

# 基于二项熵和邻域节点间范德华力的关键节点识别方法

梁威, 孙鹏, 张杰勇, 肖越文

(空军工程大学信息与导航学院, 西安, 710077)

**摘要** 对节点重要性进行排序是复杂网络中识别关键节点的一种常用分析方法, 分析网络中节点的重要性, 有助于深入了解网络特性。在现有方法上为进一步提升节点评估精准度, 引入二项熵概念来量化节点在网络中的重要性, 通过邻域相似度衡量节点间的相互影响力, 同时采用范德华力抽象节点之间的相互作用关系, 提出一种基于二项熵和邻域节点间范德华力的关键节点识别方法, 该方法从网络的整体信息流和相邻节点之间的位置和交互关系, 综合考虑节点的局部和全局特征, 并选取3个同类算法通过3个评价指标验证性能优劣, 实验结果表明该算法对重要节点的判断具有良好的性能。

**关键词** 复杂网络; 二项熵; 邻域拓扑; 范德华力; 节点识别

**DOI** 10.3969/j.issn.2097-1915.2024.04.011

**中图分类号** TP301.6 **文献标志码** A **文章编号** 2097-1915(2024)04-0072-07

## A Key Node Identification Method Based on Binomial Entropy and Van der Waals Forces between Neighboring Nodes

LIANG Wei, SUN Peng, ZHANG Jieyong, XIAO Yuewen

(Information and Navigation School, Air Force Engineering University, Xi'an 710077, China)

**Abstract** Ranking importance in nodes is a commonly used analysis method for identifying key nodes in complex networks, and importance of analyzing nodes in the network is a great help to a deeper understanding network characteristics. In order to further improve the accuracy of node evaluation based on existing methods, this paper introduces the concept of Binomial Entropy to quantify the importance of nodes in the network, measures the mutual influence between nodes through neighborhood similarity, and simultaneously uses Van der Waals force to abstract the interaction between nodes. Therefore, a key node identification method based on binomial entropy and Van der Waals force between neighboring nodes is proposed. This method is a comprehensive consideration of the local and global characteristics of nodes from the overall information flow of the network and the location and interaction relationship between adjacent nodes, and a selection from three similar algorithms to verify the performance through three evaluation indicators. The experimental results show that the algorithm in this paper is good in performance in judging important nodes.

**Key words** complex networks; binomial entropy; neighborhood topology; Van der Waals forces; node identification

收稿日期: 2023-12-07

基金项目: 陕西省自然科学基金基础研究计划(2023-JC-QN-0728)

作者简介: 梁威(1994-), 男, 陕西西安人, 硕士生, 研究方向为复杂网络研究。E-mail: 1012644574@qq.com

**引用格式:** 梁威, 孙鹏, 张杰勇, 等. 基于二项熵和邻域节点间范德华力的关键节点识别方法[J]. 空军工程大学学报, 2024, 25(4): 72-78. LIANG Wei, SUN Peng, ZHANG Jieyong, et al. A Key Node Identification Method Based on Binomial Entropy and Van der Waals Forces between Neighboring Nodes[J]. Journal of Air Force Engineering University, 2024, 25(4): 72-78.

随着现代信息技术的快速发展,网络已经成为现代生活中不可或缺的组成部分,许多系统都可以通过网络的视角来描述,为更好地理解和分析网络,复杂网络概念被提出。复杂网络将现实网络中的实体抽象成节点,将实体之间的关系抽象成连边,通过节点和连边组成的网络结构来反映现实世界的复杂系统。在军事领域中<sup>[1]</sup>,可以通过识别关键节点,确定作战中的关键环节和薄弱点为制定更加科学合理的作战策略提供支持。因此复杂网络的关键节点研究引起了国内外广泛关注,关键节点功能受到破坏或失效,会导致网络的崩溃或功能的丧失,所以研究复杂网络重要节点对于网络的稳定性和效能具有至关重要的作用<sup>[2-5]</sup>。

现有的基础节点重要性排序方法有:度中心性、介数中心性<sup>[6]</sup>、接近中心性、K-shell等。由于基础算法排序结果过于粗粒度,在大规模网络中存在当前节点的邻节点中有2个或多个节点具有相同的赋值,比如度值、 $k_i$ 值等,导致重要性无法进一步区分。究其原因,一是缺少一种方法来细化区分节点颗粒度<sup>[7]</sup>,二是没有考虑到节点间距离远近对重要性评估的影响。因此,为了弥补上述缺点,郭程远等<sup>[8]</sup>提出基于迭代K-shell和改进信息熵的节点重要性排序算法,主要采用香农提出的信息熵来进一步提升算法的精细度,在复杂网络中节点信息熵越高,则该节点接收的信息越多,对粗粒度网络排序进行优化。邹龙等<sup>[9]</sup>提出一种度联合信息熵的网络重要节点识别方法,该算法通过当前节点及邻域节点的度值来反映本节点的重要性。将节点信息量比重称为节点之间的信任度,同时考虑节点邻域重合度来进一步评估相同赋值的节点对整个网络的重要性贡献是否一样;林欣等<sup>[10]</sup>提出一种节点度与邻域相似度标签传播算法,通过计算邻域网络相似度<sup>[11]</sup>与节点度值来度量节点重要性,从而对节点间顺序进行优化。

上述研究在基础方法上结合信息熵和邻域相似度的方法解决了原始排序过于粗粒度的问题,引入信息熵概念来量化周围节点对当前节点重要性的贡献,通过邻域相似度突出拓扑结构对重要性的影响,排序结果在基础方法上有了较大提升。因此本文受到文献<sup>[8~10]</sup>启发,在衡量节点在网络中重要性上采用复杂度较低的二项熵,在描述节点之间关系上改进领域相似度方法并引入范德华力<sup>[12]</sup>的度量方式,综合考虑节点在网络中所处的位置,利用节点的影响力以及拓扑信息兼顾节点间三角结构<sup>[13]</sup>体现节点的重要性,进一步提升节点评估精准度。综上所述,本文提出一种基于二项熵和邻域节点间范德

华力的关键节点识别方法(binomial entropy and Van der Waals forces between linkage similarity, EVL),充分挖掘节点在网络中的重要程度,并通过实验验证了其有一定的优势。

## 1 EVL方法研究

复杂网络中的关键节点识别问题是指确定网络中最具影响力的节点,在复杂网络的研究中,这些节点对于网络的稳定性和整体结构具有重要影响<sup>[14-15]</sup>。

无向无权复杂网络模型可以表示为 $G = (V, E)$ ,其中 $V = (v_1, v_2, \dots, v_n)$ 为网络节点集合, $E = (e_1, e_2, \dots, e_m)$ 为节点间的连边集合, $n = |V|$ 为网络的节点总数, $m = |E|$ 为网络的连边总数。

### 1.1 EVL方法构建

本文为了进一步优化复杂网络节点重要度排序,将3种算法合并考虑,在提升区分节点颗粒度上引入二项熵的概念,二项熵相比于香农信息熵复杂度较低,节省计算时间;在考虑节点之间关系和位置上引入改进邻域相似度和范德华力概念,综合考虑网络全局和局部的节点特性,进一步提高整体排序的准确性和可靠性。

#### 1.1.1 二项熵

熵构建的过程首先需要考虑构建熵所需的概率集,一般情况下使用网络结构中某一特性参数,为了降低计算复杂度使用节点度 $k_i$ 作为参数,可以将概率 $p_i$ 定义为:

$$p_i = \frac{k_i}{\sum_{j \in \Gamma_i} k_j} \quad (1)$$

式中: $k_i$ 和 $k_j$ 分别为节点 $i$ 和 $j$ 的度数; $\Gamma_i$ 为节点 $i$ 的一阶邻居。

信息熵被用来衡量事件中的信息量<sup>[16-17]</sup>,本文将每次节点熵的计算视为待测节点和其余所有节点,实现二项分布特征,引入二项熵<sup>[18-19]</sup>这一概念来衡量节点的重要性。则定义节点 $i$ 的二项熵 $E_i$ 为:

$$E_i = - [p_i \log p_i + (1 - p_i) \log(1 - p_i)] \quad (2)$$

式中:熵值 $E_i$ 越大,表示该节点的信息量越多、越重要。

#### 1.1.2 邻域相似度

通过计算待测节点的一阶邻居节点在网络拓扑结构上的重合度<sup>[20]</sup>,来确定节点是否容易被替代,重合度越低,说明节点越不容易被取代。节点相似度指标定义为Jaccard指标值,表示为

$$\text{sim}(i, j) = \begin{cases} \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|}, & i, j \text{ 之间无连接} \\ 1, & i, j \text{ 之间有连接} \end{cases} \quad (3)$$

式中:  $N(i)$  为节点  $i$  的一阶邻居节点集合;  $|N(i)|$  为节点  $i$  的一阶邻居节点个数,  $|N(i) \cap N(j)|$  和  $|N(i) \cup N(j)|$  分别为节点  $i$  与  $j$  的共同邻居节点数和全部邻居节点数。  $\text{sim}(i, j)$  的值介于 0 与 1 之间。在文献[19]所提 LLS 算法的基础上, 增加考虑节点与节点之间度和连边的关系, 优化得出邻域相似度评估公式为:

$$\text{LS} = \sum_{i, j \in N(i)} \left( \frac{k_i(k_i - 1)}{2} - \text{sim}(i, j) \right) \quad (4)$$

式中:  $\frac{k_i(k_i - 1)}{2}$  为节点  $i$  与其他节点之间可能存在的边的最大数量。LS 值越大说明  $\frac{k_i(k_i - 1)}{2}$  和  $\text{sim}(i, j)$  差异越大, 即相邻节点在功能上或拓扑上相似度越低, 节点越不容易被替代, 反之, 则表示 2 个集合在网络中有更多的共同边存在。

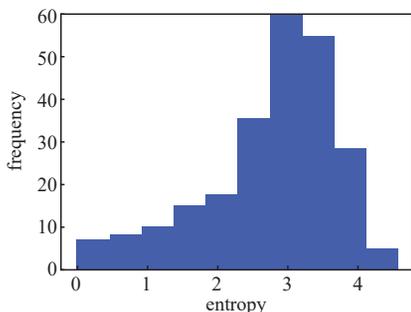
### 1.1.3 范德华力

范德华力<sup>[21]</sup>指的是分子间的作用力, 将范德华力引入到复杂网络中, 将网络中的节点映射为分子或原子, 并将节点的度值表示为分子或原子的范德华常量, 由于节点间范德华力兼具引力和斥力的特点, 提高了节点的评估精度。用范德华力来定义节点的影响力, 表示为:

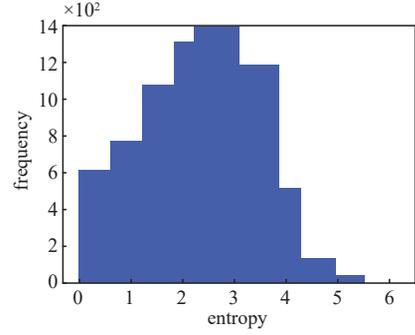
$$V(i) = \sum_{j \neq i} \left( \frac{k_i}{d_{ij}^{12}} - \frac{k_j}{d_{ij}^6} \right) \quad (5)$$

式中:  $i$  为网络中的节点; 度值为  $k_i$ ;  $j$  为非  $i$  节点, 度值为  $k_j$ ;  $d_{ij}$  为节点  $i$  与节点  $j$  之间的最短路径长度。2 个节点之间的作用力与它们之间的距离成反比, 由于范德华排斥力与  $d_{ij}^{12}$  成反比, 范德华吸引力与  $d_{ij}^6$  成反比, 所以当节点的范德华力为正时, 表现为斥力, 反之为引力。

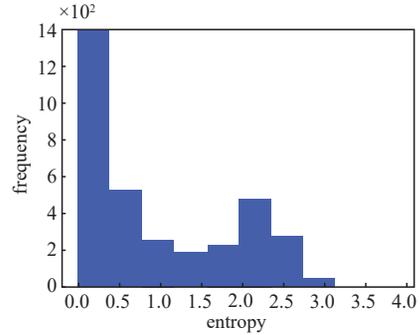
由于网络越大, 会导致需要处理的节点数越多, 为了降低算法的时间复杂度, 通过二项熵对节点进行预处理, 实验网络节点熵的分布如图 1 所示。



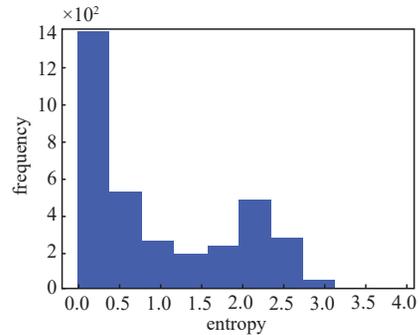
(a) Bn-macaque-rhesus brain\_1 node entropy



(b) FB-pages-government node entropy



(c) P2P-Gnutella25 node entropy



(d) P2P-Gnutella25 node entropy

图 1 节点熵的分布

通过图 1, 可以将节点熵的阈值区间设置为  $(3, \infty)$ , 熵小于 3 的节点不进入下一轮的计算。因此, 本文通过借助线性整流函数 ReLU (rectified linear unit) 函数对上述节点熵进行过滤, 对网络中的节点重要性等级进行进一步修正, 其计算式为:

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x > 3 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

同时, 在后续的计算过程中需要考虑到 3 种算法数值融合, 为统一不同算法的量纲, 将  $E_i$ 、LS、 $V(i)$  得出的值分别通过线性归一化公式, 将数据映射到  $[0, 1]$  区间, 归一化公式表示为:

$$x_i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

式中:  $x_{\max}$  和  $x_{\min}$  分别为原始数据中的最小值和最大值。通过式 (7), 将原始数据  $x$  缩放到  $[0, 1]$  之间, 使得归一化后  $0 < x_i < 1$ 。

至此, 本文通过综合考虑从 3 个角度共同计算

节点的重要性,将归一化得到的结果求和,最终得到节点重要性排序数值,表达式为:

$$\text{EVL}(i) = x_{E_i} + x_{v(i)} + x_{LS} \quad (8)$$

## 1.2 算法流程

EVL算法的具体步骤如下:

**步骤 1** 根据式(2)对输入网络进行处理,初步得到网络中节点的二项熵。

**步骤 2** 二项熵小于3的节点,不进入后续计算,二项熵大于3的节点用式(4)~式(5)计算相应数值。

**步骤 3** 为了消除3种不同方法所在不同量纲和取值范围的影响,以便下一步进行指标间融合,使用式(8)对3种结果进行线性归一化。将3个计算结果映射到同一个区间,得到3个处理后的节点重要度指标。

**步骤 4** 将3个指标数值求和,得到最后节点重要度评价数值,根据该公式结果对网络节点进行重要度排序。

图2为EVL算法流程。

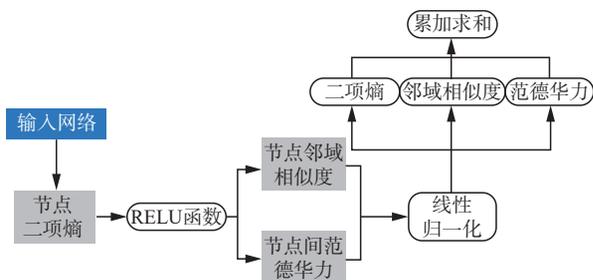


图2 EVL算法流程

## 2 评价指标

本文评价节点的重要性主要遴选网络剩余边、网络效率和网络敏感率这3个指标。按照节点重要性顺序依次删除节点,通过分析3个指标的变化,得出节点删除对网络结构和功能的影响程度,从而反映出节点对网络的重要程度。

### 2.1 网络剩余边

网络剩余边代表当某节点被破坏后的网络剩余连通能力,当节点删除后,对网络连通造成的破坏越大,节点就越重要,计算方法为:

$$P_{\text{Edges}} = \frac{E}{M} \quad (9)$$

式中: $E$ 为每次去除节点后剩余网络的总边数; $M$ 为初始网络的连边数量。

### 2.2 网络效率

不同功能的网络具有不同特性。网络的连通规模不一定能很好地反映该网络的效能。因此,基于全局效率EF提出了可以对包含多个子网络的网络群评估整体效率的方法,设网络整体效率TE为:

$$\text{TE} = \sum_{i=1}^T n_i^2 \text{EF}(i) = \sum_{i=1}^T \frac{n_i}{(n_i - 1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (10)$$

$$\text{EF} = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (11)$$

式中: $T_i$ 为连通子图 $i$ ;  $n_i$ 为第 $i$ 个连通子图的规模(节点数量)。若节点处于孤立地位,则网络服务效率 $\text{EF}=0$ 。若删除节点后,网络效率曲线降幅越快,说明被攻击节点的重要性越高。若节点处于孤立地位,则网络服务效率 $\text{EF}=0$ 。

### 2.3 网络敏感度

删除个别节点不会完全破坏网络,但会影响网络的稳定性。可以通过网络敏感率来侧面反映网络的稳定性。被攻击的节点越重要,网络将产生较大的波动,就越先出现峰值,计算式为:

$$S = \sum_{s < \sigma} \frac{n_s s^2}{n} \quad (12)$$

式中: $n_s$ 为大小为 $s$ 的组件的个数; $n$ 为网络节点数量; $\sigma$ 用于判断网络被摧毁后的剩余碎片阈值。随着网络节点的逐渐移除,网络被分解为许多小尺寸的网络碎片。当网络解体至一定程度后,会在图像上出现敏感度 $S$ 的峰值,出现峰值意味着已经将原网络最大限度地拆解成小于等于阈值 $\sigma$ 的碎片群。

## 3 实验结果与分析

本文仿真采用PyCharm Community Edition软件搭建的Python 3.10环境,使用4个真实数据集,分别为Bn-macaque-rhesus-brain、FB-pages-government、P2P-Gnutella25和Road-Italy来验证EVL算法对节点重要度评估的效果。实验参数见表1。

表1 实验的网络参数

数据集	节点数	边数	平均度	平均聚类系数	三元组占比/%
Bn-macaque-rhesus-brain	242	4 100	12	0.532 6	36.15
FB-pages-government	7 100	89 400	25	0.410 9	22.38
P2P-Gnutella25	22 700	54 700	4	0.053 1	0.454
Road-Italy	6 700 000	700 000	2	0.010 2	0.269

图 3 中 4 组图代表本文选取的 4 个数据集,为了使实验结果具有一定普适性,图 3(a)代表稀疏网络,图 3(b)代表高密度社区网络,图 3(c)代表点对点的分布式去中心化网络,图 3(d)组代表大规模网络。曲线分别为用 CEKS、DE、LLS-LPA 和 EVL

算法对数据集网络的节点进行重要性排序,按照各方法的顺序,从最重要的节点依次删除。再通过网络服务效率、网络剩余连通能力和网络敏感性 3 个指标给出网络变化。从实验结果不难看出,本文算法在复杂网络重要节点排序上具有优势。

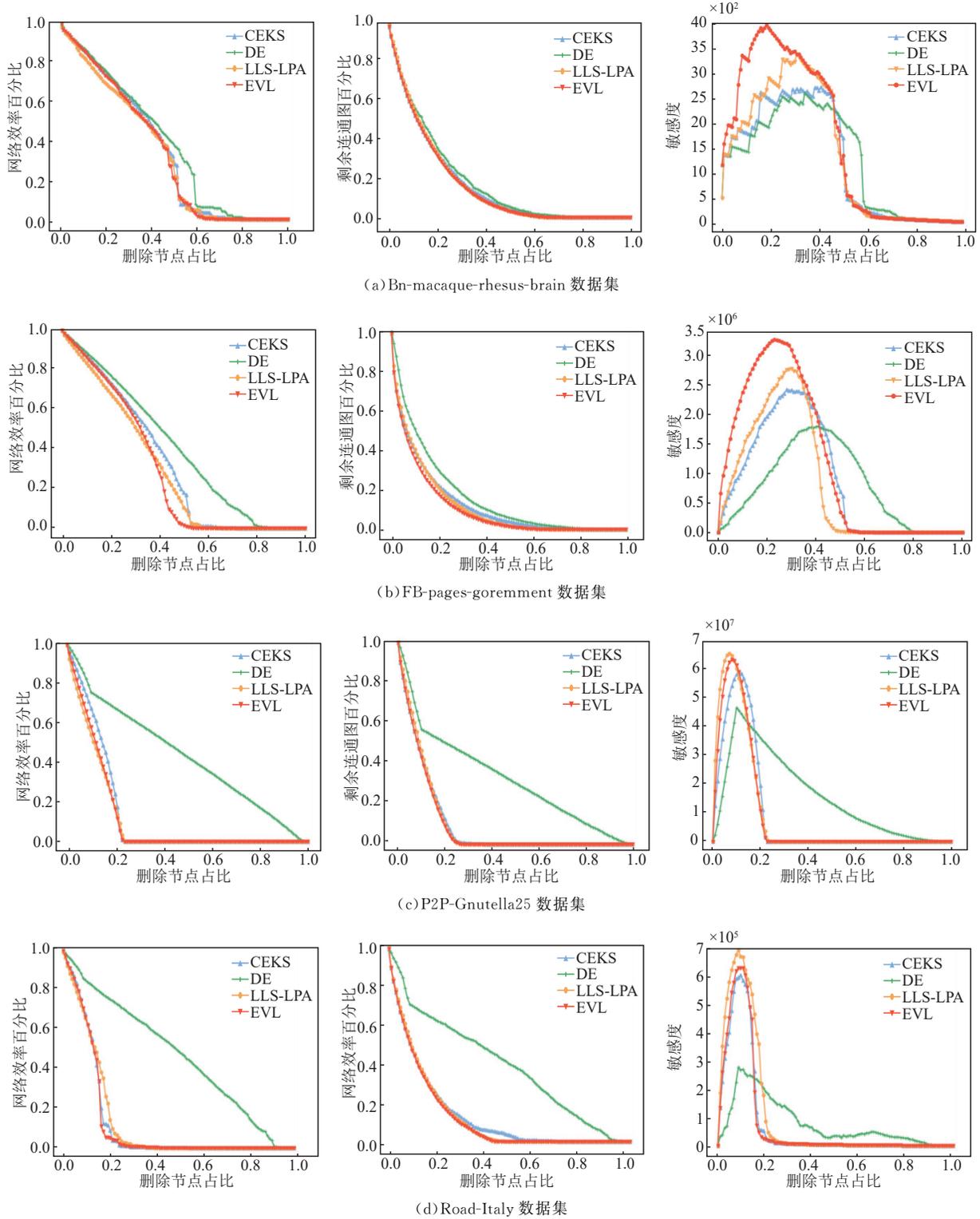


图 3 4 组数据集结果对比

图 3 中每组左中右 3 幅图分别为网络服务效率、网络剩余连通能力和网络敏感性 3 个指标。从图 3 结果可知,在网络效率指标方面,可以观察到

EVL 算法在刚开始删除节点时,能力稍逊于 LLS-LPA 算法,但是短时间内效果转好且优于其他算法,并且率先趋于平稳;EVL 算法的网络剩余连通

能力下降速度比其他方法快;在网络敏感性方面,按照EVL算法排序后删除节点能使网络快速产生较大的波动,并且很快达到峰值。总的来说,本文所提EVL算法在网络数量级为 $10^3$ 的效果表现优异,在大规模网络中,优势不是很明显,还有提升空间。总体来看,本文算法对节点排序相较于其余3个方法更能反映网络节点重要性。下一步计划研究重点放在动态网络场景下的多层复杂网络,结合深度学习<sup>[24]</sup>的方法,更好地处理大规模的复杂网络。

从4个网络的参数和图3中的曲线分析网络参数对网络整体鲁棒性的影响。当三元组占比较高时,表示网络中相似节点之间有较强的连接集群化趋向,从而增加了网络传导信息的能力,也增强了网络对攻击的抗性,而较高的平均聚类系数往往意味着较高的图密度,因为聚类节点之间的连接增加了网络的总连接数,意味着网络鲁棒性较好。根据网络参数对4组网络三元组和平均聚类系数占比的高低进行排序,可以得到(a)>(b)>(c)>(d)。结合4组曲线,可以看出4组网络在节点遭受删除,网络效率下降达到平稳态时,4组图(左)节点删除占比排序为(a)>(b)>(c)>(d);在网络剩余连通能力上,可以清楚地看出4组图(中)曲线斜率由低到高为(a)<(b)<(c)<(d);在网络敏感度达到峰值时,4组图(右)节点删除占比排序为(a)>(b)>(c)>(d)。综上可得网络参数中三元组和平均聚类系数占比高低对网络的鲁棒性<sup>[25]</sup>起到了较为关键的作用。

## 4 结语

本文通过二项熵、改进邻域相似度和范德华力3种方法的结合,提出一种EVL算法来识别网络中重要节点。首先通过计算网络节点二项熵来进行预处理,不仅能精细区分节点颗粒度,还能降低时间复杂度。再考虑邻居节点的信息以及节点间拓扑结构,综合利用节点的影响力以及拓扑信息来体现节点的重要性。最后通过实验,结合网络剩余边、网络效率和网络敏感度3个指标对算法效能进行评价。实验结果表明本文算法虽然在处理网络的时间复杂度上相比于现有的改进算法没有大幅度减少,但是在分析和识别网络重要节点的精确性上有较为明显的提高,对后续的研究有一定借鉴意义。

采用网络攻击的方式分析节点的重要性,主要用于无权无向网络,对于现实网络中的复杂模型还过于简单。因此在下一步研究中,一方面继续提升算法在大规模网络的能力,另一方面将侧重于研究

多层以及超网络。多层网络突破了单层网络中节点和连边同质性的限制,考虑了多种类型节点及其连边关系,多层网络的拓扑结构以及相互依存的鲁棒性和抗毁性如何,本文方法是否能够继续适用,是否存在更合理的评价指标,可否引入深度学习等方法强化算法的韧性,都将是下一阶段的研究方向。

## 参考文献

- [1] 杨倩. 基于复杂网络的作战体系节点重要性评估方法研究[D]. 绵阳:西南科技大学,2023.
- [2] 石福丽. 基于超网络的军事通信网络建模、分析与重构方法研究[D]. 长沙:国防科学技术大学,2013.
- [3] LI J C, JIANG J, YANG K W, et al. Research on Functional Robustness of Heterogeneous Combat Networks[J]. IEEE Systems Journal, 2019, 13(2): 1487-1495.
- [4] SHANG Q Y, DENG Y, CHEONG K H. Identifying Influential Nodes in Complex Networks: Effective Distance Gravity Model[J]. Information Sciences, 2021, 577: 162-179.
- [5] 钱榕,徐雪飞,刘晓豫,等. 引入度中心性选择攻击节点的对抗攻击算法[J/OL]. 计算机工程与应用. (2023-08-18)[2023-12-01]. <https://kns.cnki.net/kcms/detail/11.2127.TP.20230817.1423.014.html>.
- [6] FREEMAN L C. A Set of Measures of Centrality Based on Betweenness[J]. Sociometry, 1977, 40(1): 35.
- [7] BONACICH P. Factoring and Weighting Approaches to Status Scores and Clique Identification[J]. The Journal of Mathematical Sociology, 1972, 2(1): 113-120.
- [8] 郭程远,陈鸿昶,王庚润,等. 基于迭代K-shell和改进信息熵的节点重要性排序算法[J]. 信息工程大学学报, 2022, 23(5): 556-562.
- [9] 邹龙,李晨璞,刘凯,等. 度联合信息熵的网络重要节点识别[J]. 福建电脑, 2023, 39(5): 30-34.
- [10] 林欣,吴玉芹,冯玮,等. 节点度与邻域相似度标签传播算法[J]. 宁德师范学院学报(自然科学版), 2023, 35(3): 254-259.
- [11] ZHANG J P, DING X Y, YANG J. Revealing the Role of Node Similarity and Community Merging in Community Detection[J]. Knowledge-Based Systems, 2019, 165: 407-419.
- [12] 任晓龙,吕琳媛. 网络重要节点排序方法综述[J]. 科学通报, 2014, 59(13): 1175-1197.
- [13] 韩忠明,陈炎,李梦琪,等. 一种有效的基于三角结构的复杂网络节点影响力度量模型[J]. 物理学报, 2016, 65(16): 289-300.
- [14] YANG Y Z, YU L, WANG X, et al. A Novel Method to Evaluate Node Importance in Complex Networks[J]. Physica A: Statistical Mechanics and Its Applications,

