 37 = 44 = 51 = 55$ ;  $K$ -核分解法认为 1, 2, 7, 8, 9 等 37 个节点的重要性最高且相等.

与度中心性、Ren 法、Chen 法、特征向量法、Katz 指标、PageRank 算法和 CI 法不同, 加权  $K$ -阶传播数法认为节点 37 相对重要, 排在第 7 位. 文献 [27] 将该海豚网络划分成了若干小社区, 指出 37 号海豚曾消失了一段时间, 此时不同海豚社区间的互动受到了限制, 当 37 号海豚再次出现时, 社区又恢复了普遍联系, Lusseau 等 [27] 指出 37 号海豚是社区间信息交流的关键个体. 因此, 加权  $K$ -阶传播数法认为节点 37 相对重要的结论是较为合理的. 然而, 37 号海豚处于社区边缘, 因此介数中

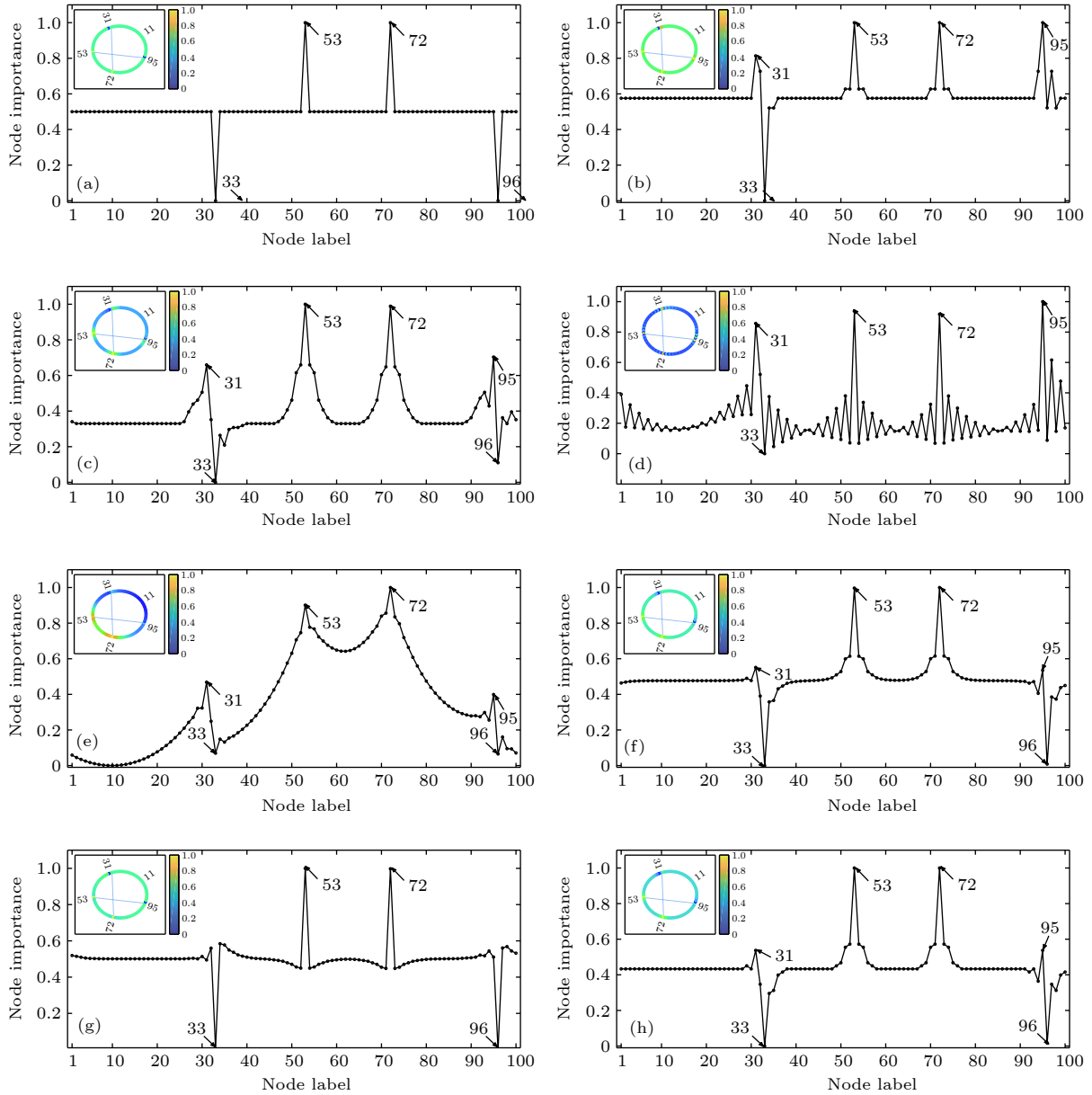


图 3 基于若干传统算法的 WS 小世界网络节点重要性评价结果 (a) 度中心性; (b) Ren 法; (c) Chen 法; (d) 介数中心性; (e) 特征向量法; (f) Katz 指标法; (g) PageRank 算法; (h) CI 法

Fig. 3. Node importance of the WS small-world network based on some traditional algorithms: (a) Degree centrality; (b) Ren method; (c) Chen method; (d) betweenness centrality; (e) eigenvector method; (f) Katz index; (g) PageRank; (h) CI method.

心性认为节点 37 重要性最高的结论是值得商榷的。

此外, 加权  $K$ -阶传播数法提高了对节点 2 和 21 的重视程度, 分别排在第 11 和第 4 位, 而 2 和 21 处于各自海豚社区偏中心位置, 又与 37 等负责信息交流的节点直接相连, 因此该结论是较为合理的; Ren 法、Chen 法、特征向量法、Katz 指标法和 CI 法认为节点 22 相对重要, 但该节点的度中心性与介数均是较低的; 最后, 度中心性法和  $K$ -核分解法存在节点重要性相同的情况, 排序不细致。可见, 加权  $K$ -阶传播数法在评价海豚网络节点重要性时

适当地提高了对海豚社区间信息交流起到关键作用的节点的重视程度, 结果更为合理。

### 3.3 基于节点重要性的蓄意攻击研究

为了进一步验证加权  $K$ -阶传播数法的优越性, 本文对美国西部电网 [25]、芝加哥公路网络 [28]、网络科学家合著网络 [29] 以及小鼠神经纤维束网络 (Kasthuri\_graph\_v4) [30] 进行了节点重要性研究。其中美国西部电网共有 4941 个节点和 6594 条边, 描述了美国西部的高压输电电网结构; 芝加哥公路

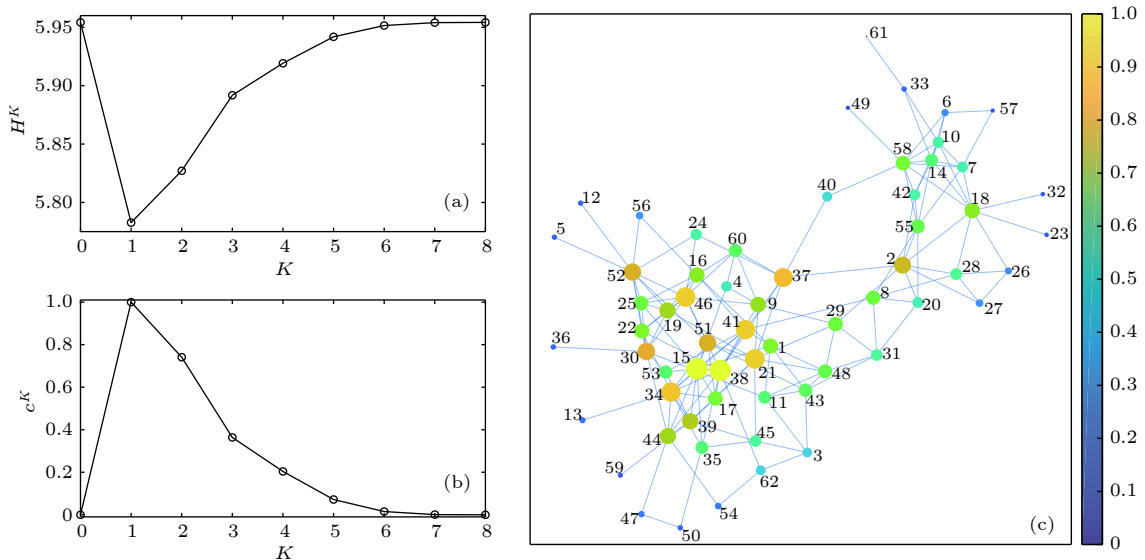


图 4 基于加权  $K$ -阶传播数法的海豚网络节点重要性结果 (a)  $K$ -阶结构熵; (b) 权重系数; (c) 节点重要性  
 Fig. 4. Node importance of the dolphin network based on the weighted  $K$ -order propagation number algorithm: (a) The  $K$ -order structure entropy; (b) the weight coefficient; (c) the importance of nodes.

表 2 海豚网节点重要性排序结果  
 Table 2. Node importance ordering result of the dolphin network.

节点重要性	加权 $K$ -阶传播数法	Ren法	Chen法	介数中心性法	特征向量法	Katz指标法	PageRank算法	CI法
1	15	15	15	37	15	15	15	15
2	38	38	38	2	38	38	18	38
3	46	46	46	41	46	46	52	46
4	21	34	34	38	34	34	58	34
5	41	51	51	8	51	52	38	41
6	34	41	41	18	30	30	46	30
7	37	21	17	21	52	41	34	52
8	30	30	22	55	17	21	30	21
9	52	52	19	52	41	51	14	51
10	51	44	30	58	22	39	2	19
11	2	39	21	40	19	19	21	39
12	39	19	25	29	39	17	39	17
13	19	37	39	30	25	22	10	22
14	44	22	52	44	44	44	41	44
15	9	18	44	15	21	25	55	25

网络共有 1467 个节点和 1298 条边, 描述了美国芝加哥地区的公路运输情况; 网络科学家合著网络共有 1589 个节点和 2742 条边, 描述了从事网络理论和实验的科学家合著关系; 小鼠神经纤维束网络共有 1029 个节点和 1700 条边, 描述了小鼠部分脑皮层中神经元间的轴突束连接. 美国西部电网、芝加哥公路网络和网络科学家合著网络均是无权无向的; 小鼠神经纤维束网络是有权无向的, 为了适应加权  $K$ -阶传播数法, 本文忽略了该网络的边权信

息. 以上网络的  $K$ -阶结构熵  $H^K$  如图 5 所示.

由于以上四个网络的节点数较多, 本文拟采用蓄意攻击策略对节点重要性进行研究<sup>[31-33]</sup>. 蓄意攻击策略是指基于节点重要性由高到低的排序依次攻击网络, 即移除与节点相连的所有边, 通过网络结构随攻击次数的变化情况来评价节点重要性算法的各自特点. 由于在蓄意攻击后网络结构发生了变化, 因此仅仅依据原始网络的节点重要性进行研究会存在偏差. 为了解决此问题, 本文在每次蓄



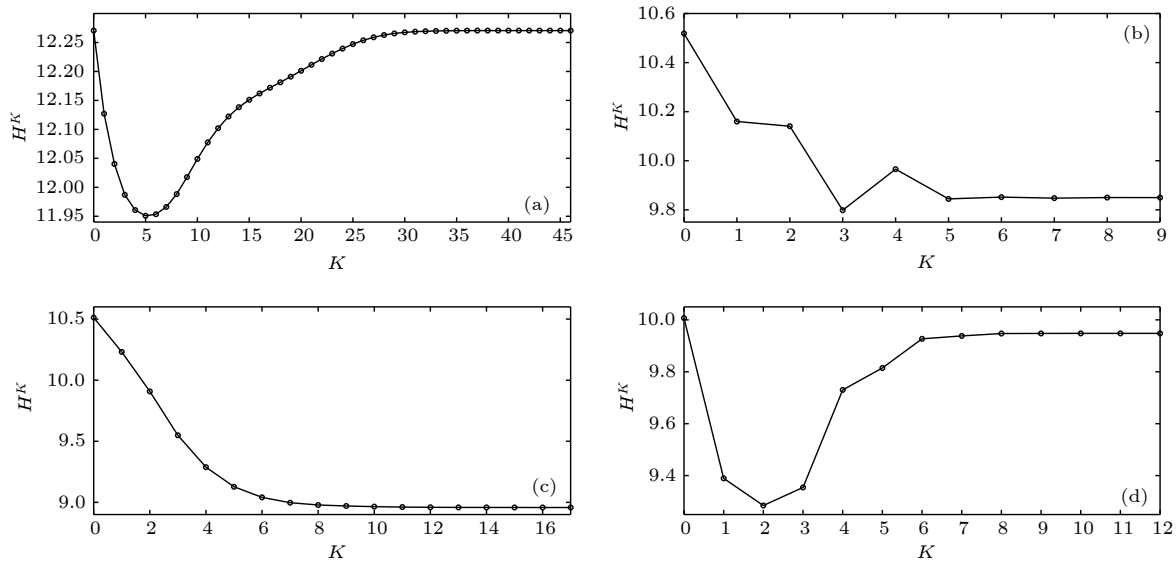


图 5  $K$ -阶结构熵 (a) 美国西部电网; (b) 芝加哥公路网络; (c) 科学家合作网络; (d) 小鼠神经纤维束网络  
 Fig. 5. The  $K$ -order structure entropy: (a) The western power grid of the United States; (b) the road transportation network of the Chicago region; (c) the co-authorship network in network science; (d) the axonal tracts network between neurons of mouse.

意攻击后更新节点排序结果. 此外, 若存在重要性相等的情况, 将选择编号最小的节点进行攻击.

由于网络被攻击后可能会出现孤立节点, 因此选用网络效率来评价网络的连通性. 网络效率的表达式为

$$e = \frac{1}{n(n-1)} \sum_{v_i \neq v_j} \frac{1}{d_{v_i v_j}}, \quad (6)$$

其中  $d_{v_i v_j}$  是指节点  $v_i$  和  $v_j$  之间的最短路径长度. 由 (6) 式可知,  $e$  值越大, 网络效率越高; 当网络由孤立节点组成时  $e = 0$ , 取值最小. 攻击节点可能造成网络连接通路的中断, 节点间的最短路径将有所增大, 网络效率则会随之降低. 为了更为直观地反映被攻击后网络效率的降低情况, 依照文献 [9] 定义网络效率下降率  $\varepsilon$  为

$$\varepsilon = 1 - \frac{e}{e_0}, \quad (7)$$

其中  $e_0$  为未被攻击的原始网络的网络效率. 由 (7) 式可见,  $\varepsilon$  的取值为  $[0, 1]$ , 当网络未被攻击时  $\varepsilon = 0$ , 当网络被攻击时  $\varepsilon$  会随之升高; 最终, 当所有的边被删除时网络效率降至最低, 此时  $\varepsilon = 1$ . 此外, 为了进一步分析攻击前后网络拓扑结构的变化情况, 依照文献 [33] 将网络最大子图节点数设为  $\gamma$ . 图 6 和图 7 分别绘制了美国西部电网、芝加哥公路网络、网络科学家合著网络以及小鼠神经纤维束网络基于不同节点重要性的排序方法, 其网络效率下降率  $\varepsilon$  和最大子图节点数  $\gamma$  随攻击次数的变化曲

线. 其中, Katz 指标的权重衰减因子为 1.5 倍邻接矩阵最大特征值的倒数; PageRank 算法的阻尼系数取 0.85; CI 法的权重系数取 0.5.

对美国西部电网而言, 依照加权  $K$ -阶传播数法和介数中心性法的排序结果进行攻击, 网络效率下降得最为迅速, 仅攻击约 100 次后, 网络效率便下降了近 90%; 而依据度中心性、Ren 法、Chen 法、特征向量法、Katz 指标、PageRank 算法和 CI 法则需约 300 次,  $K$ -核分解法需约 550 次, 才能达到同等效果. 此外, 基于加权  $K$ -阶传播数法和介数中心性对网络进行攻击, 最大子图节点数降低的速率远远高于其他方法. 以加权  $K$ -阶传播数法为例, 当网络被攻击 200 次后, 最大子图节点数仅为 154, 为原始网络节点数的 3%, 可认为网络基本瘫痪. 而度中心性、Ren 法、Chen 法、特征向量法、Katz 指标和 CI 法则需攻击近 400 次, PageRank 算法需 650 次,  $K$ -核分解法需 1000 次才能达到相同效果; 对芝加哥公路网络而言, 依据加权  $K$ -阶传播数法、介数中心性、Chen 法、特征向量法、Katz 指标和 CI 法进行蓄意攻击, 网络被破坏的程度较为接近, 且优于度中心性、Ren 法、PageRank 算法和  $K$ -核分解法. 其中, 依据加权  $K$ -阶传播数法进行蓄意攻击, 网络效率下降得最为迅速; 对网络科学家合著网络而言, 依照加权  $K$ -阶传播数法和介数中心性进行蓄意攻击, 网络被破坏的程度较为接近, 且优于其他算法; 对小鼠神经纤维束网络而言, 依

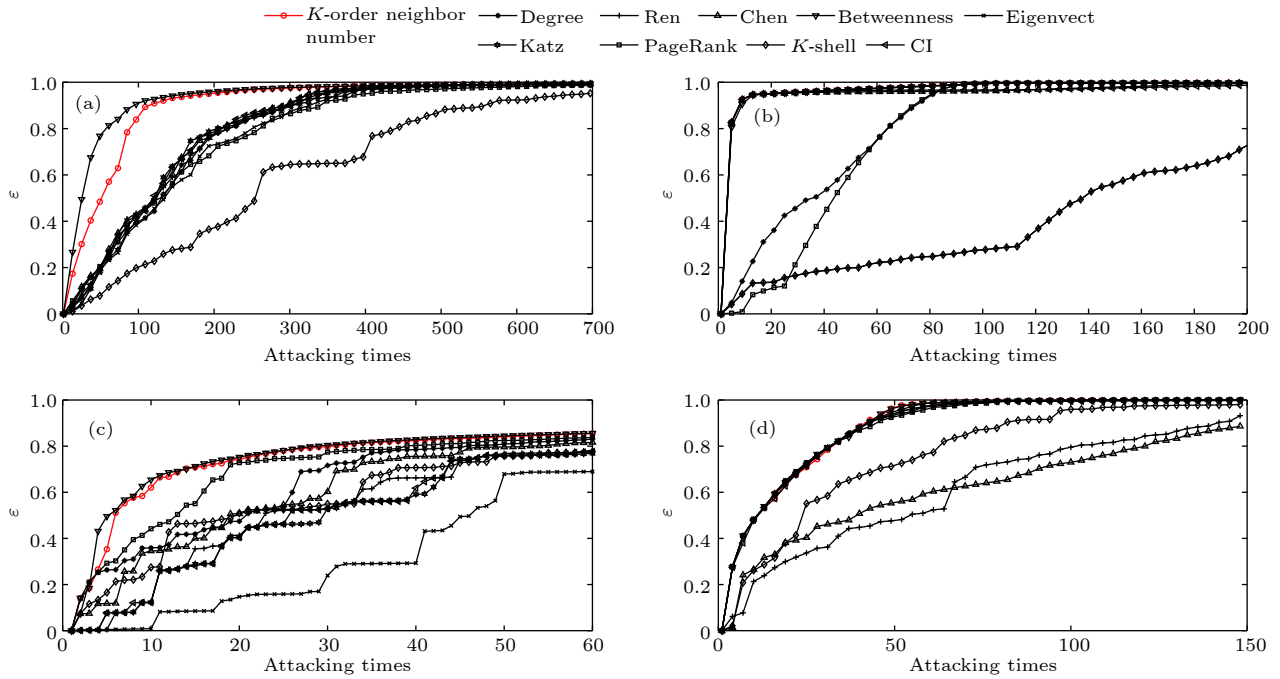


图 6 网络效率下降率  $\varepsilon$  随攻击次数的变化情况 (a) 美国西部电网; (b) 芝加哥公路网络; (c) 网络科学家合著网络; (d) 小鼠神经纤维束网络

Fig. 6. Change of the network efficiency decrease rate  $\varepsilon$  with attacking times: (a) The western power grid of the United States; (b) the road transportation network of the Chicago region; (c) the co-authorship network in network science; (d) the axonal tracts network between neurons of mouse.

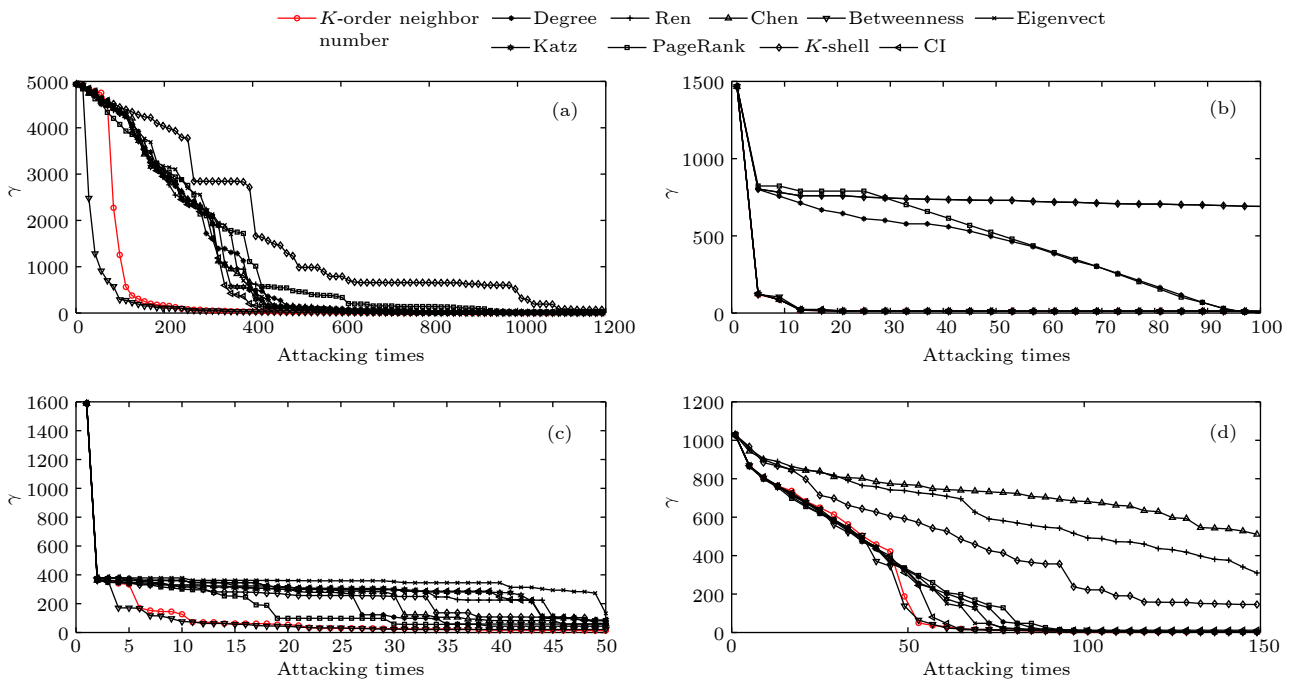


图 7 最大子图节点数  $\gamma$  随攻击次数的变化情况 (a) 美国西部电网; (b) 芝加哥公路网络; (c) 网络科学家合著网络; (d) 小鼠神经纤维束网络

Fig. 7. Change of the node number of the maximum sub-graph  $\varepsilon$  with attacking times: (a) The western power grid of the United States; (b) the road transportation network of the Chicago region; (c) the co-authorship network in network science; (d) the axonal tracts network between neurons of mouse.

据加权  $K$ -阶传播数法、度中心性、介数中心性、特征向量法、Katz 指标、PageRank 算法和 CI 法进行蓄意攻击, 网络被破坏的程度较为接近, 且优于 Ren 法、Chen 法和  $K$ -核分解法. 其中, 依据加权  $K$ -阶传播数法和介数中心性进行蓄意攻击, 网络最大子图节点数下降得最为迅速, 略优于其他算法.

综上所述, 无论对美国西部电网、芝加哥公路网络、网络科学家合著网络还是小鼠神经纤维束网络而言, 基于加权  $K$ -阶传播数法进行蓄意攻击, 仅需移除少量重要节点便可实现对网络结构的充分破坏.

## 4 结 论

本文考虑到个体间的相关关系, 基于网络拓扑结构重新描述了传染病模型, 并将各个节点分别设置为传染源进行疾病传播, 提出了加权  $K$ -阶传播数法. 通过对 WS 小世界网络的仿真分析发现, 加权  $K$ -阶传播数法相较于传统的节点重要性评价指标能够更为综合地考量长程连接对网络长、短程通信的影响; 此外, 加权  $K$ -阶传播数法不仅认为海豚网络社区中心的重要性较高, 且适当地提高了对社区间的信息交流起到关键作用的海豚的受重视程度. 最后, 本文还基于蓄意攻击策略以美国西部电网、芝加哥公路网络、网络科学家合著网络以及小鼠神经纤维束网络为实例进行了研究. 然而, 由于目前加权  $K$ -阶传播数法的求解依赖于网络邻接矩阵的  $K$ -阶多项式, 需要进行  $K-1$  次矩阵乘法和  $K$  次矩阵加法, 时间复杂度较高. 因此需要进一步探索  $K$ -阶传播数的物理意义以提出更为便捷的求解方案. 此外, 小鼠神经纤维束网络原为有权网络, 为了适应加权  $K$ -阶传播数法, 本文忽略了该网络的边权信息. 因此将加权  $K$ -阶传播数法进行改进, 使其适用于有向有权网络, 也是后续研究的重点.

## 参考文献

- [1] Zhou X, Zhang F M, Zhou W P, Zhou W, Yang F 2012 *Acta Phys. Sin.* **61** 190201 (in Chinese) [周漩, 张凤鸣, 周卫平, 邹伟, 杨帆 2012 物理学报 **61** 190201]
- [2] Zhou T, Bai W J, Wang B H, Liu Z J, Yan G 2005 *Physics*

- 34** 31 (in Chinese) [周涛, 柏文洁, 汪秉宏, 刘之景, 严钢 2005 物理 **34** 31]
- [3] Liu J G, Wang Z T, Dang Y Z 2005 *Mod. Phys. Lett. B* **19** 785
- [4] Wang X F, Li X, Chen G R 2006 *Complex Network Theory and Application* (Beijing: Tsinghua University Press) p8 (in Chinese) [汪小帆, 李翔, 陈关荣 2006 复杂网络理论及其应用 (北京: 清华大学出版社) 第8页]
- [5] Dunne J A, Williams R J, Martinez N D 2002 *Ecol. Lett.* **5** 558
- [6] Samant K, Bhattacharyya S 2004 *Proceedings of the 37th Annual Hawaii International Conference on System Sciences* Washington, USA, January 5-8, 2004 p289
- [7] Newman M E J, Forrest S, Balthrop A 2002 *Phys. Rev. E* **66** 035101
- [8] Jeong H, Mason S, Barabasi A L 2001 *Nature* **411** 41
- [9] Ren Z M, Shao F, Liu J G, Guo Q, Wang B H 2013 *Acta Phys. Sin.* **62** 128901 (in Chinese) [任卓明, 邵凤, 刘建国, 郭强, 汪秉宏 2013 物理学报 **62** 128901]
- [10] Chen D B, Lü L Y, Shang M S, Zhang Y C, Zhou T 2012 *Physica A* **391** 1777
- [11] Freeman L C 1977 *Sociometry* **40** 35
- [12] Sabidussi G 1966 *Psychometrika* **31** 581
- [13] Bavelas A 1950 *J. Acoustical Soc. Amer.* **22** 725
- [14] Stephenson K, Zelen M 1989 *Soc. Netw.* **1** 11
- [15] Borgatti S P 2005 *Soc. Netw.* **27** 55
- [16] Katz L 1953 *Psychometrika* **18** 39
- [17] Zhang J, Xu X K, Li P, Zhang K, Small M 2011 *Chaos* **21** 016107
- [18] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
- [19] Zeng A, Zhang C J 2013 *Phys. Lett. A* **377** 1031
- [20] Bryan K, Leise T 2006 *SIAM Rev.* **48** 569
- [21] Lü L Y, Zhang Y C, Yeung C H, Zhou T 2011 *PLoS One* **6** e21202
- [22] Zhong L F, Liu Q H, Wang W, Cai S M 2018 *Physica A* **511** 78
- [23] Allen L J S 1994 *Math. Biosci.* **124** 83
- [24] Huang L Y, Huo Y L, Wang Q, Cheng X F 2019 *Acta Phys. Sin.* **68** 018901 (in Chinese) [黄丽亚, 霍有良, 王青, 成谢锋 2019 物理学报 **68** 018901]
- [25] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [26] Lusseau D, Schneider K, Boisseau O J, Haase P, Slooten E, Dawson S M 2003 *Behav. Ecol. Sociobiol.* **54** 396
- [27] Lusseau D, Newman M E J 2004 *Proc. Biol. Sci.* **271** 477
- [28] Eash R W, Chon K S, Lee Y J, Boyce D E 1983 *Transport. Res. Rec.* **994** 30
- [29] Newman M E J 2006 *Phys. Rev. E* **74** 036104
- [30] Ryan A R, Nesreen K A 2015 *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* Austin, Texas, January 25-30, 2015 p4292
- [31] Li P X, Ren Y Q, Xi Y M 2004 *Syst. Eng.* **22** 13 (in Chinese) [李鹏翔, 任玉晴, 席西民 2004 系统工程 **22** 13]
- [32] He N, Li D Y, Gan W Y, Zhu X 2007 *Comput. Sci.* **34** 1 (in Chinese) [赫南, 李德毅, 淦文燕, 朱熙 2007 计算机科学 **34** 1]
- [33] Zhao Z Y, Meng X R, Sun R N 2018 *Comput. Eng.* **44** 62 (in Chinese) [赵志远, 孟相如, 孙瑞男 2018 计算机工程 **44** 62]

# Node importance based on the weighted $K$ -order propagation number algorithm\*

Huang Li-Ya<sup>1)2)†</sup> Tang Ping-Chuan<sup>1)</sup> Huo You-Liang<sup>1)</sup>

Zheng Yi<sup>1)</sup> Cheng Xie-Feng<sup>1)2)</sup>

1) (*College of Electronic and Optical Engineering, College of Microelectronics, Nanjing University of*

*Posts and Telecommunications, Nanjing 210023, China*)

2) (*National and Local Joint Engineering Laboratory of RF Integration and Micro-Assembly Technology, Nanjing 210003, China*)

( Received 16 January 2019; revised manuscript received 5 April 2019 )

## Abstract

The measurement of node importance is significant for analyzing a network structure. To comprehensively reflect the global and local network features, in this paper we abstract the propagating process of epidemic diseases based on the network topology structure, and then respectively sets each node as an infection source. After a certain dissemination time  $K$ , the number of infected nodes in the network is defined as the  $K$ -order propagation number, and the weighted sum of  $K$ -order propagation numbers under different values of  $K$  is taken as the important index of nodes. The simulation experiments of Watts-Strogatz small-world networks and a dolphin network show that the weighted  $K$ -order propagation number algorithm is more effective than the traditional method in evaluating the importance of nodes. For the Watts-Strogatz small-world networks, it can reflect the influence of long-distance connections on information transmission in detail. For the dolphin network, the weighted  $K$ -order propagation number algorithm significantly raises the profile of those nodes which play a key role in the information communication between different dolphin communities. In addition, in this paper we use a deliberate attacking method to analyze the western power grid of the United States, the road transportation network of the Chicago region, the co-authorship network in network science and the axonal tracts' network between neurons of mouse. According to the specific order of the node importance from high to low, network nodes are attacked in turn, that is, all edges of the attacked nodes are removed. The analysis results of network parameters such as the network efficiency and the node number of the maximum sub-graph changing with the attacking times show that comparing with other evaluation indices of node importance such as degree, Ren method, Chen method, eigenvector method, Katz index, PageRank, CI method and  $K$ -shell, the weighted  $K$ -order propagation number algorithm focuses much on destroying the major structure, and all of the above four networks will break down if only a small number of important nodes are attacked. For example, after attacking only 100 nodes, the network efficiency of the western power grid of the United States is down by 90%, and after attacking 200 nodes, the network scale of the maximum sub-graph is nearly 3% of the original network.

**Keywords:** complex network, node importance, propagation model

**PACS:** 89.70.Cf, 89.75.Fb

**DOI:** 10.7498/aps.68.20190087

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 71671093, 61271334).

† Corresponding author. E-mail: [huangly@njupt.edu.cn](mailto:huangly@njupt.edu.cn)