

## 基于邻域信息的复杂网络节点重要性评估方法设计

李幸<sup>1</sup>, 詹杰<sup>1</sup>, 任保全<sup>2</sup>, 朱思奇<sup>1</sup>

(1. 湖南科技大学物理与电子科学学院, 湖南 湘潭 411201; 2. 中国电子设备系统工程公司研究所, 北京 100141)

**摘要:** 在复杂网络中准确识别影响力节点, 对网络管理和网络安全至关重要。局部中心性方法简明易用, 但忽略了邻居节点间的拓扑关系, 不能提供全局最优结果。提出了一种关联节点连边关系和拓扑结构的重要节点评估方法, 该方法首先综合应用 H 指数和信息熵对节点进行评估, 在此基础上代入节点的结构洞特征, 即在关注节点自身质量和邻居节点信息量的同时, 考虑了其“桥接”属性。采用疾病传播模型对算法进行验证, 应用肯德尔 (Kendall) 相关系数、互补累积分布函数以及传播影响力来验证算法的有效性与适用性。在 6 个真实网络数据集上的仿真结果表明, 在识别和排序网络中的关键节点上, 所提方法比传统中心性方法更准确。

**关键词:** 复杂网络; 节点重要性; SIR 模型; 信息熵; H 指数; 结构洞

中图分类号: O157.5

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2024.00335

## Design of nodes importance assessment method for complex network based on neighborhood information

LI Xing<sup>1</sup>, ZHAN Jie<sup>1</sup>, REN Baoquan<sup>2</sup>, ZHU Siqi<sup>1</sup>

1. School of Physics and Electronic Sciences, Hunan University of Science and Technology, Xiangtan 411201, China

2. China Electronics System Engineering Corporation Research Institute, Beijing 100141, China

**Abstract:** Accurate identification of influential nodes in complex networks is crucial for network management and network security. The local centrality method is concise and easy to use, but ignores the topological relationship between neighboring nodes and cannot provide globally optimal results. A node importance assessment method was proposed to correlate the node edge relationship and topology, which firstly applied the H-index and information entropy to assess the nodes, then combined the structural holes of the nodes with the structural characteristics of the nodes, and took into account the attribute of “bridging” while focusing on the node’s own quality and the amount of information about the neighboring nodes. The algorithm was validated by simulating the propagation process using the SIR model, and the Kendall correlation coefficient, complementary cumulative distribution function and propagation influence were applied to validate the validity and applicability of the method. Comparison of the experimental results on six real network datasets shows that the proposed method is more accurate than the traditional centrality methods in identifying and ordering the key nodes in the network.

**Key words:** complex network, node importance, SIR model, information entropy, H-index, structural hole

收稿日期: 2023-11-14; 修回日期: 2024-01-24

通信作者: 詹杰, jiezhawl@163.com

基金项目: 国家自然科学基金项目 (No.61875054); 移动专用网络国家工程研究中心项目 (No.BJTU20221102)

**Foundation Items:** The National Natural Science Foundation of China (No.61875054), Mobile Private Network National Engineering Research Center Project (No. BJTU20221102)

## 0 引言

随着网络信息技术在各行业的渗透, 复杂网络应用越来越广泛, 在复杂网络研究中, 如何准确地评估网络中节点重要性一直是网络研究的热点话题, 引起了众多研究者的关注<sup>[1-3]</sup>。大量与人类日常生活伴随的各种复杂系统都可以抽象成复杂网络, 如科学合作网络<sup>[4]</sup>、疾病传播网络<sup>[5]</sup>、信息传播网络<sup>[6]</sup>等, 这些网络都有评价和识别关键节点的需要, 因此有效识别复杂网络中有影响力的节点具有重要的理论和现实意义。

经典识别网络关键节点的算法有度中心性 (DC, degree centrality)<sup>[7]</sup>、介数中心性 (BC, betweenness centrality)<sup>[8]</sup>、接近度中心性 (CC, closeness centrality)<sup>[9]</sup>、K 壳分解法 (KS, k-shell)<sup>[10]</sup>和 H 指数 (H-index)<sup>[11]</sup>等。DC 简单方便, 但没有探讨节点在网络中的位置, 不能准确地评估节点影响力。BC 和 CC 需遍历整个网络才能得到结果, 计算复杂度高。KS 得出的排序结果区分度不高, 在某些网络 (如星型图、BA 人工网络) 中表现不佳。H 指数也因其粗粒度问题限制了应用范围。近年来, 针对经典算法的局限性有了大量的研究, 提出了一系列的改良方法<sup>[12-16]</sup>。

随着网络与人类社会信息交互程度的逐渐深入, 应用信息熵的概念识别网络中的关键节点成为新的研究方向, 出现了一系列的研究成果。Nie 等<sup>[17]</sup>考虑节点所有邻居的相关性, 提出映射熵 (ME, mapping entropy) 概念来识别节点重要性。Qiao 等<sup>[18]</sup>考虑了节点一跳邻居的直接和两跳邻居的间接影响, 提出了一种基于熵的子图度中心性 (SDC, subgraph degree centrality) 算法来评估网络节点的传播能力。Zareie 等<sup>[19]</sup>采用香农熵和 Jensen-Shannon 散度识别关键节点。卢鹏丽等<sup>[20]</sup>结合节点的介数中心性和度中心性, 提出介度熵概念来确定高影响力节点。Ezeh 等<sup>[21]</sup>利用 Dehmer 熵模型<sup>[22]</sup>量化最近邻节点的影响, 构造了子图度信息中心性 (SDI, subgraph degree information centrality)。

出于实现成本的考虑, 大量的算法都采用了局部中心性的设计思路, 只考虑了节点本身或邻居节点的相关属性, 而忽略了网络整体拓扑结构, 从而造成准确度不高。在上述研究工作的基础上, 我们提出了一种局部 H 指数熵中心性算法 (LHE, local

H-index entropy centrality), 该方法从节点邻域信息出发, 通过 H 指数和信息熵初步筛出网络中的重要节点, 然后应用结构洞特征代入网络拓扑结构信息, 提高算法的准确性。在 6 个真实网络数据集上应用 SIR 模型验证所提算法的有效性, 实验结果表明, 与一些中心性方法如 DC、BC、CC、H-index、ME、SDI 和 SDC 相比, 新的算法能更准确地识别关键节点。

## 1 相关工作

无向无权网络  $G = (V, E)$ ,  $V$  和  $E$  分别表示网络的节点集合和边集合。网络的邻接矩阵表示为  $A = (a_{ij})_{n \times n}$ , 若节点  $v_i$  连接到节点  $v_j$ , 则  $a_{ij} = 1$ , 否则  $a_{ij} = 0$ 。

### 1) 传统中心性

度中心性是最经典的中心性, 度数越大, 节点越重要, 计算式为

$$DC_i = \sum_{j=1}^N a_{ij} \quad (1)$$

其中,  $a_{ij}$  是网络所对应的邻接矩阵第  $i$  行第  $j$  列的元素值。

介数中心性认为如果一个节点出现在网络中所有节点对的最短路径上, 则该节点的重要性很高, 计算式为

$$BC_i = \sum_{s \neq u \neq i \in V} \frac{\delta_{su}(i)}{\delta_{su}} \quad (2)$$

其中,  $\delta_{su}$  表示网络中任意两个节点  $v_s$  和  $v_u$  之间的最短路径个数,  $\delta_{su}(i)$  表示的是节点  $v_s$  和  $v_u$  之间经过节点的最短路径数。

接近度中心性认为一个节点与网络其余节点的平均距离越小, 则该节点越重要。计算式为

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

其中,  $d_{ij}$  表示节点  $v_i$  到网络中任意一点  $v_j$  的最短路径长度。

### 2) H 指数

H 指数<sup>[23]</sup>最初用于科学计量学中评估研究人员的学术产出数量与学术产出水平。Lu 等<sup>[11]</sup>将其引用到复杂网络中, 揭示了度、H 指数和核数之间的内在关系, 指出在分析关键节点时, H 指数可视为核中心性与度中心性之间的良好平衡。若一个节点

的H指数为 $h$ ，说明该节点至少有 $h$ 个邻居节点，且它们的度都不小于 $h$ ，表示为

$$H_i = \zeta(k_{j_1}, k_{j_2}, \dots, k_{j_s}, \dots, k_{j_k}) \quad (4)$$

其中， $k_{j_s}$ 表示节点 $v_i$ 的第 $s$ 个邻居的度数。算子 $\zeta$ 返回最大整数 $h$ ，使得节点 $v_i$ 至少有 $h$ 个邻居的度数不少于 $h$ 。

### 3) 映射熵中心性

映射熵中心性<sup>[17]</sup>考虑节点与其邻居之间的映射关系，定义为

$$ME_i = -k_i \sum_{j=1}^N \log k_j \quad (5)$$

其中， $N$ 代表节点 $v_i$ 的邻居节点， $k_i$ 是 $v_i$ 的度， $k_j$ 是 $v_i$ 相邻节点 $v_j$ 的度。

### 4) 子图度信息中心性

子图度信息中心性<sup>[21]</sup>通过计算节点一跳邻域的度分布，并用Dehmer熵模型来量化节点的影响力。SDI模型如下：

$$SDI_i = -\sum_{i=1}^N P_i \lg P_i, P_i = \frac{\lambda_i}{\sum_{j \in \Gamma(i)} \lambda_j} \quad (6)$$

其中， $\lambda_i = k_i/N$ ， $v_j$ 为节点 $v_i$ 的一阶邻居， $k_i$ 是 $v_i$ 的度， $N$ 是节点总数。

### 5) 子图度中心性

子图度中心性<sup>[18]</sup>将节点影响力分解为直接影响和间接影响。直接影响 $LI_i$ 是指节点 $v_i$ 对一跳邻居的局部影响，间接影响 $II_i$ 是指节点 $v_i$ 对两跳内邻居的局部影响，表达式为

$$SDC_i = \omega_1 LI_i + \omega_2 II_i \quad (7)$$

根据文献可知， $\omega_1$ 取0.4， $\omega_2$ 取0.6。直接影响 $LI_i = -\sum_{i=1}^N P_i \lg P_i$ ， $P_i = k_i / \sum_{j=1}^N k_j$ ， $k_i$ 为节点 $v_i$ 的度。间接影响 $II_i = \sum_{k=1}^{M_i} II_{ik} / M_i$ ， $M_i$ 表示 $v_i$ 的两跳邻居节点数， $II_{ik} = \sum_{k=1}^{N_k} (LI_i \times LI_j / N_{ik})$ ， $N_{ik}$ 表示 $v_i$ 到 $v_k$ 的路径数。

### 6) 结构洞理论

结构洞<sup>[24]</sup>指网络结构中非冗余联系人之间存在的缺口，占据结构洞位置的个体可以带来累加而非重叠的竞争优势。Burt等<sup>[25]</sup>提出用网络约束系数来衡量节点形成结构洞所受到的约束，表示为

$$c_i = \sum_j \left( p_{ij} + \sum_{q \neq i, j} p_{iq} p_{qj} \right)^2, p_{ij} = \frac{z_{ij}}{\sum_{j \in \Gamma(i)} z_{ij}} \quad (8)$$

其中， $\Gamma(i)$ 表示节点 $v_i$ 的邻居集合，当 $v_i$ 与 $v_j$ 存在连边时， $z_{ij}$ 为1，反之 $z_{ij}$ 为0。节点 $q$ 表示 $i$ 和 $j$ 之间的共同邻居。节点的度越大， $p_{ij}$ 值越小，网络约束系数 $c_i$ 越小，说明节点与邻居之间的闭合程度低，这样的节点更容易形成结构洞，对于获取新的关系资源更有利，因此约束系数小的节点在信息传播中的影响力越大<sup>[26]</sup>。

### 7) 信息熵

“熵”通常用于描述复杂网络整体结构的复杂性以及相关统计特性。Shannon<sup>[27]</sup>提出用熵量化信息的不确定程度，熵越大，代表信息的不确定性越大。若 $X = \{x_1, x_2, \dots, x_n\}$ 是 $n$ 个可能事件的集合，对应发生概率 $P = \{p_1, p_2, \dots, p_n\}$ ， $0 \leq p_i \leq 1$ ， $\sum_{i=1}^n p_i = 1$ ，则 $X$ 的熵定义为

$$H(X) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (9)$$

Wang等<sup>[28]</sup>将熵的概念推广到复杂网络中，称为节点信息熵。节点信息熵考虑了邻居节点的传播效应，信息熵越大，其影响越容易传播到邻居节点，因此节点的影响越大。信息熵的定义为

$$E_i = -\sum_{j \in \Gamma(i)} I_j \cdot \ln I_j, I_i = \frac{k_i}{\sum_{j=1}^N k_j} \quad (10)$$

其中， $k_i$ 为节点 $v_i$ 的度。

## 2 局部H指数熵中心性算法

算法参考信息熵的定义，设置每个节点的信息熵为

$$E_i = -\sum_{j \in \Gamma(i)} I_j \cdot \ln I_j, I_i = \frac{k_i}{\sum_{j=1}^N k_j} \quad (11)$$

其中， $k_i$ 为节点 $v_i$ 的度。

使用单一因素判断节点重要性易造成结果过于粗糙的问题，且在大多数情况下，H指数比度或核数更能量化节点影响<sup>[11]</sup>，因此算法使用H指数对信息熵进行改进。由于信息熵和H指数具有不同的维度，两者不能直接融合，通过设置平衡因子 $\lambda$ 来融合两者的结构特征，平衡因子 $\lambda$ 定义为网络的平均H指数 $\langle h \rangle$ 与平均信息熵 $\langle e \rangle$ 的比值，定义为

$$\lambda = \frac{\left\langle \sum_{i=1}^N H_i \right\rangle}{\left\langle \sum_{i=1}^N E_i \right\rangle} = \frac{\langle h \rangle}{\langle e \rangle} \quad (12)$$

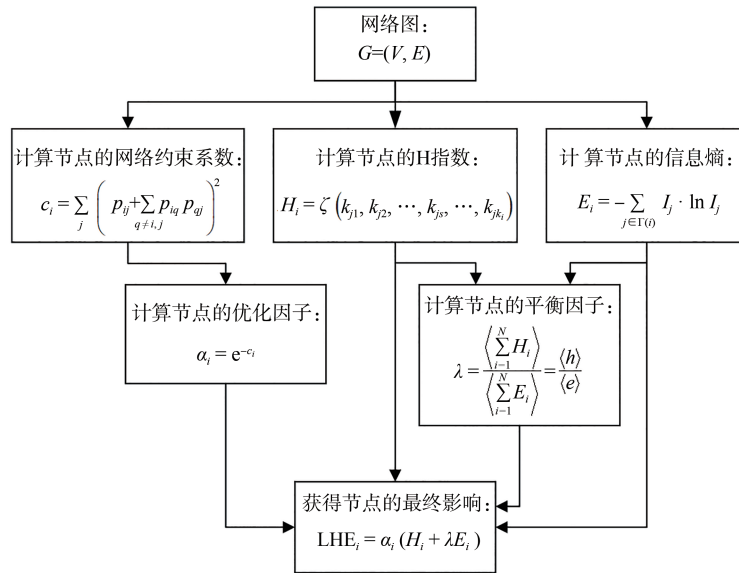


图1 LHE算法流程

H指数作为“优中选优”的方法，能挑选出网络中高质量的节点，但未考虑网络的结构信息。结构洞特征关注到邻域节点间的拓扑关系，因此算法使用结构洞的网络约束系数作为调整算法的优化因子，使计算节点影响力更具有效性。优化因子设置为 $\alpha_i = e^{-c_i}$ ， $c_i$ 代表节点 $v_i$ 形成结构洞的网络约束系数。 $c_i$ 值越小，优化因子 $\alpha_i$ 值越大，代表节点影响力越好， $\alpha_i$ 取值范围为 $(0, 1]$ 。

本方法采用H指数筛选高质量节点，将节点的邻域信息应用信息熵表示，拓扑结构用结构洞特征表达，以计算节点的最终影响力。局部H指数熵中心性LHE的定义为

$$LHE_i = \alpha_i (H_i + \lambda E_i) \quad (13)$$

LHE算法流程如图1所示。

图2对所提算法进行说明，示例网络为9个节点、10条边组成的网络。

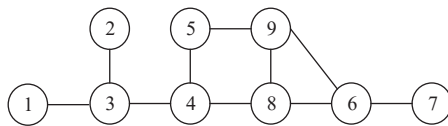


图2 示例网络

由图2可知，节点4的一阶邻居为节点3、节点5和节点8，它们的度分别为3、2和3，有两个度大于2的节点，因此节点4的H指数为2。节点4的度为3，节点4的重要度为 $I_4 = k_4 / \sum_{j=1}^N k_j = 3/20$ ，由此得到节点4的信息熵为 $E_4 = 0.799$ 。根据网络中

各节点的连边情况得出平衡因子 $\lambda = \langle h \rangle / \langle e \rangle = 2.897$ 。网络约束系数 $c_4 = 0.333$ ，优化因子 $\alpha_4 = e^{-0.333} = 0.716$ 。根据式(13)，可得节点4的值为 $LHE_4 = 3.092$ 。网络中其他节点的计算可参照上述过程。

表1为不同识别方法在示例网络中的排名结果，表中加粗字体为识别方法与SIR排序的吻合情况。从表1可以看出，DC和H-index均将示例网络中的节点分为极少的层数，意味着两种方法区分度低。BC、CC和SDC方法的分辨效果好，却无法正确识别关键节点的顺序。ME和SDI方法可以判断出前两个节点的重要性，但不能区分节点4和节点9的顺序。而所提的LHE方法在保证分辨率的前提下，还能精准识别排名前4的节点，同时，节点的排名顺序与SIR模型相差较小。综上所述，LHE比其他算法能更准确地识别高影响力节点。

### 3 仿真实验设置

实验选取了6个不同类型的真实网络数据集。Word<sup>[29]</sup>是查尔斯·狄更斯小说《大卫·科波菲尔》中常见形容词和名词的邻接网络。USAir<sup>[30]</sup>是美国的航空运输网络。Celegans<sup>[31]</sup>是秀丽隐杆线虫（代谢）网络。EEC<sup>[32]</sup>是一家大型欧洲研究机构成员之间的电子邮件交换网络。PB<sup>[33]</sup>是美国政治博客网络。WV<sup>[34]</sup>是一个描述了投票记录的维基百科网络。实验均使用各个真实网络的最大连通子图。表2为6个真实网络数据集的统计特征，包括网络的节点数 $n$ 、网络

表1 不同识别方法在示例网络中的排名结果

排名	1	2	3	4	5	6	7	8	9
SIR	8	4	9	6	3	5	7	1	2
DC	3, 4, 6, 8, 9	5	1, 2, 7	—	—	—	—	—	—
BC	4	3	8	6	9	5	1, 2, 7	—	—
CC	4	8	3, 5	6, 9	1, 2	7	—	—	—
H-index	8	4, 5, 6, 9	1, 2, 3, 7	—	—	—	—	—	—
ME	8	4, 9	6	5	3	1, 2, 7	—	—	—
SDI	8	4, 9	6	3	5	1, 2, 7	—	—	—
SDC	6	8	9	4	3	5	7	1, 2	—
LHE	8	4	9	6	5	3	1, 2, 7	—	—

表2 6个真实网络数据集的统计特征

网络	$n$	$m$	$c$	$d$	$\beta_{th}$
Word	112	425	0.173	2.472	0.078
USAir	332	2 126	0.749	2.738	0.023
Celegans	453	2 025	0.647	2.455	0.025
EEC	986	16 064	0.451	2.587	0.013
PB	1 222	16 714	0.360	2.738	0.012
WV	7 066	100 736	0.209	3.248	0.007

的连边数  $m$ 、网络的集聚系数  $c$ 、网络的平均距离  $d$ 、SIR 模型的传播阈值  $\beta_{th}$ ，实验均使用各个真实网络的最大连通子图。为验证所提算法的性能，LHE 与 DC、BC、CC、H-index、基于熵模型的 ME、SDI 和 SDC 这 7 种算法进行对比，分别从准确性、区分度和影响力 3 个方面来检验 LHE 方法的识别效果。

### 3.1 SIR 模型

为评估网络中各个节点的真实影响力，本文采用 SIR 疾病传播模型<sup>[35]</sup>。SIR 被广泛用于模拟信息传播和疾病感染的过程，模型中的节点有易感 (S)、感染 (I) 和恢复 (R) 3 种状态。若要获得节点  $v_i$  的影响力，首先将  $v_i$  设置为感染状态，其他节点设为易感状态。在接下来的每个步骤中，感染状态的节点以感染概率  $\beta$  感染其易感状态的邻居，且每个被感染的节点以  $\gamma$  的概率转为恢复状态，此后不再被感染。最后，当网络中没有处于感染状态的节点时，传播过程结束。传播结束时恢复状态的节点数作为估算最初感染节点  $v_i$  影响力的指标。SIR 模型中传播阈值  $\beta_{th} = \langle k \rangle / \langle k^2 \rangle$ ，其中， $\langle k \rangle$  和  $\langle k^2 \rangle$  分别代表网络的平均度和二阶平均度。如果  $\beta$  远小于  $\beta_{th}$ ，源节点的传播可能会限制在一个相对有限的范围内，如果  $\beta$  远大于  $\beta_{th}$ ，传播将会覆盖整个网络。为了减小数据误差，本文进行 1 000 次独立的模拟实

验，将恢复节点数的平均值作为节点的真实影响力。

### 3.2 肯德尔相关系数

为评估识别方法的准确性，本文采用肯德尔 (Kendall) 相关系数<sup>[36]</sup>测量不同识别方法得到的节点影响力排名与 SIR 模型获得的节点传播能力排名之间的相关性，表达式为

$$\tau(X, Y) = \frac{2(n_c - n_d)}{n(n-1)} \quad (14)$$

其中， $X$  和  $Y$  代表不同识别方法的排序列表， $n$  代表  $X$  或  $Y$  的排序结果数量， $n_c$  代表计算后相关性一致的数量， $n_d$  代表计算后相关性不一致的数量。通过相关系数可以较好地反映节点重要性评估方法的优劣，相关系数  $\tau$  越大，代表算法和真实传播影响力越接近。

### 3.3 互补累积分布函数

区分具有不同传播能力节点的能力是度量复杂网络中节点重要性评估方法性能的重要指标之一。为量化不同方法的区分度，本文使用互补累积分布函数 (CCDF, comprehensive cumulative distribution function)<sup>[37]</sup> 作为度量指标。CCDF 可以充分描述每种方法获得值的概率分布，表达式为

$$CCDF(r) = 1 - \frac{\sum_{i=1}^r n_i}{n} \quad (15)$$

其中， $n$  是网络的总节点数， $\sum_{i=1}^r n_i$  表示数值排名的数量。如果排名列表中的排名越接近  $n$ ，即节点在同一排名中的累积较少，则区分效果越好。

## 4 仿真实验与结果分析

本节从准确性、区分度、影响力和时间复杂度 4 个方面对识别方法进行性能分析。准确性用于验

表3 不同识别方法与SIR模型模拟得出的肯德尔相关系数 $\tau$ 值

网络	DC	BC	CC	H-index	ME	SDI	SDC	LHE
Word	0.868 0	0.657 6	0.842 8	0.869 6	0.894 5	0.790 7	0.882 1	<b>0.904 9</b>
USAir	0.764 5	0.573 0	0.792 5	0.789 6	0.803 3	0.715 9	0.800 2	<b>0.834 8</b>
Celegans	0.688 8	0.508 7	0.649 0	0.703 9	0.737 7	0.500 7	0.746 2	<b>0.797 7</b>
EEC	0.857 2	0.727 6	0.850 8	0.813 4	0.866 3	0.839 5	0.868 9	<b>0.894 9</b>
PB	0.864 1	0.687 6	0.777 3	0.882 2	0.873 0	0.849 4	0.887 7	<b>0.892 5</b>
WV	0.807 4	0.740 8	0.799 5	0.812 6	0.803 7	0.714 9	0.810 3	<b>0.813 1</b>

证排序结果的合理性，区分度旨在避免大量节点被定义为相同值，影响力用于检验排序节点的传播能力相关性，时间复杂度为不同算法的运算时间。

#### 4.1 准确性分析

本实验使用肯德尔相关系数验证准确性。为了更接近真实疾病的传播过程，探索感染率变化对算法排名准确性的影响。表3为感染概率 $\beta = \beta_{th}$ 时，不同识别方法与SIR模型模拟得出的肯德尔相关系数 $\tau$ 值，表中加粗字体对应网络中的最优值。从表3可看出，LHE算法的准确性优于其他方法。

为了进一步说明准确性，本实验将感染区间设置为 $[0.5\beta_{th}, 1.5\beta_{th}]$ ，在区间内取11个均匀间隔的感染概率 $\beta$ ，得到相应的相关系数 $\tau$ 。实验结果如图3所示，灰色虚线表示网络的传播阈值 $\beta_{th}^{[38]}$ 。由图可知，尽管全局指标BC和CC耗费时间长，但并不能很好地衡量节点在网络中的影响。在WV网络中，当 $\beta < \beta_{th}$ 时，LHE方法不及H-index和DC，这是因为在 $\beta$ 较小时，传播通常局限于初始感染节点的周围，因此度大的节点具有更强的传播能力。当 $\beta$ 增大时，LHE开始展示出更好的性能。整体来看，LHE的 $\tau$ 值最高，说明LHE识别的节点与节点的真实传播能力之间有显著相关性。

#### 4.2 区分度分析

一个优秀的识别方法应该具有高分辨率，本节通过CCDF检验不同排名累积的节点。图4展示了CCDF曲线，通过观察曲线的下降趋势，可以直观地了解每种方法的区分度。DC和H-index的曲线近似直线下降，说明这两种方法定义了很多相同值的节点。代表BC和CC方法的曲线下降速度低于上述3种方法。特别在WV网络中，CC在前半段的区分能力最出色，这是因为全局指标在区分度上能够比局部指标更有效，但随着节点数量增加，CC在后期的识别上乏力。相较传统的中心性，BC在

Word网络中表现良好，该方法旨在找到位于网络枢纽位置的节点，因此BC在节点数量小的网络中展示出较好的区分能力。在EEC上，LHE识别能力与ME一致。在Word、PB和Celegans网络中，LHE明显表现出比其他方法更慢的下降速度，表示分配给同一排名的节点数最少。由上述分析可得，LHE方法具有优秀的区分性能。

#### 4.3 影响力分析

本节进一步分析不同评估方法的节点中心性度值与SIR模型传播能力之间的相关性。为了更直观地说明结果，选取了Celegans、EEC和WV这3个网络，且着重对比ME、SDI、SDC和LHE这4个与信息熵相关的指标。实验设置感染概率为 $\beta = \beta_{th}$ ，时间步长 $t$ 为10， $F(t)$ 表示SIR模型计算得出的节点传播能力，并将实验重复100次以减小误差。

图5显示了3个真实网络上采用4种不同评估方法获得的节点影响值与SIR模型仿真获得的值之间影响一致性的散点图，图中的每个散点表示一个节点。散点图所示曲线趋势的单调性越好，散点越近，相关性越高，代表传播能力与真实水平越接近。

在3个网络中，ME、SDI、SDC和LHE均与节点影响 $F(t)$ 呈强正相关。但与LHE相比，ME、SDI和SDC图中的点相对离散，图中左上角计算得出的相关系数(corre)也证明了上述推论。相对而言，所提出的LHE方法得到的节点度值与节点影响的相关性最高，表明该方法能更好地检测出网络中具有更强传播能力的节点。

#### 4.4 复杂度分析

不同识别算法的时间复杂度见表4。从表4中可以看出，DC的时间复杂度最低，BC和CC具有最高的时间复杂度，尽管LHE的时间复杂度并非最低，但其准确性显著高于其他方法，节点包含的网络信息也高于其他方法。因此，LHE方法适用于大型网络。

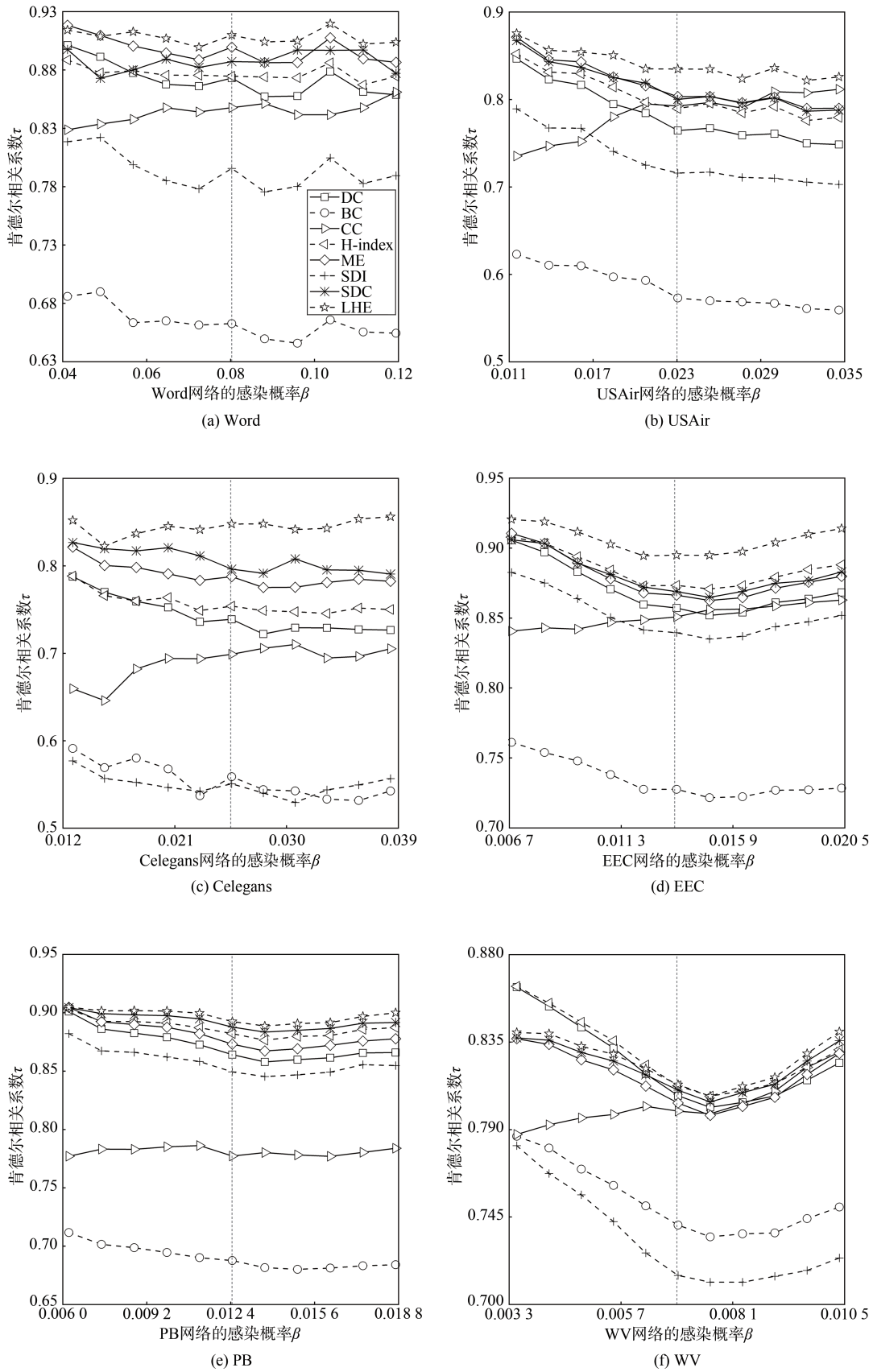


图3 6个真实网络上8种不同识别方法的肯德尔(Kendall)相关系数 $\tau$ 变化趋势(结果为SIR模型1000次独立实验后的平均值,横坐标代表不同的感染概率,纵坐标代表各算法得出的肯德尔相关系数,垂直虚线代表传播阈值)

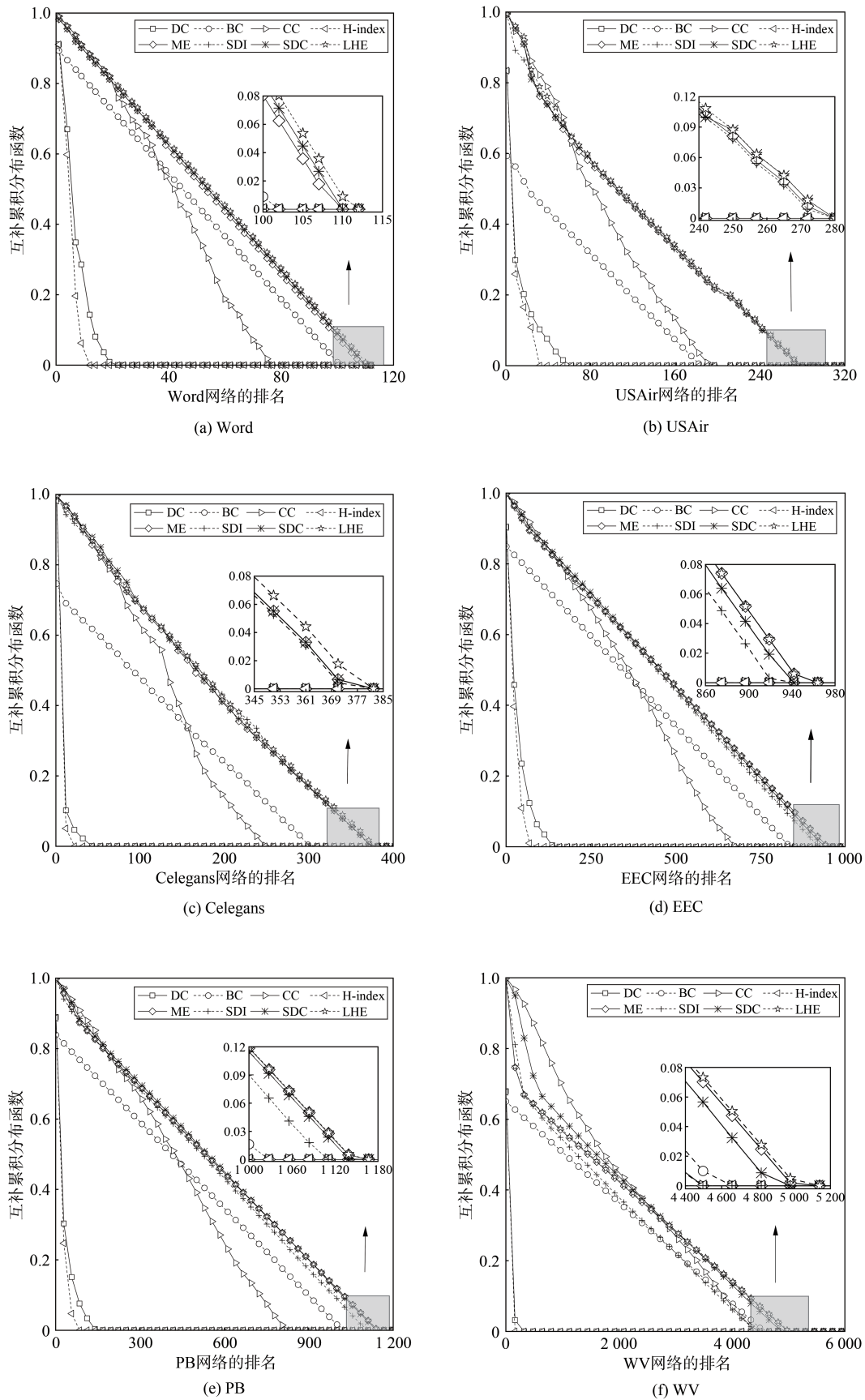


图4 6个真实网络上8种不同评估方法的排序结果互补累积分布函数(横坐标代表各算法得出的排名数量,纵坐标代表算法得出的CCDF值)

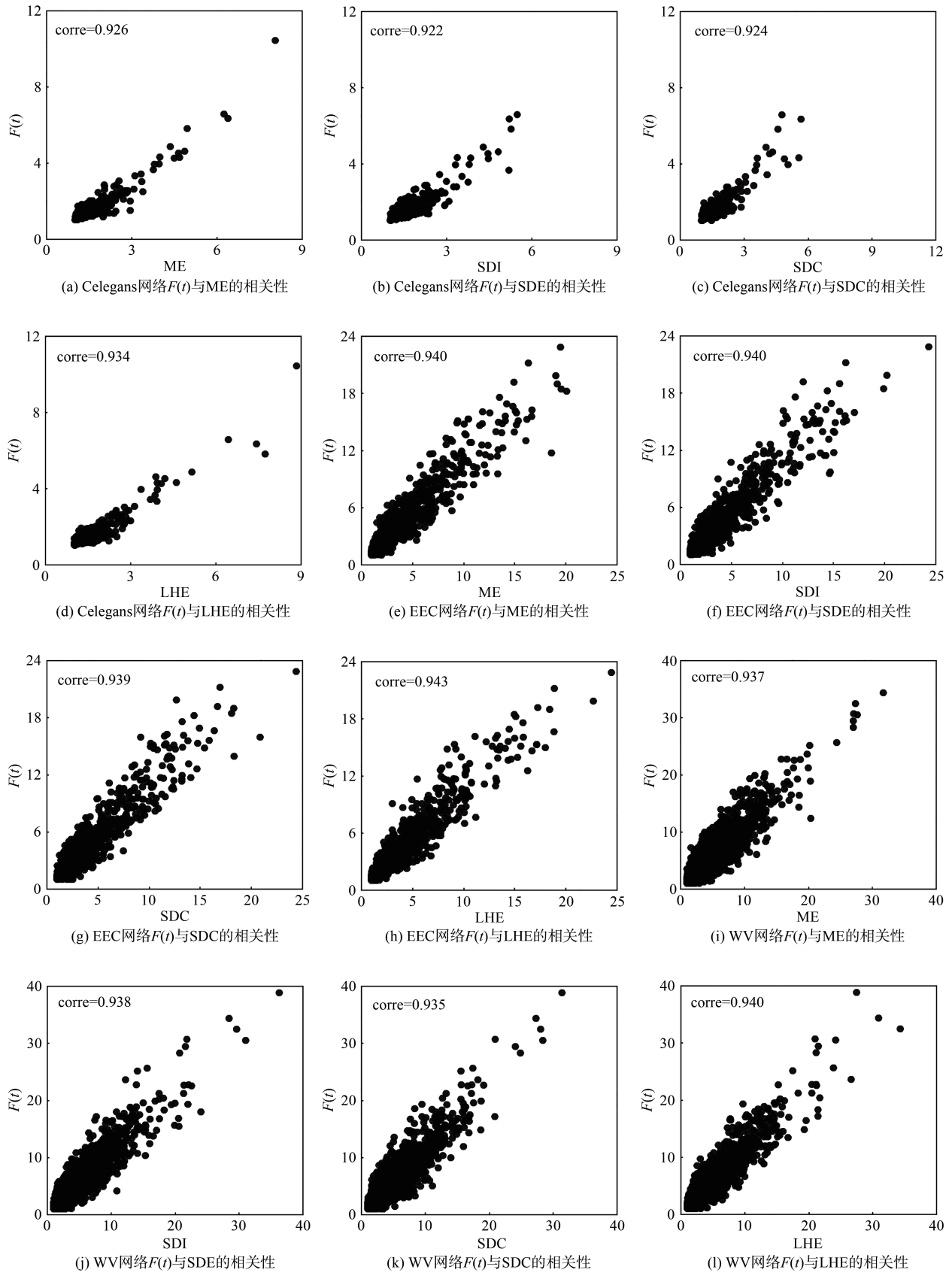


图5 3个真实网络上采用4种不同评估方法与其传播能力的相关性(横轴对应于评估方法节点的中心度值,纵轴表示传播能力 $F(t)(t=10)$ ,结果由SIR模型的100个独立实验进行平均得到)

表 4 不同识别算法的时间复杂度

算法	时间复杂度
DC	$O(n)$
BC	$O(nm)$
CC	$O(nm)$
Hindex	$O(n \log n)$
ME	$O(n \log n)$
SDI	$O(n \log n)$
SDC	$O(n \log n)$
LHE	$O(n^2)$

### 5 结束语

复杂网络中高影响力节点的识别问题对网络性能和安全的研究至关重要。本文提出了一种基于节点邻域信息的局部中心性算法 LHE，该方法通过 H 指数和信息熵识别网络的关键节点，并利用结构洞特征度量拓扑结构。在 6 个真实网络上的对比实验表明，与已知的度量方法如 DC、BC、CC、H-index、ME、SDI 和 SDC 相比，LHE 方法在识别关键节点方面表现突出，在不同规模的网络中都能很好地区分节点间的差异。虽然仿真环境为算法在无向无权网络的静态测试，与真实世界中复杂有向加权网络的大小、密度和动力特性有一些区别，但本方法对研究有向网络、加权网络、多级网络和时态网络中关键节点的影响力排序有很好的借鉴作用。

#### 参考文献:

[1] GAO J, ZHANG Y C, ZHOU T. Computational socioeconomics [J]. *Physics Reports*, 2019, 817: 1-104.

[2] LIU Y Y, SLOTINE J J, BARABÁSI A L. Controllability of complex networks[J]. *Nature*, 2011, 473: 167-173.

[3] VERMEULEN R, SCHYMANSKI E L, BARABÁSI A L, et al. The exposome and health: where chemistry meets biology[J]. *Science*, 2020, 367(6476): 392-396.

[4] JUNG H, PHOA F K H. A mixture model of truncated zeta distributions with applications to scientific collaboration networks[J]. *Entropy*, 2021, 23(5): 502.

[5] BARABÁSI A L, GULBAHCE N, LOSCALZO J. Network medicine: a network-based approach to human disease[J]. *Nature Reviews Genetics*, 2011, 12(1): 56-68.

[6] BORGE-HOLTHOFER J, MORENO Y. Absence of influential spreaders in rumor dynamics[J]. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 2012, 85(2): 026116.

[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] FREEMAN L C. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977, 40(1): 35.

[9] SABIDUSSI G. The centrality of a graph[J]. *Psychometrika*, 1966, 31(4): 581-603.

[10] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. *Nature Physics*, 2010, 6: 888-893.

[11] LU L Y, ZHOU T, ZHANG Q M, et al. The H-index of a network node and its relation to degree and coreness[J]. *Nature Communications*, 2016, 7: 10168.

[12] CHEN D B, LÜ L Y, SHANG M S, et al. Identifying influential nodes in complex networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2012, 391(4): 1777-1787.

[13] REN Z M, SHAO F, LIU J G, et al. Node importance measurement based on the degree and clustering coefficient information[J]. *Acta Physica Sinica*, 2013, 62(12): 128901.

[14] LIU Y, WEI B, DU Y X, et al. Identifying influential spreaders by weight degree centrality in complex networks[J]. *Chaos, Solitons & Fractals*, 2016, 86: 1-7.

[15] BAE J, KIM S. Identifying and ranking influential spreaders in complex networks by neighborhood coreness[J]. *Physica A: Statistical Mechanics and Its Applications*, 2014, 395: 549-559.

[16] XU G Q, MENG L, TU D Q, et al. LCH: a local clustering H-index centrality measure for identifying and ranking influential nodes in complex networks[J]. *Chinese Physics B*, 2021, 30(8): 088901.

[17] NIE T Y, GUO Z, ZHAO K, et al. Using mapping entropy to identify node centrality in complex networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2016, 453: 290-297.

[18] QIAO T, SHAN W, ZHOU C. How to identify the most powerful node in complex networks? A novel entropy centrality approach[J]. *Entropy*, 2017, 19(11): 614.

[19] ZAREIE A, SHEIKHAHMADI A, JALILI M. Influential node ranking in social networks based on neighborhood diversity[J]. *Future Generation Computer Systems*, 2019, 94: 120-129.

[20] 卢鹏丽, 郭旭东, 董瑞, 等. 基于介度熵的复杂网络节点重要度识别方法[J]. *兰州理工大学学报*, 2020, 46(2): 111-115.

LU P L, GUO X D, DONG M, et al. Importance identification method of complex network nodes based on betweenness and degree entropy[J]. *Journal of Lanzhou University of Technology*, 2020, 46(2): 111-115.

[21] EZEHC, REN T, XU Y J, et al. Entropy and structural-hole based node ranking methods [J]. *J Internet Technol*, 2021, 22(5): 1009-1017.

[22] DEHMER M. Information processing in complex networks: graph entropy and information functionals[J]. *Applied Mathematics and Computation*, 2008, 201(1/2): 82-94.

[23] HIRSCH J E. An index to quantify an individual's scientific re-

- search output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46): 16569-16572.
- [24] SALANCIK G R, BURT R S. WANTED: a good network theory of organization[J]. Administrative Science Quarterly, 1995, 40(2): 345.
- [25] BURT R S, KILDUFF M, TASSELLI S. Social network analysis: foundations and frontiers on advantage[J]. Annual Review of Psychology, 2013, 64: 527-547.
- [26] SU X P, SONG Y R. Leveraging neighborhood “structural holes” to identifying key spreaders in social networks[J]. Acta Physica Sinica, 2015, 64(2): 020101.
- [27] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.
- [28] WANG M, LI W C, GUO Y N, et al. Identifying influential spreaders in complex networks based on improved k-shell method[J]. Physica A: Statistical Mechanics and Its Applications, 2020, 554: 124229.
- [29] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2006, 74(3 Pt 2): 036104.
- [30] COLIZZA V, PASTOR-SATORRAS R, VESPIGNANI A. Reaction-diffusion processes and metapopulation models in heterogeneous networks[J]. Nature Physics, 2007, 3: 276-282.
- [31] DUCH J, ARENAS A. Community detection in complex networks using extremal optimization[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2005, 72(2 Pt 2): 027104.
- [32] YIN H, BENSON A R, LESKOVEC J, et al. Local higher-order graph clustering[J]. KDD: Proceedings International Conference on Knowledge Discovery & Data Mining, 2017, 2017: 555-564.
- [33] ADAMIC L A, GLANCE N. The political blogosphere and the 2004 U.S. election: divided they blog[C]//Proceedings of the 3rd international workshop on Link discovery. New York: ACM, 2005: 36-43.
- [34] LESKOVEC J, HUTTENLOCHER D, KLEINBERG J. Predicting positive and negative links in online social networks[C]//Proceedings of the 19th International Conference on World Wide Web. New York: ACM, 2010: 641-650.
- [35] SHARKEY K J. Deterministic epidemic models on contact networks: correlations and unbiological terms[J]. Theoretical Population Biology, 2011, 79(4): 115-129.
- [36] KENDALL M G. A new measure of rank correlation[J]. Biometrika, 1938, 30(1/2): 81-93.
- [37] HELTON J C. Probability, conditional probability and complemen-

tary cumulative distribution functions in performance assessment for radioactive waste disposal[J]. Reliability Engineering & System Safety, 1996, 54(2/3): 145-163.

- [38] CASTELLANO C, PASTOR-SATORRAS R. Thresholds for epidemic spreading in networks[J]. Physical Review Letters, 2010, 105(21): 218701.

#### [作者简介]



李幸(1999- ), 女, 湖南科技大学物理与电子科学学院硕士生, 主要研究方向为复杂网络、节点影响力评估。



詹杰(1973- ), 男, 湖南科技大学物理与电子科学学院教授、博士生导师, 主要研究方向为复杂网络评估、近距离无线通信技术、物联网。



任保全(1974- ), 男, 博士, 中国电子设备系统工程公司研究所高级工程师, 主要研究方向为复杂环境下物联网技术及应用。



朱思奇(1997- ), 男, 湖南科技大学物理与电子科学学院硕士生, 主要研究方向为复杂网络、复杂系统控制理论。