

DOI:10.16356/j.1005-2615.2025.02.020

## 基于邻域信息熵与有效距离的网络节点识别

张正勇, 苏健生, 姜敏勤, 杨 钰

(南京财经大学管理科学与工程学院, 南京 210023)

**摘要:** 为了克服现有关键节点识别技术存在的计算复杂性大、评估维度单一和应用范围有限等缺点, 构造了一个适用于关键节点评估的新算法。该算法首先通过分析节点的信息熵以及其邻居节点的影响力贡献, 评估节点的局部影响力, 从而消除了传统仅仅依赖节点度量为评估标准的瑕疵。其次, 该算法通过衡量节点间距离的相关性来确定节点的全局影响力, 有效解决了因考虑过多路径而导致的计算量激增的问题。为了充分论证算法的实用性, 借助单调性实验、传染病模型实验以及鲁棒性实验, 对 4 个规模各异的真实网络以及 6 种比较算法展开分析。最终结果显示该算法在准确性、有效性和识别能力等方面均有一定改善, 同时, 其计算复杂度较低, 可应用于稀疏的网络。

**关键词:** 复杂网络; 关键节点; 节点信息熵; 全局信息; 局部信息

中图分类号: TP393

文献标志码: A

文章编号: 1005-2615(2025)02-0387-10

## Network Node Recognition Based on Neighborhood Information Entropy and Effective Distance

ZHANG Zhengyong, SU Jiansheng, JIANG Minqin, YANG Yu

(School of Management Science and Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China)

**Abstract:** In order to overcome the shortcomings of the existing key node recognition technologies, such as high computational complexity, single evaluation dimension, and limited application scope, this paper constructs a novel algorithm suitable for key node evaluation. This algorithm first evaluates the local influence of a node by analyzing its information entropy and the influence contribution of its neighboring nodes, thereby eliminating the shortcomings of traditional evaluation criteria that rely solely on node metrics. Secondly, this algorithm determines the global influence of nodes by measuring the correlation of distances between them, effectively solving the problem of excessive computation caused by considering multiple paths. In order to fully demonstrate the practicality of the algorithm, four real networks of different scales and six comparative algorithms are analyzed using monotonicity experiments, infectious disease model experiments, and robustness experiments. The final results show that the algorithm has certain improvements in accuracy, effectiveness, and recognition ability. At the same time, its computational complexity is low and it can be applied to sparse networks.

**Key words:** complex networks; key nodes; node information entropy; global information; local information

**基金项目:** 国家自然科学基金(61602217); 江苏高校“青蓝工程”; 江苏省研究生科研与实践创新计划(KYCX23\_1794)。

**收稿日期:** 2024-08-22; **修订日期:** 2025-01-11

**通信作者:** 张正勇, 男, 副教授, E-mail: zyzzhang@nufe.edu.cn。

**引用格式:** 张正勇, 苏健生, 姜敏勤, 等. 基于邻域信息熵与有效距离的网络节点识别[J]. 南京航空航天大学学报(自然科学版), 2025, 57(2): 387-396. ZHANG Zhengyong, SU Jiansheng, JIANG Minqin, et al. Network node recognition based on neighborhood information entropy and effective distance[J]. Journal of Nanjing University of Aeronautics & Astronautics(Natural Science Edition), 2025, 57(2): 387-396.

复杂网络可用以模拟和表征真实世界中的多种复杂系统,如社交网络、电力网络和交通规划网络等<sup>[1]</sup>,这些网络结构在各自领域扮演着核心角色,包括且不限于引导社会舆论的扩散、控制疾病的蔓延以及优化交通流动等。复杂网络相关研究旨在揭示网络的普遍规律和内在机制<sup>[2]</sup>,涉及网络的动态性,包括节点和边的增加、删除,网络拓扑结构的演化过程<sup>[3]</sup>,以及网络鲁棒性和脆弱性研究,有利于更好地维护网络的安全和工作环境等<sup>[4]</sup>。在复杂网络中,一些特定的节点扮演着核心的角色,它们对网络的整体框架发挥着至关重要的作用<sup>[5]</sup>。在复杂网络研究中关键节点的准确识别极其重要,有助于揭示网络的特性、架构和功能。这项工作的重心在于定位那些扮演独特角色或具有特定功能的节点,比如,网络中的主要影响者、核心枢纽或异常行为的节点。近几十年来,它已成为各个领域研究者关注的共同焦点<sup>[6]</sup>,比如,在社交网络中管控舆论领袖的言辞,可以有效地减少谣言的扩散;在交通网络中施行对枢纽的人流管理,可以显著减轻整个交通系统的拥挤状况;在电力网络中养护核心变电站则有望预先避免级联故障<sup>[7-9]</sup>。因此,在网络科学领域,通过精准地识别这些关键要素、核心节点,人们能够更深入地揭示网络的运作原理和潜在规律。

迄今为止,学者们针对不同的现实需求情况提出了多种形式的关键节点识别算法,包括度中心性(Degree centrality, DC)、介数中心性(Betweenness centrality, BC)、特征向量中心性(Eigenvector centrality, EC)、PageRank 中心性(PageRank centrality, PRC)、接近中心性(Closeness centrality, CC)和 K-shell(Ks)分解法等<sup>[10-14]</sup>。上述算法中描述节点的中心值越大,表明节点越重要。此类算法计算相对较为简单、具有良好的借鉴意义,但由于在设计该类算法时,都各自关注网络中的不同方面,可能忽略了一些其他的重要因素。如 DC 算法、EC 算法、PRC 算法等,更多聚焦于节点的局部信息,对全局信息有所忽视。CC 算法和 BC 算法更关注网络的连通性,但它们在网络中高度集中的节点上可能存在敏感性问题的,以及存在计算复杂度高的问题。与同类算法相比,Ks 算法同时包含了单个节点和相邻节点之间的连接情况,从而可以更加全面的描述节点的重要性。然而,对于具有同样 Ks 数值的节点则无法进行区分,故许多学者针对 Ks 方法的缺点开展了改进研究,如 Wang 等<sup>[15]</sup>基于 Ks 值的迭代信息来区分相同 Ks 值的节点, Sun 等<sup>[16]</sup>提出了一种基于社区的 Ks 值分解方法来缓解重叠问题。Lai 等<sup>[17]</sup>利用节点和邻居节点度,考察了 3

个不同改进模型用以评估交通网络中的关键节点情况,结果显示加权 Ks 邻域识别模型的综合评估效果最好,具有一定的指导意义。

随着识别方法研究的不断深化,研究者们逐渐认识到,仅依赖单一中心性指标来衡量节点的重要性存在缺陷,于是,人们开始从多个维度探索和描述网络中的关键节点:(1)采用多维度评价体系来整合多样的节点中心性度量标准,Zhang 等<sup>[18]</sup>研究了 H 指数、CC、Ks 等多种属性,进而设计出了一种多属性网络决策指标。Yang 等<sup>[19]</sup>综合多重中心性测度,结合灰色关联分析和易感、感染、恢复模型,构建了一种基于理想解相似性排序(Technique for order preference by similarity to an ideal solution, TOPSIS)偏好的动态加权技术模型。研究表明,该方法比传统的加权 TOPSIS 方法具有更高的准确性。此类方法适用性较广,但由于不同属性间的权重可能会受到数据和指标选择的影响,某些属性可能会被过度强调或忽略,若综合过多指标,则可能会增加计算复杂度,尤其是面向大规模网络时,计算资源和时间的消耗常会显著增加。(2)在现有传统算法的基础上,开发一套全新的节点评价指标体系,比如, Singh 等<sup>[20]</sup>定义了图傅里叶变换中心性来量化复杂网络中节点的重要性,而 Wang 等<sup>[21]</sup>通过迭代去除网络内的节点来引入效率中心性。Bian 等<sup>[22]</sup>在已有的证据中心性(Evidential centrality, EVC)基础上,通过修正 EVC 的基本概率赋值强度,提出了一种新的证据中心性。此类方法相对传统的中心性指标,可以针对特定问题和网络数据的特点,提高节点重要性识别的有效性,但这种有效性可能会出现只适用于某些特定网络类型的情况,在不同的网络类型或数据集上可能会有一定的差异。(3)综合运用不同的机器学习算法来捕捉关键节点,Zhang 等<sup>[23]</sup>设计了 1 个命名为 CGNN 的深度学习方法,其合并了卷积神经网络(Convolutional neural network, CNN)和图神经网络(Graph neural network, GNN)。Asgharian 等<sup>[24]</sup>利用机器学习来克服启发式算法的约束,以此来识别网络中的重要节点。机器学习方法能够从数据中学习节点的特征表示,从而更好地捕捉节点的重要性,但其需要较多的数据来训练模型,故对于规模较小的网络或节点间分布较为稀疏的网络可能会表现不佳。(4)借助对节点全局信息或局部信息的剖析,来判定网络中的关键节点, Ma 等<sup>[25]</sup>从万有引力原理中获得灵感,提出了一种重力中心性模型,该模型将节点的连接数视为其“质量”,并将节点间的最短路径长度作为衡量它们“距离”的依据。但在现实网络中,每个节点的吸引力并不完全

一致,为了修复这一瑕疵,Yang等<sup>[26]</sup>引入了一个基于Ks的改进重力模型。此外,Tu等<sup>[27]</sup>构建了一种基于位置和邻域的新型重力模型,同时考虑了节点的位置和它们的局部信息。在此类算法中,仅依赖于节点的局部特征或全局特征可能无法全面捕捉节点在网络中的实际影响力,进而,需要综合局部信息和全局信息的识别方法来减少单一视角导致的评估偏差,以提高节点重要性识别的准确性,比如,Hu等<sup>[28]</sup>利用Ks层次信息获取全局影响,用节点的度和邻居节点的度及Ks层次信息确定局部影响,提出一种综合性的重要节点识别方法。

受现有研究的启发,为克服已有算法识别复杂网络关键节点的局限,本文研究构造了一种关键节点识别方法——FI(Final impact)算法<sup>[29]</sup>。节点信息熵的计算与节点的邻居或连接的信息有关,用来衡量系统的不确定程度,对于一个节点,如果其信息传递的不确定性越高,则节点的信息熵越高,表明节点值相近的邻居节点越多。此外,节点间的有效距离更符合现实网络的情况。据此,FI算法通过研究节点信息熵以及它的邻居节点影响,计算出节点的局部影响力。随后,该算法通过评估节点间距离的相关性,来确定节点的全局影响力。最后,在4个真实网络上,通过单调性实验、传染病模型实验和鲁棒性实验,分析了其识别出的重要节点的区分度、有效性和准确性,实验结果证实了该方法的有效性。

## 1 相关理论基础

对于给定的无向图 $G=(V,E)$ , $V$ 为节点集, $E$ 代表连接集,图中的总节点数为 $n,n=|V|$ , $G$ 的邻接矩阵集 $A=\{a_{ij}\}$ ,当节点 $i$ 与节点 $j$ 间有连接时, $a_{ij}=1$ ,否则 $a_{ij}=0$ 。

节点中心性是网络分析中用于评估节点相对重要性的关键指标,它可以通过以下几种方式来定义。

(1) DC:网络拓扑中,节点的度是一个核心的度量指标,DC是一种常见的节点重要性评估技术,对应节点 $i$ 的计算公式为

$$DC_i = \frac{k_i}{n-1} \quad (1)$$

式中 $k_i$ 表示节点 $i$ 的度值。

(2) CC:旨在考虑节点 $i$ 在网络中与其他节点间的距离,其计算公式为

$$CC_i = \frac{n-1}{\sum_{j=1}^n d_{ij}} \quad (2)$$

式中: $d_{ij}$ 为节点 $i$ 至节点 $j$ 间的最短距离,如果节点 $i$ 与节点 $j$ 之间没有可达路径,则 $d_{ij}=\infty$ 。

(3) EC:在网络中,一个节点的影响力不仅由其直接连接的邻居数量决定,还受到这些邻居节点自身重要性的影响,计算公式为

$$EC_i = \lambda^{-1} \sum_{j=1}^n a_{ij} e_j \quad (3)$$

式中: $\lambda$ 为邻接矩阵 $A$ 的最大特征值, $e_j$ 为对应于 $\lambda$ 的第 $j$ 项特征的向量值。

(4) Ks:包括基于其拓扑位置划分网络结构,有效地将网络分层为相对于度中心性的子结构。根据节点度的更新然后从网络中迭代地删除节点来实现,最终确定每个节点的Ks值,通常表示为Ks中心性。Ks值越大,表示它越接近网络的核心层。

(5) 局部和全局的中心性(Local and global centrality,LGC):通过考虑节点度的比例和网络中两个节点间的最短距离来量化节点的重要性,计算公式为

$$LGC_i = \frac{d_i}{n} \times \sum_{i \neq j} \frac{\sqrt{d_j + \alpha}}{d_{ij}} \quad (4)$$

式中 $\alpha$ 表示一个可调参数。

(6) 局部引力模型(Local gravity model,LGM):仅考虑截断半径 $R$ (近似与网络平均距离成线性关系)范围内的节点相互作用,计算公式为

$$LGM_i = \sum_{d_{ij} \leq R, j \neq i} \frac{k_i k_j}{d_{ij}^2} \quad (5)$$

## 2 本文算法

目前,越来越多的算法基于节点度来获取节点的局部影响,但这一指标只考虑了最近邻居节点的信息,在挖掘网络中的关键节点时具有一定局限性。Qiu等<sup>[30]</sup>研究指出,若一个节点有较多的邻居节点,邻居节点的度值较高且分布均匀,则该节点具有较强的影响力,在这种情况下可以利用节点信息熵来克服节点度指标的缺陷。故本文引入节点信息熵来计算节点的局部影响力。下面,将对FI算法所需的相关定义展开阐述。

**定义1** 信息熵。以信息量来度量特定事件所带来的信息,其期望值即为信息熵<sup>[31]</sup>,熵越大,信息的不确定性越大。假设 $X=\{x_1, x_2, \dots, x_n\}$ 为 $n$ 个可能的事件的集合, $P_i$ 为事件 $x_i$ 发生的概率, $X$ 的信息熵为 $H(X)$ ,计算公式为

$$H(X) = - \sum_{i=1}^n P_i \log_2 P_i \quad (6)$$

**定义2** 节点信息熵。将熵的概念推广到复

杂网络中,称为节点信息熵(Information entropy, IE)<sup>[32]</sup>。节点  $i$  的信息熵计算公式为

$$IE_i = - \sum_{j \in \Gamma_i} P_j \log_2 P_j \quad (7)$$

式中:  $\Gamma_i$  为节点  $i$  的相邻节点的集合,  $P_j$  为节点  $x_j$  的发生概率。

**定义 3** 邻域熵影响力。每个邻居节点对节点  $i$  的影响力贡献不同。如果一个节点有大量的邻居节点,同时邻居节点的度中心性较高,可以认为节点  $i$  是一个重要性较高的节点<sup>[33]</sup>。然而,度中心性只考虑节点的最近邻节点,所以为了考虑高阶邻居节点的影响,本文结合节点信息熵,给出了节点  $i$  的邻居节点的度,从而反映了节点  $i$  在信息扩散时在邻居节点中的扩展能力。节点  $i$  邻域信息熵(Neighborhood information entropy, NEI)影响的计算公式为

$$NEI_i = IE_i \times \sum_{j \in \Gamma_i} k_j \quad (8)$$

式中  $k_j$  为节点  $j$  的度。该定义得到了每个邻居节点对节点整体影响的细微贡献,同时结合了相邻节点的度和信息熵,从而为评估该节点在网络中的重要性提供了更全面的评估。

**定义 4** 有效距离。计算节点间距离最基本的方法就是将任意相邻节点间的距离设置为 1,但是这种方法不符合现实网络中信息流相互影响的原则<sup>[34]</sup>。在真实网络中,影响节点通信的因素除了与边数和节点间的距离有关外,还与节点间信息流传输的大小有关。根据节点与相邻节点的连通性和节点固有属性的重要性,将信息流分配给不同的路径,故任何节点之间的距离都不应该总是为 1。Brockmann 和 Helbing<sup>[35]</sup>提出了有效距离的概念,以克服传统的节点间距离计算方法的局限。节点间的有效距离是根据自身节点的流入量和相邻节点的总流出量计算的,不仅考虑了全局网络结构,还考虑了周围相邻节点的局部信息。任意两个节点之间的有效距离计算公式为

$$d(i, j) = 1 - \log_2 p_{ji} \quad (9)$$

式中:  $p_{ji}$  为节点  $i$  与节点  $j$  之间的信息流之比,并且,  $p_{ji} = \frac{a_{ij}}{\sum_{j \in \Gamma_i} a_{ij}}$ , 当节点  $i$  和相邻节点  $j$  之间存在连

接时,  $a_{ij} = 1$ , 否则,  $a_{ij} = 0$ 。同时,任意两点之间的距离都大于 1 且  $d(i, j) \neq d(j, i)$ 。这个定义总结了信息流路径的微妙本质,同时考虑了节点的连接性和内在属性,从而提供了网络交互中更真实的距离描述。

**定义 5** 全局影响。将节点的度作为一个内在属性来衡量该节点的信息分布。度越高,意味着

接收信息流的可能性越大。此外,节点间的距离因素也必须被纳入,因为它与该节点的影响成反比。由此,本文将节点的全局影响(Global influence, GI)定义为节点的度与节点间有效距离的比值,计算公式如下

$$GI_i = \sum_{i \neq j} \frac{\sqrt{k_i + 1}}{d(i, j)} \quad (10)$$

式中:平方根用于规范化节点的影响力。通过结合节点的度和有效距离,该定义总结了信息分布和网络拓扑之间的相互作用,提供了一个节点全局影响的整体度量。

**定义 6** 最终影响。节点的最终影响取决于其在网络中邻域熵的影响力和全局影响力,最终可通过以下公式计算

$$FI_i = NEI_i \times GI_i \quad (11)$$

算法 FI 步骤如下:

输入:  $G = (V, E)$

网络  $G$  包含  $n$  个节点  $E$  条边。

(1) 对于  $i = 1 \sim n //$  主循环,计算所有节点的最终影响(FI):

(a) 确定节点  $v_i$  的度并计算  $P(v_i)$ ;

(b) 利用式(7)计算节点  $v_i$  的信息熵( $IE(v_i)$ );

(c) 利用式(8)计算节点  $v_i$  的邻域熵影响力( $NEI(v_i)$ );

(d) 利用式(9)计算节点  $v_i$  与相邻节点间的有效距离;

(e) 利用式(10)计算节点  $v_i$  的全局影响( $GI(v_i)$ );

(f) 利用式(11)计算节点  $v_i$  的最终影响( $FI(v_i)$ );

(2) 对所有节点的最终影响(FI)进行排序。

输出:  $FI(v_i)$

### 3 实验分析

研究所涉及的实验都是在一个标准化的计算机配置上执行的。该实验使用了一个 16 核的 Rip-tide 5600U 处理器,在微软 Windows 10 操作系统下运行,并配备了 32 GB 内存。

#### 3.1 数据集

为了分析 FI 算法的有效性,在 4 个不同的真实数据集上进行了一系列的实验研究。这些数据集信息介绍如下:(1) 电子邮件网络(E-mail)。该网络由 1 133 个节点和 5 451 条边组成,是指基于人们之间电子邮件的交流形成的网络结构,属于社交网络的一种,其中节点代表人,连边代表电子邮件往来,在 E-mail 网络中,有少数节点具有较高的连接

度,而大部分节点具有较低的连接度,这也代表了E-mail网络的无标度特性。(2)Karate网络。由34个节点和78条边组成,在Karate网络中节点可以明确的分为两个社区,该网络也是社团结构网络的典型代表。(3)成渝城际轨道交通网络(Chengdu Chongqing intercity rail transit network, C-Y TN)。由75个节点和110条边组成,该网络是交通网络的具体实例,节点代表车站,边代表车站之间的轨道连接,交通网络大多数节点之间的最短路径较短,这也体现了交通网络的小世界特性。(4)Jazz音乐家的合作网络。包含198个节点和2742条边,该网络代表了一种特定的社交网络结构。选取这4个网络,可以很好地反映复杂网络在不同领域的应用和特性,从社交网络到交通网络,每种网络都有其独特的结构特征和动态行为。以上数据集可以公开访问,具体可从“http://snap.stanford.edu/data/as-caida.html”和“http://networkrepository.com”获得。各网络的拓扑特征如表1所示。表中: $|E|$ 和 $|V|$ 为边数和节点数, $\langle K \rangle$ 为平均度,CC为网络的平均聚类系数, $\rho$ 为网络密度, $\beta_{th}$ 为传染病模型的传染率阈值。

表1 4个真实世界网络的统计特征

Table 1 Statistical characteristics of four real-world networks

网络	$ V $	$ E $	$\langle K \rangle$	CC	$\rho$	$\beta_{th}$
E-mail	1 133	5 451	9.62	0.220 2	0.008 5	0.14
Karate	34	78	4.59	0.047 0	0.039 6	0.13
C-Y TN	75	110	2.93	0.570 6	0.139 0	0.41
Jazz	198	2 742	27.67	0.617 5	0.140 6	0.27

### 3.2 评价指标

#### 3.2.1 单调性指数

节点单调性是区分复杂网络中节点传播能力和等级的关键度量指标之一。在对节点的重要性进行排序的情况下,如果每个节点都表现出不同的排序值,则该方法可以精确地区分节点的重要性。节点的重要性值相同的节点越少,表示该方法的分辨率越高。在实验分析中,采用单调性指数 $M(R)$ 来度量不同方法的分辨率。单调性度量的计算公式为

$$M(R) = \left[ 1 - \frac{\sum_{r \in R} n_r(n_r - 1)}{n(n-1)} \right]^2 \quad (12)$$

式中: $R$ 代表体现每个节点重要程度的排序向量, $n$ 代表网络节点的总数, $n_r$ 代表在节点重要性排序结果中具有相同重要性得分的节点数量。

#### 3.2.2 传染病(SI)模型

不同的传染病模型在复杂网络中的传播机制也各不相同,其中SIR模型和SI模型被广泛用于模拟信息传播,分别对应易感(Susceptible, S)、感染(Infected, I)和康复(Recovered, R)。当SIR模型不考虑恢复状态时,它就变成了SI模型,它也可以用于评估节点在网络中的传播能力,以反映节点的影响。由于本文只考虑节点的影响,不考虑节点的恢复状态,故选择SI模型来评估节点的传播能力更为合理。

当SI模型受感染的节点无法恢复时,导致随着时间 $t$ 的推移,受感染的节点数量逐渐增加,同时易感的节点也会减少,最终网络中没有易感的节点。本文将时间 $t$ 内受感染节点的比例记为 $F(t)$ ,计算公式为

$$F(t) = \frac{N_r}{N} \quad (13)$$

式中: $N_r$ 表示 $t$ 时刻状态下被感染的节点数, $N$ 为网络的总节点数。可以看出,当节点的初始影响越大时, $F(t)$ 值也越大,节点就越重要。参考文献研究得到了传播阈值 $\beta_{th}$ ,  $\beta_{th} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$ ,  $\langle k^2 \rangle$ 代表度平方的平均值<sup>[36]</sup>。

#### 3.3 鲁棒性评价指标

该评价指标能够衡量网络在受到攻击或扰动时保持其结构和功能的能力大小,并且可以用来衡量在一个网络中节点的重要性程度,研究从鲁棒性的角度出发,使用最大连通子图比例作为一个标准来评估和比较各个算法的有效性。最大连通子图表示通过去除节点将网络分解为子网络后的最大连通分量<sup>[37]</sup>。本文通过比较删去特定节点前后网络中最大连通子图的规模变化来评估这些节点在整体网络结构中的重要性,最大连通子图的比例将使用如下计算公式获得。

$$S = \frac{N'}{N} \quad (14)$$

式中: $S$ 表示最大连通子图的比例; $N'$ 表示最大连通子图所含有的节点数; $N$ 为网络的总节点数。 $S$ 越小,网络崩溃的速度越快,说明被删除的节点越重要。

#### 3.4 实验结果

计算复杂度较低也是FI算法的优势所在,如表2所示,给出了FI算法与其他算法的计算复杂度比较情况。DC的复杂度为 $O(n)$ ,是最低的,但研究表明DC的实验性能并不理想。本文提出的FI算法主要分为两个部分:(1)全局影响力,通过计算节点与其邻居节点之间的有效距离来实现,这

表 2 FI 算法与其他算法的复杂度比较

Table 2 Comparison of complexity of the FI algorithm with the other algorithms

方法	复杂度
CC	$O(nm)$
EC	$O(n^2)$
Ks	$O(n)$
LGC	$O(n^2)$
DC	$n^2$
LGM	$O(nk^2)$
FI	$O(nk^2)$

需要遍历网络中所有节点的邻居节点,其时间复杂度为  $O(nk^2)$ ; (2) 通过计算节点信息熵得到节点局

部影响力,这一步的时间复杂度为  $O(nk)$ ,  $k$  为每个节点的邻居数量。因此,FI 算法的计算复杂度为  $O(nk^2)$ 。

实验通过使用单调性指标  $M(R)$ 、SI 模型和鲁棒性评价指标  $S$  来验证 FI 算法的有效性,将获得的结果与 DC、CC、EC、Ks、LGC、LGM 等算法进行比较分析。

3.4.1 单调性实验分析

对各类算法的单调性进行了度量分析。表 3 记录了 4 个真实网络中不同算法的单调性值,其中粗体表示最优值。此外,图 1 给出了一个可视化的比较。

表 3 不同排序算法的单调性指标  $M$

Table 3 Monotonicity index  $M$  of different sorting algorithms

超网络	$M(\text{DC})$	$M(\text{CC})$	$M(\text{EC})$	$M(\text{Ks})$	$M(\text{LGC})$	$M(\text{LGM})$	$M(\text{FI})$
Karate	0.707 9	0.899 3	<b>0.954 2</b>	0.549 9	<b>0.954 2</b>	0.947 2	<b>0.954 2</b>
C-Y TN	0.444 0	0.987 1	<b>0.999 9</b>	0.100 1	0.959 3	0.923 7	0.947 4
Jazz	0.965 9	0.987 8	0.989 3	0.895 1	<b>0.999 2</b>	0.998 1	0.998 2
E-mail	0.887 4	0.998 8	0.995 9	0.852 1	0.997 0	0.985 6	<b>0.999 7</b>

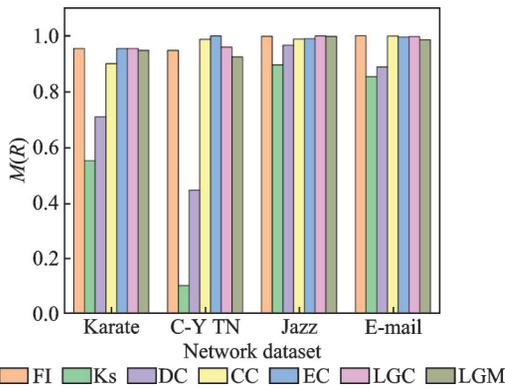


图 1 不同排序算法的单调性指标  $M$

Fig.1 Monotonicity index  $M$  of different ranking algorithms

可以看出,FI 算法除了在成渝城际交通网络中的表现略为逊色外,在其余 3 个真实网络中的表现都较好,其单调性较强。而 Ks 算法在各个网络中的表现均不佳,这可能是因为 Ks 算法的本质是将节点按其度数进行递减排序,导致具有相同度数的节点被分配到同一个 Ks 层次,故造成单调性较差。在 Karate 网络中,FI 算法与 EC、LGC 算法具有相同的单调性。在 C-Y TN 中,EC 算法的单调性最接近于 1,其次是 CC 算法和 LGC 算法,FI 的表现稍差。在 Jazz 网络中其单调性仅次于 LGC 算法,且单调性值接近于 1,表明 FI 算法能够较好地将节点的重要性区分开来。E-mail 网络中 FI 算法的单调性最强,其次是 CC 算法,LGC 算法不如 FI 算法的原因可能是由于在考虑节点的局部影响力时,FI 在节点本身的基础上还考虑了网络的邻居

节点的影响力,而 LGC 仅考虑了节点本身的影响力。另外,在节点的全局影响力部分,FI 算法对节点间的距离做出了改进,一定程度上提高了算法的有效性,故 FI 算法的单调性在 E-mail 网络上稍优于 LGC 算法。

3.4.2 SI 模型实验分析

在本实验中,采用 SI 模型来评估 FI 算法所识别的关键节点对网络传播力的影响情况。在 4 个真实世界的网络中进行传播过程的模拟,对比不同算法在稳定状态时感染节点的比例,以此评估节点的重要性。实验依据各个算法所得到的重要节点,选取排序在前 10% 的节点为初始感染节点,通过 SI 模型感染整个网络。为了使方法更具有合理性,实验将每种方法重复 100 次。最后,通过计算最终的感染节点所占比例  $F(t)$  来确定初始节点的影响。为确保传播过程能顺利进行,实验将传播概率设置为  $\beta = 1.5\beta_m$ 。实验结果如图 2 所示。由图 2 可知,曲线的快速上升和最终感染节点数量的增加表明初始感染节点具有较大的影响力,这些节点在网络中的重要性也更高。相应地,排序算法的性能也被认为更优越。图 2(a,c,d) 中显示出在实验传播的早期阶段,这些方法之间的差异并不显著,说明在短时间内这些方法所选取的前 10% 的节点影响基本一致,不过随着时间的增加,可以明显看出 FI 算法逐渐超过了其他算法,且在 Jazz 网络中与其他算法拉开了较大的距离,表明 FI 算法较其他算法更有效率。在图 2(b) 中,FI 算法较之其他

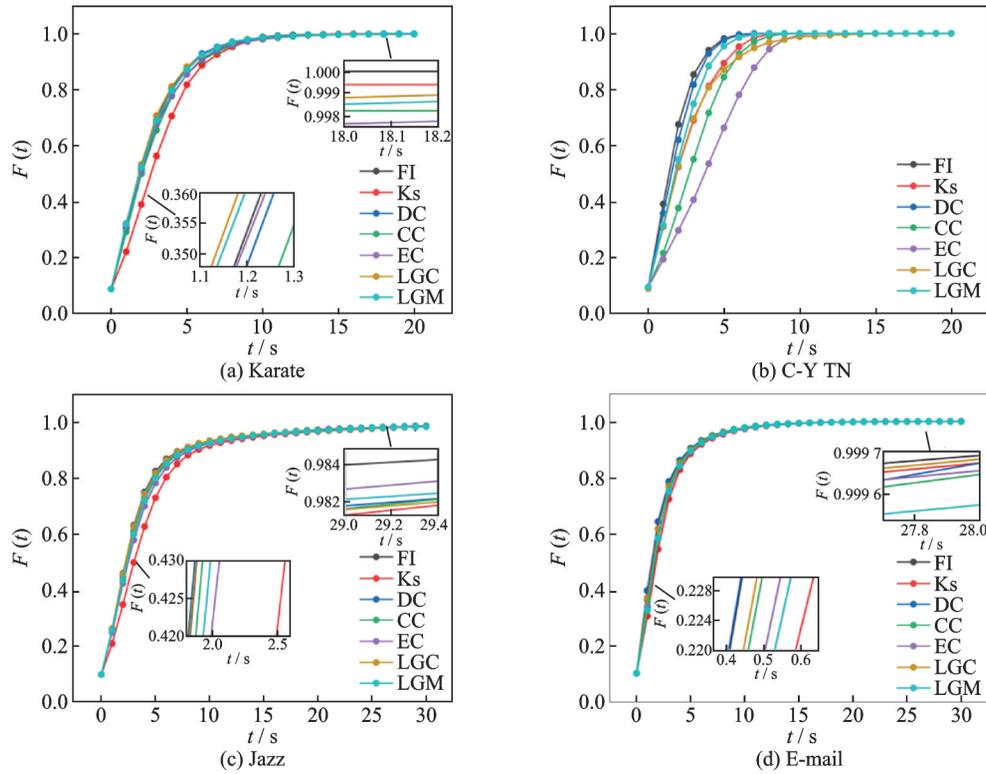


图 2 使用FI和其他算法选择的前10%个节点对网络的影响情况  
 Fig.2 Effect of the top 10% of nodes selected using FI and other algorithms on the network

的算法显示得更加明显的差异。而DC算法的相应曲线在图2(a,c,d)中上升速度最慢,显示出DC算法计算复杂度虽然较低,但性能并不佳。

### 3.4.3 鲁棒性实验分析

为进一步讨论评估算法识别关键节点的有效性,本组实验分别将前文6种算法与FI算法,在4个真实网络上得到的重要节点进行排序,按重要性排序从网络中依次删去节点,通过分析后续网络最大连通子图比例的变化,评估7种不同节点重要性评价算法的效果。删去节点后网络的最大连通子图比例变化曲线如图3所示。Jazz网络和E-mail网络规模大,节点的数量太多,为了清晰展示出网络最大连通子图比例的变化趋势,在绘图时,设置Jazz网络每隔3个点显示1个节点,E-mail网络每隔15个点显示1个节点。

图3(a)分析可以发现,在Karate网络中,当删去2个节点后,FI算法的下降曲线与EC、LGM、Ks的曲线相吻合,CC对应的曲线下降速度最快。然而,在删去4个节点时,FI算法相关网络的最大连通子图比例数值从76.5%快速下降到29.4%,网络的连通性受到了显著的影响,删去17个节点时,最大连通子图比例数值仅为5.9%,网络几乎崩溃,这一结果显示对于小型网络,只需删去少量FI算法识别出的排序靠前节点,便能够显著影响网络的连通性,凸显了FI算法的效能。此外,DC算法在

删去16个节点后,最大连通子图比例数值也可以降到5.9%,表面上看该算法似乎比FI算法更早导致网络解体,不过据3.4.2实验可知DC算法的表现并不理想,且仅领先一个节点的优势也并不显著。故总体而言,FI算法在识别关键节点方面的表现优异。就LGC算法来说,它在本实验中表现不佳,直到累计删去达到32个节点以后,才对网络的连通性造成较大影响,表明该算法虽然在单调性实验和SI模型实验中表现较好,但从连通性角度衡量其性能时并不占优势。

如图3(b)所示,在C-Y TN网络中,当删去前10个节点,7种算法的最大连通子图比例数值比较接近,随着删去节点到11个,算法间的曲线差异逐渐显现,FI算法对应的最大连通子图比例数值迅速降到38.7%,此时,DC和LGM算法曲线也显示出相似情况。继续删去12到21个节点,FI算法曲线显著下降,随后FI算法曲线开始保持持续领先。

如图3(c)所示,在Jazz网络中,7种算法的表现均不可观且差异不明显。当累计删去105个节点时,CC算法对应的最大连通子图比例曲线与其他算法拉开了较为明显的距离,最大连通子图比例数值锐减至21.2%,而其他算法对应的最大连通子图比例数值仍保持在44%左右,这可能是由于CC算法可以很好捕捉整个网络的连通性,具有较高接近中心性的节点通常在网络中担任桥梁角色,这些

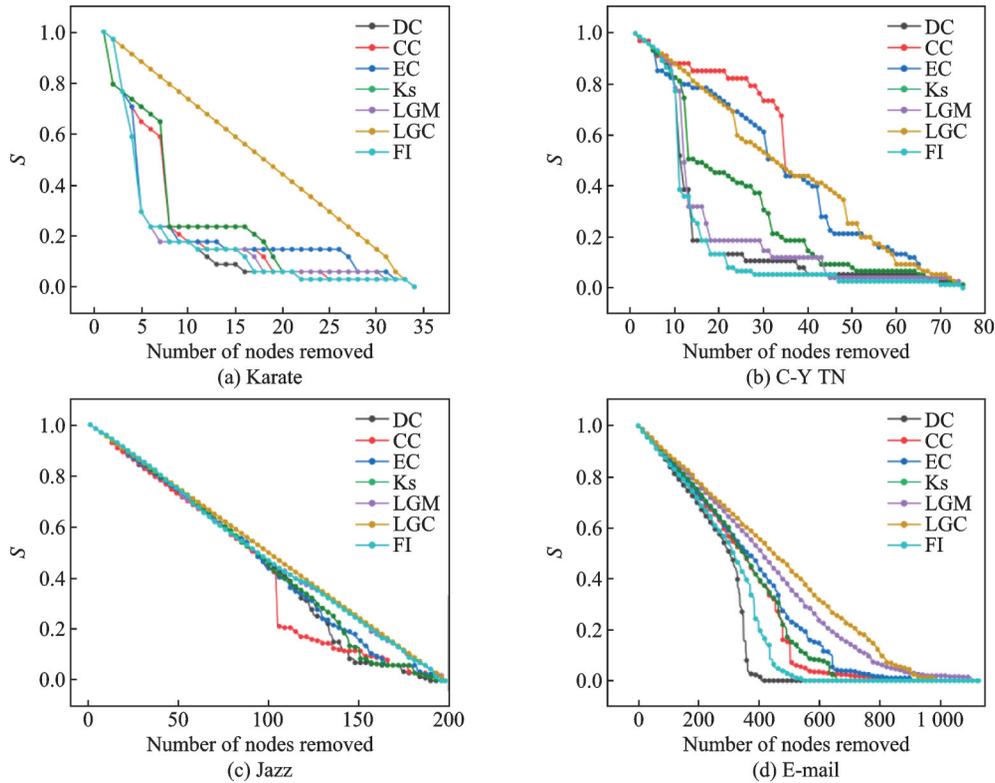


图3 不同算法得到排序结果后进行节点删去产生的最大连通子图比例变化情况

Fig.3 Proportion change of the largest connected subgraph generated by node removal after obtaining sorting results through different algorithms

节点在信息传播和影响力传播中起到关键作用。另外,可能 Jazz 网络的拓扑结构分布更适应 CC 算法,该网络聚集性较高、有较为明显的幂律特征且存在一些距离较远的节点,CC 算法考虑了节点与其他节点的平均距离,可以较好地应对这些问题。不过,在大型网络中,计算 CC 可能会变得非常耗时和复杂,且 CC 算法以节点间的距离为衡量标准,在某些特定场景下可能不是最合适的选择。比如,在社交网络中,节点间的距离可能并不直接关系到节点的重要性,此外,CC 算法仅考虑节点与其他节点的最短路径,忽略了可能通过多个中间节点到达其他节点的情况,可能会略去一些节点在传播中的突出作用。综上,尽管 CC 算法在 Jazz 网络中的表现相对略好,但随着删去节点的数量增加,各算法间的差异并不明显。LGC、LGM 和 FI 算法得出的结果基本一致,可能是因为 LGC 与 FI 算法都是从网络的局部特征和全局特征出发,而 Jazz 网络结构比较稠密,其局部特征表现更为明显,故按此类算法得到的节点的重要性值可能相对均衡,删去这些节点,并不会对网络的连通性造成较大影响。就 LGM 算法而言,可能是因为实验在进行算法对比时,其截断半径近似取为网络平均距离的  $1/2$ ,这就导致只有那些比较近的邻居节点对节点的重要性值计算有实质的贡献,这样,较远的节点

的影响会被削弱,从而导致在删去节点后,对网络连通性的影响不大。

由图 3(d)可以看出,在 E-mail 网络中,当删去排序靠前的 200 个节点时,各算法对应的曲线下降趋势没有太大的差异。但是当删去的节点数超过 300 时,则产生了较为明显的差异,在所讨论的网络中,DC 算法导致的最大连通子图比例曲线下降速度最为显著。当累计删去 365 个节点后,该算法对应的网络最大连通子图比例数值急剧下降到 3.8%,网络达到崩溃状态。FI 算法对应的最大连通子图比例曲线与 DC 算法得到的曲线下降趋势大致相同,但 FI 的下降速率较慢。在网络中删去累计 473 个节点后,FI 算法对应的最大连通子图比例数值降低到了 3.8%,说明从网络的连通度来看,DC 算法得到的重要节点,具有较高的连接数量和影响,它们被删去后会导致网络分裂成多个较小的子图,因此,导致了网络连通度的下降。而 FI 算法得到的重要节点被删去后对网络连通性的破坏程度不及 DC 算法,其原因可能是由于 FI 算法得到的排序较为靠前的节点间可以互相弥补,即使删去了其中一部分重要节点,但其他重要节点仍然可以维持网络的连通性,因此,需要删去一定规模的重要节点,减少网络中具有互补作用的节点数量,才会使网络的连通性发生较为明显的变化。就

LGC与LGM算法而言,随着重要节点的删去,网络的连通性被破坏的速度较为平缓,这可能是因为LGM算法所引入的截断半径在计算节点重要性值时,只考虑了一定距离内的邻居节点,节点的重要性值只受到近邻节点的影响,导致节点删去后,网络连通性所受到的影响较小。对于LGC算法来说,可能是因为在计算节点的重要性值时,考虑了节点度数的比例和邻居节点的重要性,在一定程度上平衡了节点的重要性,不会过度依赖度值较高的节点。此外,当一个节点度值较大,即使与邻居节点的最短路径较长,其重要性值仍然可能相对较大,从而保持了网络的一定连通性。而FI算法对应的最大连通子图比例曲线下降速度领先于LGC和LGM算法,原因可能是因为FI算法考虑了节点与邻居节点的信息熵,如果某个节点与其邻居节点的关联程度较低,即信息熵较大时,删去此类节点后,可能会导致部分信息传递能力的损失,使得网络连通度快速下降。这种算法间的差异也强调了网络的复杂性和节点重要性的多样性,不同节点重要性的计算方法和网络的拓扑结构都会导致不同的连通性变化,故应从多个角度对重要节点的识别算法的性能进行评估。

## 4 结 论

本文首先从熵的概念出发,分析了节点本身的集聚特征,然后从整个网络的角度出发,引入了有效距离的概念来改进节点全局影响的模型。在此基础上,构建了一种新型的复杂网络关键节点识别算法FI,该算法同时考虑了网络的局部特征和全局特征。计算复杂度为 $O(nk^2)$ ,算法从多个角度出发,充分考虑了节点本身的影响力、与邻居节点间的关联程度以及信息的传播能力来综合评估网络中节点的重要性度量。通过单调性实验、传染病模型实验及鲁棒性实验,在4个真实网络数据集对FI算法的有效性和准确性进行评估,结果表明,FI算法在真实网络中具有较高的区分度,且具有明显的扩散优势,能够跨不同类型的网络识别出具备强大传播力的节点。在网络的连通性方面,受网络拓扑结构特征及节点重要性互补的影响,总体而言,FI算法依旧取得了较好的效果,进一步体现了本文算法的可行性与有效性。

### 参考文献:

[1] LYNN C W, PAPADOPOULOS L, KAHN A E, et al. Human information processing in complex networks[J]. Nature Physics, 2020, 16(9): 965-973.  
 [2] 刘涛,陈忠,陈晓荣.复杂网络理论及其应用研究概述[J].系统工程,2005,23(6):1-7.

LIU Tao, CHEN Zhong, CHEN Xiaorong. A brief review of complex networks and its application[J]. Systems Engineering, 2005, 23(6): 1-7.  
 [3] 周建林,樊瑛,狄增如.复杂网络进展[J].北京师范大学学报(自然科学版),2023,59(5):691-704.  
 ZHOU Jianlin, FAN Ying, DI Zengru. Progress in complex networks[J]. Journal of Beijing Normal University (Natural Science), 2023, 59(5): 691-704.  
 [4] 楼洋,李均利,李升,等.复杂网络能控性鲁棒性研究进展[J].自动化学报,2022,48(10):2374-2391.  
 LOU Yang, LI Junli, LI Sheng, et al. Recent progress in controllability robustness of complex networks [J]. Acta Automatica Sinica, 2022, 48(10): 2374-2391.  
 [5] 刘小洋,何道兵,刘超,等.基于时变阻尼运动的社交网络信息传播动力学分析[J].中国科学:信息科学,2021,51(11):1867-1884.  
 LIU Xiaoyang, HE Daobing, LIU Chao, et al. Social network information propagation dynamic analysis based on time-varying damping motion[J]. Scientia Sinica Informationis, 2021, 51(11): 1867-1884.  
 [6] 姜敏勤,石小晶,姚安阳,等.基于CSSCI数据的复杂网络研究热点、知识演进与趋势探析[J].武汉商学院学报,2022,36(2):74-79.  
 JIANG Minqin, SHI Xiaojing, YAO Anyang, et al. A probe into the hotspots, knowledge evolution and research trends of complex networks based on CSSCI data[J]. Journal of Wuhan Business University, 2022, 36(2): 74-79.  
 [7] ZHANG W, CHEN Y, FIUMARA G. Research on evolutionary model and dynamic analysis for internet public opinion spreading[J]. Complexity, 2021(1): 8788960.  
 [8] LIU Y Y, SONG A B, SHAN X, et al. Identifying critical nodes in power networks: A group-driven framework[J]. Expert Systems with Applications, 2022, 196: 116557.  
 [9] 姜敏勤,石小晶,杨钰,等.基于节点重要度动态评估的复杂网络级联失效分析[J].科学技术与工程,2023,23(31):13482-13491.  
 JIANG Minqin, SHI Xiaojing, YANG Yu, et al. Complex network cascade failure analysis based on dynamic evaluation of node importance[J]. Science Technology and Engineering, 2023, 23(31): 13482-13491.  
 [10] PAGE L, BRIN S, MOTWANI R, et al. The pagerank citation ranking: Bringing order to the web: SIDL-WP-1999-0120[R]. [S.l.]: Stanford Infolab, 1999.  
 [11] FREEMAN L. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40: 35-41.  
 [12] BONACICH P. Factoring and weighting approaches to status scores and clique identification[J]. The Journal of Mathematical Sociology, 1972, 2(1): 113-120.  
 [13] KITSACK M, GALLOS L K, HAVLIN S, et al.

- Identification of influential spreaders in complex networks[J]. *Nature Physics*, 2010, 6(11): 888-893.
- [14] OKAMOTO K, CHEN W, LI X Y. Ranking of closeness centrality for large-scale social networks [C]//Proceeding of Frontiers in Algorithmics. Lecture Notes in Computer Science. Berlin Heidelberg: Springer, 2008: 186-195.
- [15] WANG Z X, ZHAO Y, XI J K, et al. Fast ranking influential nodes in complex networks using a K-shell iteration factor[J]. *Physica A: Statistical Mechanics and its Applications*, 2016, 461: 171-181.
- [16] SUN P G, MIAO Q G, STAAB S. Community-based K-shell decomposition for identifying influential spreaders[J]. *Pattern Recognition*, 2021, 120: 108130.
- [17] LAI Q, ZHANG H H. Analysis of identification methods of key nodes in transportation network[J]. *Chinese Physics B*, 2022, 31(6): 68905.
- [18] ZHANG Y H, LU Y L, YANG G Z, et al. Multi-attribute decision making method for node importance metric in complex network [J]. *Applied Sciences*, 2022, 12(4): 1944.
- [19] YANG P L, XU G Q, CHEN H P. Multi-attribute ranking method for identifying key nodes in complex networks based on GRA[J]. *International Journal of Modern Physics B*, 2018, 32(32): 1850363.
- [20] SINGH R, CHAKRABORTY A, MANOJ B S. GFT centrality: A new node importance measure for complex networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2017, 487: 185-195.
- [21] WANG S S, DU Y X, DENG Y. A new measure of identifying influential nodes: Efficiency centrality[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2017, 47: 151-163.
- [22] BIAN T, DENG Y. A new evidential methodology of identifying influential nodes in complex networks[J]. *Chaos, Solitons & Fractals*, 2017, 103: 101-110.
- [23] ZHANG M, WANG X J, JIN L, et al. A new approach for evaluating node importance in complex networks via deep learning methods [J]. *Neurocomputing*, 2022, 497: 13-27.
- [24] ASGHARIAN R A, MUNOZ J, JALILI M, et al. A machine learning-based approach for vital node identification in complex networks[J]. *Expert Systems with Applications*, 2023, 214: 119086.
- [25] MA L L, MA C, ZHANG H F, et al. Identifying influential spreaders in complex networks based on gravity formula[J]. *Physica A: Statistical Mechanics and its Applications*, 2016, 451: 205-212.
- [26] YANG X, XIAO F Y. An improved gravity model to identify influential nodes in complex networks based on K-shell method[J]. *Knowledge-Based Systems*, 2021, 227: 107198.
- [27] TU D Q, XU G Q, MENG L. GPN: A novel gravity model based on position and neighborhood to identify influential nodes in complex networks[J]. *International Journal of Modern Physics B*, 2021, 35(17): 2150183.
- [28] HU H F, SUN Z J, WANG F F, et al. Exploring influential nodes using global and local information[J]. *Scientific Reports*, 2022, 12(1): 22506.
- [29] 张正勇, 姜敏勤, 杨钰, 等. 一种基于邻域信息熵和有效距离的复杂网络关键节点挖掘方法: CN202311038200.1[P]. 2023-11-10.
- ZHANG Zhengyong, JIANG Minqin, YANG Yu, et al. A key node mining method of complex network based on neighborhood information entropy and effective distance: CN 202311038200.1[P]. 2023-11-10.
- [30] QIU L Q, ZHANG J Y, TIAN X B, et al. Identifying influential nodes in complex networks based on neighborhood entropy centrality[J]. *The Computer Journal*, 2021, 64(10): 1465-1476.
- [31] GUO C G, YANG L W, CHEN X, et al. Influential nodes identification in complex networks via information entropy[J]. *Entropy*, 2020, 22(2): 242.
- [32] 邓小龙, 王柏, 吴斌, 等. 基于信息熵的复杂网络社团划分建模和验证[J]. *计算机研究与发展*, 2012, 49(4): 725-734.
- DENG Xiaolong, WANG Bai, WU Bin, et al. Modularity modeling and evaluation in community detecting of complex network based on information entropy[J]. *Journal of Computer Research and Development*, 2012, 49(4): 725-734.
- [33] QIU L Q, LIU Y Y, ZHANG J Y. A new method for identifying influential spreaders in complex networks[J]. *The Computer Journal*, 2024, 67(1): 362-375.
- [34] 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.
- [35] BROCKMANN D, HELBING D. The hidden geometry of complex, network-driven contagion phenomena [J]. *Science*, 2013, 342(6164): 1337-1342.
- [36] CASTELLANO C, PASTOR S R. Thresholds for epidemic spreading in networks[J]. *Physical Review Letters*, 2010, 105(21): 218701.
- [37] 陆靖桥, 傅秀芬, 蒙在桥. 复杂网络的鲁棒性与中心性指标的研究[J]. *计算机应用与软件*, 2016, 33(4): 302-305.
- LU Jingqiao, FU Xiufen, MENG Zaiqiao. Research on robustness and centrality metrics of complex networks[J]. *Computer Applications and Software*, 2016, 33(4): 302-305.