

## 基于全局—局部属性的复杂网络节点综合影响力评估算法

蒋伟进<sup>1,2</sup>, 杨莹<sup>1,2</sup>, 罗田甜<sup>2,3</sup>, 周文颖<sup>1,2</sup>, 李恩<sup>1,2</sup>, 张小威<sup>1,2</sup>

(1. 湖南工商大学计算机学院, 湖南 长沙 410205;

2. 数据智能与智慧社会国家重点实验室(培育)基地, 湖南 长沙 410205;

3. 湖南工商大学前沿交叉学院, 湖南 长沙 410205)

**摘要:** 挖掘网络中的关键节点在信息传播、病毒营销、舆论控制等的演进过程中发挥着巨大的作用, 关键节点的识别可以有效地帮助控制网络攻击、检测金融风险、抑制病毒和谣言的传播、防止恐怖袭击等。为了突破现有节点影响力评估方法存在的算法复杂度高、准确度低以及评价指标内在作用机制评估角度片面的限制, 提出了一种识别关键节点的综合影响力(CI, comprehensive influence)评估算法。该算法通过同时处理网络的局部和全局拓扑来对节点重要性进行排序, 从多个角度整合网络属性信息, 提供更全面的节点重要性度量。算法中的全局属性考虑的是邻居节点以及节点之间的最短距离, 节点的信息熵用来表示节点的局部属性, 通过一个参数来调整全局和局部属性的权重比。使用 SIR (susceptible infected recovered) 模型和 Kendall 相关系数作为评价标准, 在不同规模的现实世界网络上进行实验分析, 结果表明, 所提出的方法能在识别关键节点方面优于介数中心性(BC, betweenness centrality)、接近中心性(CC, closeness centrality)、重力指数中心性(GIC, gravity index centrality)、全局结构模型(GSM, global structure model)等著名的启发式算法, 并且具有更好的排序单调性、更稳定的度量结果, 对网络拓扑的适应性更强, 适用于绝大多数具有不同结构的真实网络。

**关键词:** 关键节点; 复杂网络; 节点信息熵; 多属性综合评估

**中图分类号:** TP301.6

**文献标志码:** A

**doi:** 10.11959/j.issn.2096-3750.2022.00282

## Comprehensive influence evaluation algorithm of complex network nodes based on global-local attributes

JIANG Weijin<sup>1,2</sup>, YANG Ying<sup>1,2</sup>, LUO Tiantian<sup>2,3</sup>, ZHOU Wenyong<sup>1,2</sup>, LI En<sup>1,2</sup>, ZHANG Xiaowei<sup>1,2</sup>

1. School of Computer Science, Hunan University of Technology and Business, Changsha 410205, China

2. State Key Laboratory of Digital Intelligence and Smart Society (Cultivating) Base, Changsha 410205, China

3. School of Advanced Interdisciplinary Studies, Hunan University of Technology and Business, Changsha 410205, China

**Abstract:** Mining key nodes in the network plays a great role in the evolution of information dissemination, virus marketing, and public opinion control, etc. The identification of key nodes can effectively help to control network attacks, detect financial risks, suppress the spread of viruses diseases and rumors, and prevent terrorist attacks. In order to break through the limitations of existing node influence assessment methods with high algorithmic complexity and low accuracy, as well as one-sided perspective of assessing the intrinsic action mechanism of evaluation metrics, a comprehensive influence (CI) assessment algorithm for identifying critical nodes was proposed, which simultaneously processes the local and global topology of the network to perform node importance. The global attributes in the algorithm consider the in-

收稿日期: 2022-04-22; 修回日期: 2022-07-03

通信作者: 杨莹, 1871529103@qq.com

基金项目: 国家自然科学基金资助项目(No.61772196); 湖南省自然科学基金资助项目(No.2020JJ4249); 湖南省教育厅科学研究重点项目(No.21A0374); 湖南省研究生科研创新项目(No.CX20211108, No.CX20211151)

**Foundation Items:** The National Natural Science Foundation of China (No.61772196), The Natural Science Foundation of Hunan Province (No.2020JJ4249), The Key Scientific Research Project of Hunan Provincial Department of Education (No.21A0374), The Hunan Provincial Innovation Foundation for Postgraduate (No.CX20211108, No.CX20211151)

formation entropy of neighboring nodes and the shortest distance nodes between nodes to represent the local attributes of nodes, and the weight ratio of global and local attributes was adjusted by a parameter. By using the SIR (susceptible-infected-recovered) model and Kendall correlation coefficient as evaluation criteria, experimental analysis on real-world networks of different scales shows that the proposed method is superior to some well-known heuristic algorithms such as betweenness centrality (BC), closeness centrality (CC), gravity index centrality (GIC), and global structure model (GSM), and has better ranking monotonicity, more stable metric results, more adaptable to network topologies, and is applicable to most of the real networks with different structure of real networks.

**Key words:** node importance, complex networks, node information entropy, integrated multi-attribute evaluation

## 0 引言

复杂网络的安全性关系到经济发展和社会稳定,它的相关特性在电力网络<sup>[1]</sup>、交通规划<sup>[2]</sup>、社会科学<sup>[3]</sup>、时间序列预测<sup>[4]</sup>、生物系统等领域受到研究人员的广泛关注<sup>[5]</sup>。通常网络中的关键节点数量少,但信息传播的速度很快,一旦关键节点出现问题,整个网络系统都会瞬间受到影响<sup>[6]</sup>。因此,评估网络节点的重要性具有重要意义。例如,在犯罪网络中快速识别恐怖组织的头目<sup>[7]</sup>;在计算机网络中防御恶意网络攻击<sup>[8]</sup>;在交通网络中有效解决交通堵塞问题<sup>[9]</sup>等。节点重要性根据网络功能不同有不同的度量方法。例如,在网络攻击中,删除某个节点对网络结构的破坏程度越高,则此节点越重要;信息传播中,如果某个节点作为源节点使得信息传播范围越广或传播的速度越快,则此节点越重要。面对规模庞大的复杂网络,评估节点重要性的首要问题是快速、有效地找出网络中的关键节点。

目前,节点信息熵在许多方面的应用都取得了不错的效果。例如,信息熵应用于供应链财务风险评估,通过使用信息熵可以认识所处供应链的稳定性,吸收财务风险较低的企业加入,及时发现财务风险较高的企业并予以替换,防止供应链任一环节出现问题。它也在环境管理过程控制、水质水环境的预测预报评估等过程中发挥着重要作用。现有的关键节点挖掘算法主要是基于网络的结构特征设计的,随着基于节点信息熵的关键节点挖掘算法研究的深入,越来越多的关键节点识别算法被提出,例如,Guo等<sup>[10]</sup>提出了基于信息熵的EnRenew方法来识别关键节点的集合。Xu等<sup>[11]</sup>引入了邻接信息熵,通过整个图中的边权重和方向来识别重要节点。Maji等<sup>[12]</sup>提出了一种结合节点度、紧密度、 $k$ 壳中心性等指标的方法,实现对复杂网络中的重要节点进行排序。

但是,现有的关键节点挖掘算法通常还存在以下两个问题:节点重要性根据网络功能不同有不同

的标准,不存在一种通用指标来量化不同网络中节点的重要性;基于全局属性或可调参数的算法复杂度高,基于局部属性或无参数的算法准确度低,现有方法的效率和准确度有待进一步提高。

针对以上问题,本文提出了一种识别关键节点的综合影响力评估方法,主要贡献如下。

1) 设计了一种基于复杂网络局部和全局拓扑的节点综合影响力(CI, comprehensive influence)评估算法来量化节点的重要性,该算法同时考虑了网络的局部和全局拓扑,它使用节点之间的度数、节点信息熵、可调参数和最短路径等指标,通过引入一个调节局部和全局属性权重的参数识别网络中的重要节点。

2) 将CI评估算法与介数中心性(BC, betweenness centrality)、接近中心性(CC, closeness centrality)、重力指数中心性(GIC, gravity index centrality)、全局结构模型(GSM, global structure model)方法进行比较,实验结果表明,CI评估算法可以有效识别网络中的关键节点,并且具有更好的排序单调性和更稳定的度量结果,对网络拓扑的适应性更强,适用于绝大多数具有不同结构的真实网络。此外,CI评估算法的时间复杂度仅为 $O(n^2)$ ,并且其准确度高于BC、CC、GIC、GSM算法。

## 1 相关工作

目前,研究者根据具体应用场景,从不同角度设计了多种多样的算法来识别网络中节点的重要性<sup>[13-16]</sup>。总的来说,现有网络重要节点识别算法本质上都是基于网络结构特征设计的,网络中节点重要性的识别方法可以分为基于领域的中心性方法、基于路径的中心性方法以及基于迭代的中心性方法。

### 1) 基于领域的中心性方法

度中心性(DC, degree centrality)<sup>[17]</sup>是一种简单而典型的方法,但它没有考虑复杂网络的全局结构,使得一些看起来不重要但实际很重要的节点被忽略了。例如,删除度小的节点导致连通图断开,代表

这个节点处于重要的位置。LocalRank 算法<sup>[18]</sup>考虑了节点的4阶邻居中包含的信息。ClusterRank 算法<sup>[19]</sup>考虑的是1阶邻居的数量和节点的聚类系数。通常在1阶邻居数量相等的条件下，节点的聚类系数越大就说明节点在网络中的影响力越小。Kitsak 等<sup>[20]</sup>提出了一种称为“K-Shell”的方法，该方法考虑了网络的全局层次结构，但大量节点分布在同层，无法有效衡量同层节点的重要性。同时，由于评价视角不同，不同的中心性可能会产生不同的度量结果甚至冲突，给节点排序决策带来困难。Malang 等<sup>[21]</sup>通过比较6种重要节点识别算法的性能发现，不仅节点本身很重要，而且由有影响力的节点进行组合的一组节点同样对网络结构具有重要意义。上述算法本质上是基于邻域的，而节点的传播信息的速度和范围应该受到传播路径的影响。

### 2) 基于路径的中心性方法

接近中心性 (CC)<sup>[22]</sup>算法考虑网络每个节点之间的距离。一个节点与其他节点距离的平均值越小，节点在网络中的位置就越关键，但这种方法无法处理带有非连通图的网络。介数中心性 (BC)<sup>[23]</sup>算法可以找出网络中信息传播时路过最多的那个节点，但 BC 在节点不属于其他节点对的最短路径时会出现其值为 0 的情况。Kumar<sup>[24]</sup>考虑网络中所有的路径，路径越长，分配的权重越小。Yang 等<sup>[25]</sup>提出了一种考虑了最短路径、最短路径数和非最短路径数 3 个方面的基于全局结构的算法。Ibnoulouafi 等<sup>[26]</sup>提出了一种考虑最短路径信息和最近邻域的本地信息相结合的多属性中心性度量算法，对特定网络中的关键节点进行分类的基于重力公式的算法。此外，Li 等<sup>[27]</sup>提出了一种改进的重力模型识别传播动力学中的重要节点，并提出了一种基于节点自身及其邻域的算法。Zhuang 等<sup>[28]</sup>提出了基于结构、基于信息和基于动作的多维社会影响 (MSI, multidimensional social influence) 测量方法。

### 3) 基于迭代的中心性方法

以前的大多数方法都假设节点的影响取决于节点本身的重要性，但还有一个关键因素，即邻居的重要性，称为相互增强效应<sup>[29]</sup>。基于迭代中心性的算法，其每个节点都得到其邻居的支持。特征向量中心性 (eigenvector centrality)<sup>[30]</sup>和累积提名 (cumulative nomination)<sup>[31]</sup>算法应用于无向网络，PageRank<sup>[32]</sup>和 HITS<sup>[33]</sup>算法应用于有向网络。PageRank 最初用于对网页进行排序，为了解决悬空节点问题，PageRank

引入了一个可调参数，LeaderRank<sup>[34]</sup>假设相邻节点之间随机游走概率相同，通过令网络强连接来消除所有悬空节点，但它的假设显然是不合理的。HITS 算法从权威和枢纽的角度来评估节点的重要性。

综上所述，虽然识别关键节点算法已经取得了一定进展，但仍然存在评估角度片面、会改变网络结构从而丢失节点间的关系信息等问题。另外，大多数方法使用原始的网络结构，没有将网络的局部和全局属性进行融合，从而得到节点的综合影响力。

## 2 节点综合影响力评估算法

CI 评估算法通过同时处理网络的局部和全局拓扑对节点重要性进行排序，从多个角度整合网络属性信息，计算节点的综合影响力，更全面地度量节点重要性。算法中的全局属性考虑的是邻居节点以及节点之间的最短距离，局部属性考虑的是节点信息熵，并通过设置一个参数来调整全局和局部属性的权重比。

### 2.1 节点信息熵

目前，大部分算法都基于节点度这个指标评估节点自身的局部影响力，但是基于度设计算法只包含了节点的1阶邻居信息，并不能区分邻居节点的重要性。邻居度分布不均匀示例如图 1 所示，图 1 (a) 中节点 1 和图 1 (b) 中节点 2 的度相同，但是进一步观察节点 1 和节点 2 的邻居节点，可以看出节点 1 所处的位置比节点 2 更重要。因此，需要用更高阶的评估指标挖掘网络中的关键节点。

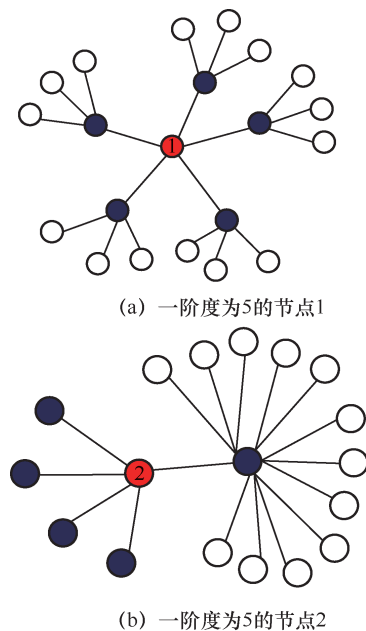


图1 邻居度分布不均匀示例

信息量在信息论中衡量的是特定事件所带来的信息，信息熵是信息量的期望值。通过将信息熵和信息量引入复杂网络中来计算节点的重要性，相较于仅使用节点的一阶信息度而言，节点信息熵考虑了更高阶的信息。一个随机变量  $X$  可能的取值为  $x_i$ ，每个  $x_i$  发生的概率为  $p_i$  ( $i = 1, \dots, n$ )，其中  $0 \leq p_i \leq 1$ ，那么  $X$  的信息熵  $H$  定义为

$$H(x) = -\sum_{i=0}^n p_i \log_b p_i \quad (1)$$

通常情况  $b = 2$ 、 $b = 10$  或  $b = e$ ，可以通过使用非负整数组  $(\lambda_1, \lambda_2, \dots, \lambda_n)$  [35] 形成概率分布  $p = (p_1, p_2, \dots, p_n)$ ，其中  $\lambda_i$  表示第  $i$  个非负整数。

$$p_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}, \quad i = 1, 2, \dots, n \quad (2)$$

那么节点  $i$  的信息熵定义为[35-36]

$$h_i = \sum_{j \in \Gamma_i} -p_i \log b(p_i) \quad (3)$$

其中， $p_i = \frac{k_j}{\sum_{l \in \Gamma_i} k_l}$ ， $\Gamma_i$  表示节点  $i$  的邻居节点集合， $k_j$  表示节点  $j$  的度。

从图 1 可以看出，节点 1 和节点 2 的度相等，即  $k_1 = k_2 = 5$ ，但就信息传播的角度而言，显然节点 1 的位置在网络中更关键，这也说明了仅仅使用节点的一阶度信息的不足，而通过式(3)可以得到节点 1 的信息熵大于节点 2，即  $h_1 = 1.61 > h_2 = 0.81$ ，因此，节点信息熵能更准确地度量节点的重要性。

## 2.2 CI 评估算法

**定义 1** 局部影响力：节点  $v_i$  的信息熵表示图  $G$  的局部信息，定义为

$$LI_{(v_i)} = \sum_{j \in \Gamma_i} -p_i \log b(p_i) \quad (4)$$

**定义 2** 全局影响力：节点不仅影响自身，还考虑相邻节点的影响。实际上度大的节点更可能是重要的，但与此同时也需要考虑任意两个节点之间的距离，它与该节点的影响成反比。所以本文使用 Dijkstra[37] 算法计算邻居节点  $j$  和节点之间的最短距离。节点的全局影响力  $GI_{(v_i)}$  可以定义为

$$GI_{(v_i)} = \sum_{i \neq j} \frac{\sqrt{d(v_i)+1}}{d_{ij}} \quad (5)$$

其中，平方根用于归一化节点  $j$  的影响。

**定义 3** 综合影响力：本文提出的 CI 评估算法是上述局部影响力和全局影响力两个方面的结合。

节点  $i$  的 CI 可以定义为

$$CI_{(v_i)} = (1-\alpha)LI_{(v_i)} + \alpha GI_{(v_i)} \quad (6)$$

将式(4)和式(5)代入式(6)可得

$$CI_{(v_i)} = (1-\alpha) \sum_{j \in \Gamma_i} -p_i \log b(p_i) + \alpha \sum_{i \neq j} \frac{\sqrt{d(v_i)}}{d_{ij}} \quad (7)$$

其中，可调参数  $\alpha$  控制度数在网络中的影响，取值为  $0 \sim 1$ 。

### 算法 1 CI 评估算法

Input:  $G = (V, E)$ ;

Network  $G$  which contains  $N$  nodes and  $E$  edge;

Initialize the information

for each node  $v$  in  $V$  do

for each node  $u$  in  $V$  do

Using Dijkstra to calculate the short-

est distance  $d_{uv}$

end for

end for

for each node  $v$  in  $V$  do

compute  $LI(v)$  using  $LI_{(v_i)} = \sum_{j \in \Gamma_i} -p_i \log(p_i)$

for each node  $u$  in  $V$  do

compute  $GI(v)$  using  $GI_{(v_i)} = \sum_{i \neq j} \frac{\sqrt{d(v_i)}}{d_{ij}}$

$\forall$  indices  $i \neq j$  for  $LI_{(v_i)} = \sum_{j \in \Gamma_i} -p_i \log(p_i)$  and

$$GI_{(v_i)} = \sum_{i \neq j} \frac{\sqrt{d(v_i)}}{d_{ij}}$$

end for

compute  $CI_{(v)}$  using  $CI_{(v_i)} = (1-\alpha)LI_{(v_i)} + \alpha GI_{(v_i)}$

end for

Rank the influence of all nodes

Return  $CI_{(v)}$

Output: Sorted  $CI_{(v)}$

## 2.3 CI 评估算法的时间复杂度

CI 评估算法的实现分为两部分：计算节点全局影响力，使用 Dijkstra[35] 计算最短路径距离，复杂度为  $O(n^2)$ ；计算节点局部影响力，结合求得的全局影响力进而计算节点的综合影响力，复杂度为  $O(n)$ 。因此，CI 评估算法的计算复杂度为  $O(n^2)$ 。可以看出 CI 评估算法的时间复杂度低于 BC 和 CC，并且其准确度优于先进的 GIC、GSM 算法。

## 2.4 示例说明

基于上述过程，本文通过一个简单的示例网络

解释 CI 评估算法的表现，示例网络如图 2 所示，由 12 个节点和 14 条边组成，以节点 6 的影响力来说明 CI 评估算法，这里设置  $\alpha = 0.5$ 。

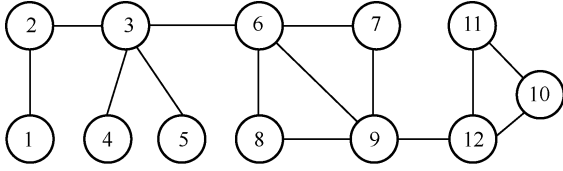


图 2 示例网络

首先，计算节点 6 的邻居节点的度数， $d_{(v_3)} = 4$ 、 $d_{(v_7)} = 2$ 、 $d_{(v_8)} = 2$ 、 $d_{(v_9)} = 4$ ，得到节点 6 的所有邻居节点度数进行求和  $\text{total\_6} = 12$ 。

其次，通过式(4)~式(5)求出每个邻居节点的局部影响力  $LI_{(v_i)}$  和全局影响力  $GI_{(v_i)}$ ： $LI_{(v_3)} = 0.16$ ， $GI_{(v_3)} = 2.24$ ； $LI_{(v_7)} = 0.13$ ， $GI_{(v_7)} = 1.73$ ； $LI_{(v_8)} = 0.13$ ， $GI_{(v_8)} = 1.73$ ； $LI_{(v_9)} = 0.16$ ， $GI_{(v_9)} = 2.24$ ；所以，求得节点  $v_6$  的局部影响力为  $LI_{(v_6)} = LI_{(v_3)} + LI_{(v_7)} + LI_{(v_8)} + LI_{(v_9)} = 0.16 + 0.13 + 0.13 + 0.16 = 0.58$ ；节点  $v_6$  的全局影响力为  $GI_{(v_6)} = GI_{(v_3)} + GI_{(v_7)} + GI_{(v_8)} + GI_{(v_9)} = 2.24 + 1.73 + 1.73 + 2.24 = 7.94$ 。

最后，通过式(6)~式(7)计算节点  $v_6$  的综合影响力，所以由上述过程可以得到节点  $v_6$  的综合影响力为  $CI_{(v_6)} = 0.5 \times LI_{(v_6)} + 0.5 \times GI_{(v_6)} = 0.5 \times 0.58 + 0.5 \times 7.94 = 4.26$ 。类似地，可以计算其他节点的影响。示例网络在 CI 评估算法下的节点排序结果见表 1，其中， $CI_{(v_6)} = 4.26$  是最大值，因此， $v_6$  是最重要的节点。 $CI_{(v_7)} = CI_{(v_8)} = 2.39$ ，表明  $v_7$  和  $v_8$  在示例网络中具有相同的影响力。同理，节点  $v_4$ 、 $v_5$  以及  $v_{10}$ 、 $v_{11}$  在示例网络中具有相同的影响力。

表 1 示例网络在 CI 评估算法下的节点排序结果

排序	节点	CI
1	6	4.26
2	9	4.14
3	3	3.66
4	12	3.08
5	7	2.39
6	8	2.39
7	10	2.01
8	11	2.01
9	2	1.94
10	4	1.12
11	5	1.12
12	1	0.87

### 3 实验分析

为了证明本文算法的优越性，与 BC、CC、GIC、GSM 4 种算法进行对比，其中 BC 与 CC 是两种经典算法，GIC 和 GSM 是近两年被提出的比较优越的算法。相关算法如下，其中， $G$  用来表示每个网络， $G = (V, E)$ ， $V$  和  $E$  分别代表网络  $G$  的节点和边。

#### 1) 介数中心性

介数中心性 (BC) 基于全局拓扑，因为它考虑通过节点的最短路径，具有较高 BC 值的节点将比其他节点更关键。节点  $i$  的 BC 可以定义为

$$BC_{(i)} = \sum_{j,k \neq i} \frac{g_{jk}(i)}{g_{jk}} \quad (8)$$

其中， $g_{jk}$  为  $j$  和  $k$  之间的路径数， $g_{jk}(i)$  为  $j$  和  $k$  之间经过  $i$  的最短路径数。

#### 2) 接近中心性

接近中心性 (CC) 是另一种基于全局的度量，它考虑图中节点的平均最短路径的倒数。节点  $i$  的 CC 定义为

$$CC_{(i)} = \frac{N-1}{\sum_{j \neq i} d_{ij}} \quad (9)$$

其中， $d_{ij}$  为  $i$  和  $j$  之间的最短距离。

#### 3) 重力指数中心性

重力指数中心性 (GIC) [38] 基于重力定律，它同时使用邻居节点的影响和路径信息来识别有影响的节点。节点  $i$  的 GIC 可以定义为

$$GIC_{(i)} = \sum_{j \in \Psi_i} \frac{Ks_{(i)} \times Ks_{(j)}}{d_{ij}^2} \quad (10)$$

其中， $\Psi_i$  是  $i$  的邻居节点的集合， $Ks_{(i)}$  和  $Ks_{(j)}$  是  $i$  和  $j$  的 K-shell 值。

#### 4) 全局结构模型

全局结构模型 (GSM) [39] 是一种新的有效且高效的全局结构模型，用于识别有影响的节点，该模型考虑了网络中节点的拓扑方面，即自身以及全局重要性。

$$GSM_{(i)} = e^{-\frac{Ks_{(i)}}{N}} \times \sum_{i \neq j} \frac{Ks_{(j)}}{d_{ij}} \quad (11)$$

其中， $Ks_{(i)}$  和  $Ks_{(j)}$  表示  $i$  和  $j$  的 K-shell 值。

### 3.1 实验数据

首先, 在一个简单的示例网络中验证 CI 评估算法的效率。然后, 在以下具有不同拓扑特性的真实网络中进行实验, 其中包括 Zebra 网络: 这是一个动物网络, 它由 27 个节点和 111 条边组成; PDZ-base 网络: 它由 212 个节点和 242 条边组成; Air-traffic 网络: 这是空中交通网络, 由 1 226 个节点和 2 615 条边组成; E-mail 网络: 该网络由 1 133 个节点和 5 451 条边组成; Friendship 网络: 这是一个友谊网络, 由 1 858 个节点和 12 572 条边组成; As-200010 网络: 该网络包含 6 474 个节点和 12 572 条边。6 个真实世界网络的统计特征见表 2, 其中  $|M|$  是网络的边数,  $|N|$  是网络中节点的数量,  $\langle K \rangle$  是平均度数,  $\alpha$  是本文使用的可调参数,  $K_{\max}$  网络中最大的度、 $\langle CC \rangle$  是网络的平均聚类系数。

表 2 6 个真实世界网络的统计特征

网络	$ M $	$ N $	$\langle K \rangle$	$\alpha$	$K_{\max}$	$\langle CC \rangle$
Zebra	111	27	8	1.0	14	0.876
PDZ-base	242	212	2	1.0	21	0.005 3
Air-traffic	2 615	1 226	4.265 9	1.0	37	0.04
E-mail	5 451	1 333	9.615	0.4	71	0.220 2
Friendship	12 534	1 858	5.745	0.2	85	0.167
As-200010	12 572	6474	3.92	0.2	1 458	0.253

### 3.2 评估指标

#### 3.2.1 SIR 模型

不同的流行病模型在复杂网络中具有不同的传播机制。为了评估识别重要节点的效率, 在实验中使用 SIR 模型模拟排名节点的传播影响。模型中的节点有易染 (S)、感染 (I) 和恢复 (R) 3 种状态。S 态代表健康, I 态指已经被感染, R 态也称为免疫态, 意味着个体从感染态转换到恢复态后就不再被感染。一开始网络中的所有种子节点都是易感的。在每个时间段内, 种子节点随机选择其直接邻居之一并以一定的概率  $\beta$  感染它。同时, 每个被感染的节点都会以概率  $\mu$  恢复, 不会再次被感染。SIR 模型状态转换示意图如图 3 所示。



图 3 SIR 模型状态转换示意图

假设网络中 3 种状态的节点所占的比例分别为

$s$ 、 $x$  和  $r$ , SIR 模型中的方程为  $\frac{ds}{dt} = -\beta sx$ ,

$\frac{ds}{dt} = \beta sx - \mu x$  和  $\frac{dr}{dt} = \mu x$ 。同时,  $s$ 、 $x$  和  $r$  满足

$$s + x + r = 1 \tag{12}$$

通过消去  $x$  求解方程可得

$$s = s_0 e^{-\beta \gamma / \mu} \tag{13}$$

其中,  $s_0$  是初始状态  $s$  的取值。然后将式(12)代入式(13)可得

$$\frac{dr}{dt} = \mu(1 - r - s_0 e^{-\beta \gamma / \mu}) \tag{14}$$

通过式(14)、式(13)、式(12)可依次求出  $r$ 、 $s$  和  $x$ 。SIR 模型的时间演变图如图 4 所示。显然, 经过足够的传播迭代, 网络可以达到没有感染的稳定阶段。

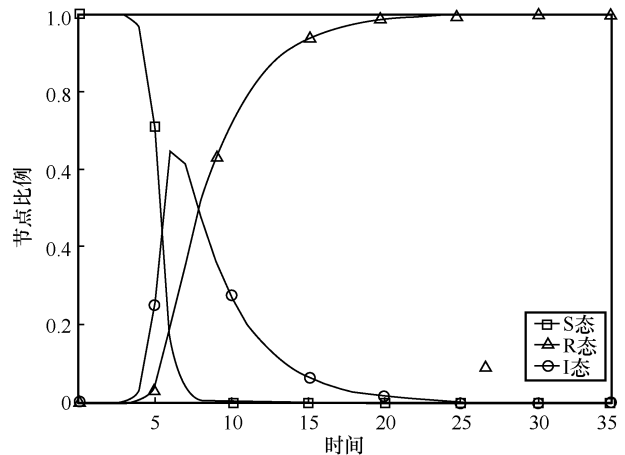


图 4 SIR 模型的时间演变图

#### 3.2.2 Kendall 系数 $\tau$

本文使用 Kendall ( $\tau$ ) 系数进一步计算 CI 评估算法的性能, 它是一种秩相关系数, 通过使用成对的观测值, 并根据成对之间的一致性和不一致模式确定关联强度。它的取值范围为  $[0,1]$ , 0 代表完全不一致, 1 代表完全一致, Kendall 系数越大就说明算法越准确。Kendall 系数的定义为

$$\tau(U, V) = \frac{n_c - n_d}{n(n-1)/2} \tag{15}$$

其中,  $n_c$  和  $n_d$  分别表示两个序列中成对之间一致和不一致对的数量。

### 3.3 实验结果

实验通过使用 SIR 和 Kendall 系数  $\tau$  来验证 CI 评估算法的有效性, 将获得的结果与 BC、CC、GIC、

GSM 这些著名的启发式算法进行比较。首先，将提出的 CI 评估算法应用在一个由 12 个节点和 14 条边组成的简单的示例网络上，示例网络的结构如图 2 所示。CI 评估算法和其他基准算法对示例网络的排序结果见表 3，其中，CI\* 和 SIR\* 分别代表节点的 CI 值和 SIR 值。此外，本文对比了 CI 评估算法和其他方法以及 SIR 在给定示例网络上生成的排序列表，可以看到 CI 评估算法的排名与 SIR 模型的排序一致，因此，CI 评估算法的 Kendall 系数等于 1，各方法在示例网络中的 Kendall 系数如图 5 所示，可以看出，就 Kendall 系数而言，CI 评估算法比其他方法表现更好。

表 3 CI 评估算法和其他基准算法对示例网络的排序结果

排序	BC	CC	GIC	GSM	CI	CI*	SIR	SIR*
1	3	6	6	9	6	8.51	6	0.437
2	6	3	9	6	9	8.28	9	0.390
3	9	9	3	3	3	7.32	3	0.333
4	12	7	12	12	12	6.15	12	0.322
5	2	8	7	7	7	4.77	7	0.318
6	1	12	8	8	8	4.77	8	0.285
7	4	2	2	2	10	4.02	10	0.281
8	5	4	10	10	11	4.02	11	0.278
9	7	5	11	11	2	3.87	2	0.256
10	8	11	4	4	4	2.24	4	0.239
11	11	10	5	5	5	2.24	5	0.229
12	10	1	1	1	1	1.73	1	0.186

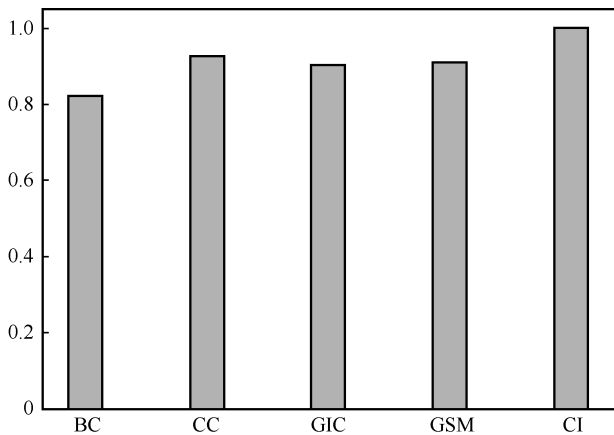


图 5 各方法在示例网络中的 Kendall 系数

### 3.3.1 基于 SIR 模型的性能评估

在实验中，本文使用了 SIR 模型来区分节点的重要性并验证其效率。为了更清楚地区分重要节点，为大型复杂网络（Friendship 和 As-20001）选择了相对较小的  $\beta$ ，设置  $\beta = 0.01$ ，因为更大的

传播率  $\beta$  将导致传播发生在几乎整个网络中；在这种情况下，很难区分不同节点的影响。对于 Zebra、PDZ-base、Air-traffic、和 E-mail 等小型复杂网络，将  $\beta$  固定为 0.1，恢复概率  $\mu = 0.9$ ，时间步长  $t = 1000$ 。首先，使用不同的中心性计算每个节点的影响力，并降序排序。其次，每个排名节点被视为一个种子节点（每次运行只有一个种子节点）。最后，通过种子节点的平均转数为 1000 次，成功获得了受感染的节点数。各方法对节点进行排序后的传播影响如图 6 所示，x 轴表示节点影响力的排序列表，y 轴表示感染的节点数。一般来说，一个更关键的节点会感染更多的节点；因此，一种有效的方法可以产生一条从左到右递减的曲线，图 7 是图 6 的部分细节展示图，从各个网络中随机取了连续的 30 个节点来展示，从图 6 可以看出所有算法都得到了一条从左到右递减的曲线，但是从图 7 可以看出 CI 评估算法的波动幅度更小，所以 CI 评估算法相比其他算法对节点的排序更准确，因此 CI 评估算法比其他算法具有更优的排序单调性。

### 3.3.2 不同网络下排名前 10 的节点性能比较

此外，本文比较了 CI 评估算法和其他基准算法选择的前 10 个节点的效果。所有前 10 个不同的节点都作为种子节点，时间步长  $t$  固定在 1~20，其中， $F(t)$  表示在时间  $t$  时感染和恢复的节点数。一个更关键的节点传播的速度会更快并且感染的节点数也会更多，SIR 传播对使用 CI 评估算法和其他基准算法的前 10 名排序效果的影响如图 8 所示，可以清楚地看出在不同的网络下 CI 评估算法相比其他算法传播数量更多、传播速度更快。CI 评估算法与 SIR 传播的结果越接近，也说明算法的准确度越高，各算法在不同网络下影响力排名前 10 的节点见表 4，可以看出 CI 评估算法对前 10 个重要节点的排序与 SIR 模型基本一致，因此 CI 评估算法在重要节点识别的准确度方面优于其他对比算法。

### 3.3.3 基于 Kendall 系数的性能评估

图 8 说明了在不同网络中对 CI 评估算法和其他算法的 Kendall 结果，感染概率  $\beta \in [0.01, 0.10]$ ， $\alpha$  是 CI 评估算法的可调参数，随着传播动力学的变化而变化。Kendall 的  $\tau$  是通过将 CI 评估算法和其他算法产生的排序结果与 SIR 在 6 个真实网络上产生的排序结果进行比较得到的。可以看出，在感染

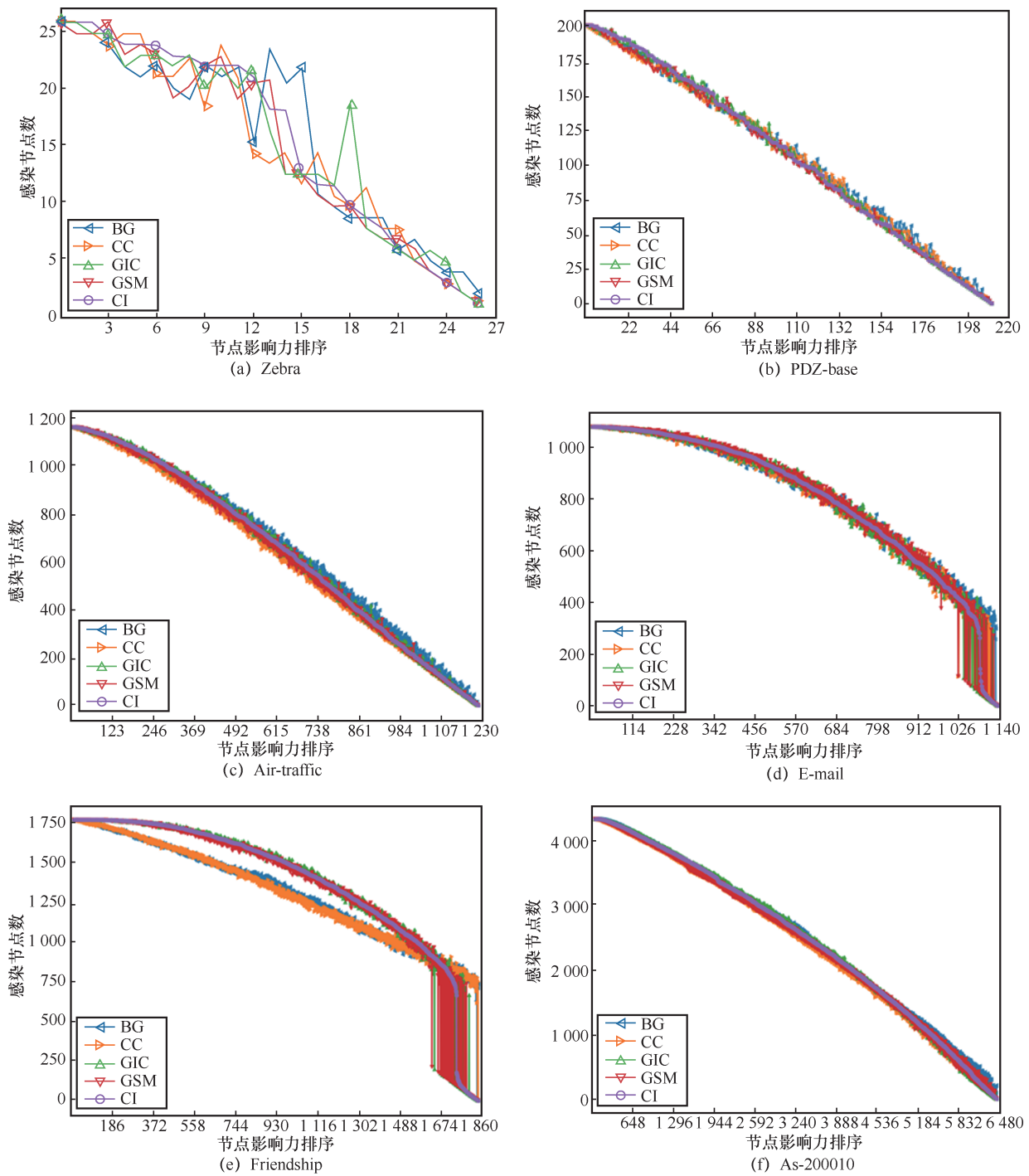


图6 各方法对节点进行排序后的传播影响

概率  $\beta \in [0.01, 0.10]$  下, CI 评估算法获得了更高的 Kendall 系数, 这表明 CI 评估算法产生的排序结果与实际的 SIR 传播过程基本相同。可调参数  $\alpha$  用于控制网络中的程度影响。在实验过程中, 本文通过改变可调参数  $\alpha$  选择一个优化值, 以达到 CI 评估算法和 SIR 之间相似度  $\tau$  (Kendall 相关系数) 的最佳值。当  $\alpha = 0.1$  时, 度数对 CI 评估

算法的影响是最小的; 当  $\alpha = 1$  时, 节点的度对 CI 评估算法的影响最大。从实验模拟中, 可以看出在 Zebra、PDZ\_base、Air-traffic 等不同的网络中, CI 评估算法的传播效率在潜在的度数情况下是理想的。而对于 Friendship 或 As-200010 等网络, 如果使用较小的  $\alpha$  来降低度数影响, 则 CI 评估算法与 SIR 也具有较高的相似度。同样,

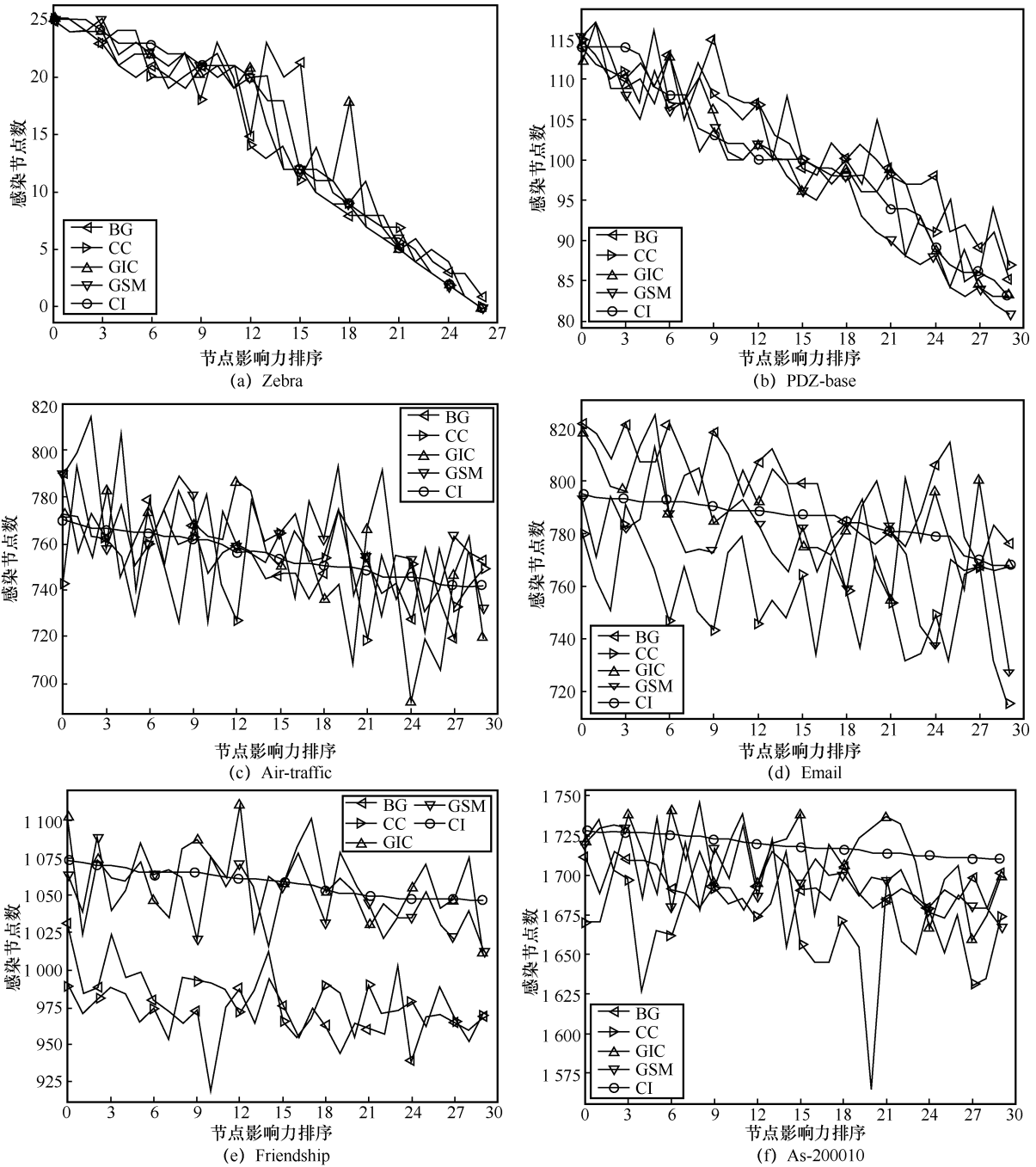


图 7 各方法对节点进行排序后部分节点的传播影响

对于 E-mail 网络，当  $\alpha = 0.4$  时，CI 评估算法和 SIR 之间的 Kendall 系数取得最佳值，各方法在不同网络下的 Kendall 系数  $\tau$  的比较如图 9 所示，可以看出，使用 CI 评估算法以及各基准算法在小型和大型复杂网络上有不同的表现，但是 CI 评估算法在不同的网络上都获得了比其他算法更高的 Kendall 系数，因此 CI 评估算法比其他对比算法具有更稳定的度量结果，对网络拓扑的适应性也更强。

### 3.3.4 实验有效性提升分析

本文提出的 CI 评估算法同时考虑了网络的局部属性和全局属性，使得算法的度量范围比其他算法更为全面，并且通过使用节点之间的度数、节点信息熵、可调参数和最短路径等多属性对节点的重要性进行综合评估，使得其对网络中节点影响力排序的结果更加准确，此外，引入一个调节局部和全局属性权重的可调参数  $\alpha$  使得 CI 评估算法具有更稳定的度量结果，对网络拓扑的适应性也更强。

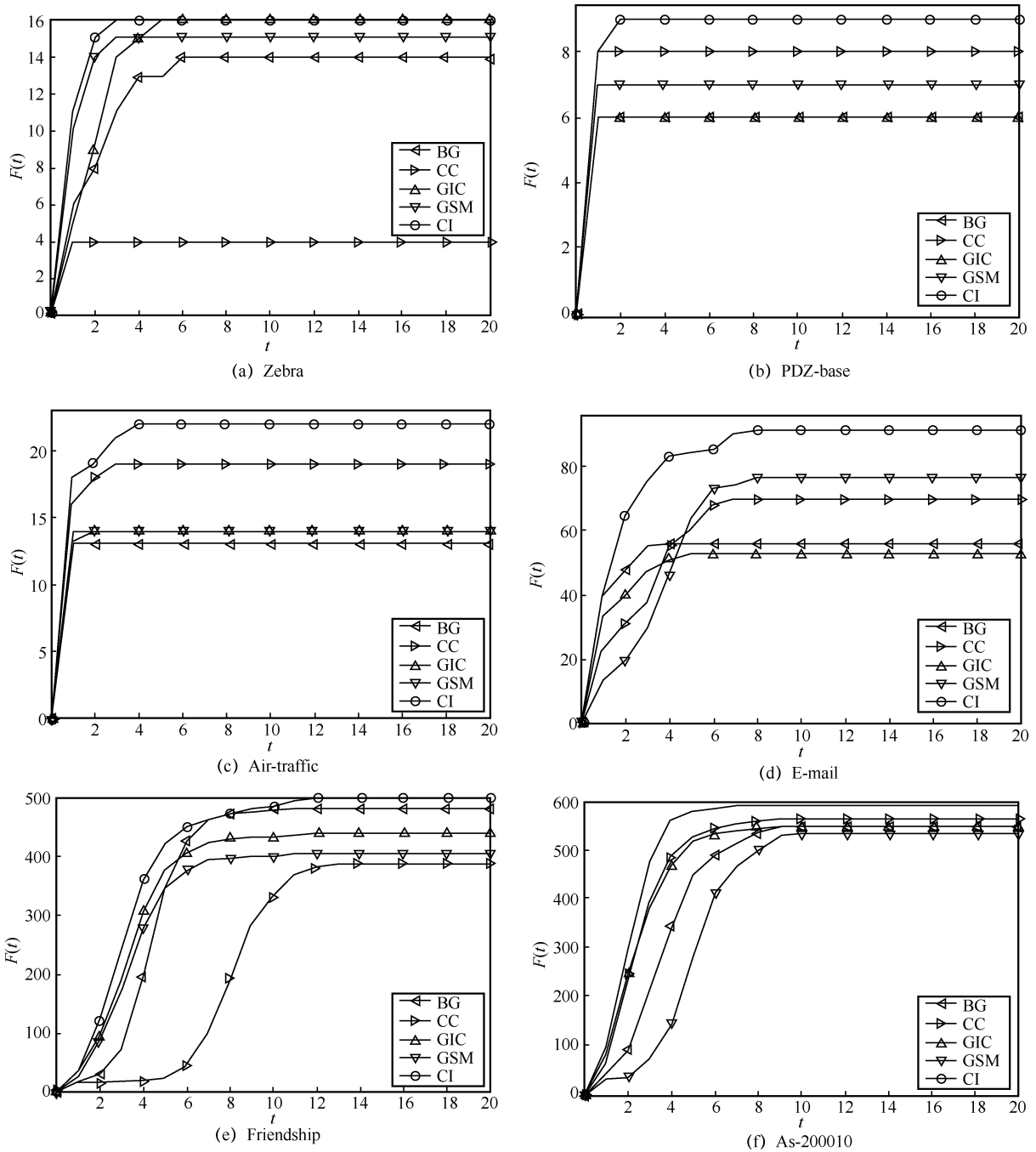


图8 SIR传播对使用CI评估算法和其他基准算法的前10名排序效果的影响

### 4 结束语

本文同时从网络的全局和局部拓扑出发研究了复杂网络中的关键节点识别问题，提出了一种识别关键节点的综合影响力评估算法，该算法利用了网络中节点的局部和全局属性信息，通过使用 SIR 模型和 Kendall 相关系数在真实世界的网络数据集中进行实验来评估其效率，结果表明，所提出的方法能在识别关键节点方面优于

BC、CC、GIC、GSM 算法，并且具有更好的排序单调性、更稳定的度量结果，对网络拓扑的适应性更强，适用于绝大多数具有不同结构的真实网络。但是，目前 CI 评估算法的设计仍然存在一些使用的限制。例如，CI 评估算法的设计仅限于未加权无向网络的情况。为了应对这些挑战，将进一步把所提方法应用于加权有向、加权无向、二分网络等网络，这将是未来的研究重点。

表 4 各算法在不同网络下影响力排名前 10 的节点

网络	Zebra						Air_traffic							Friendship				
	BC	CC	GIC	GSM	CI	SIR	BC	CC	GIC	GSM	CI	SIR	BC	CC	GIC	GSM	CI	SIR
1	16	10	2	2	2	2	68	68	52	52	52	52	237	237	237	237	237	237
2	13	13	3	3	3	3	52	52	68	68	68	68	137	137	238	238	238	238
3	10	1	9	9	9	9	213	148	113	113	113	113	169	238	168	168	168	168
4	14	2	11	11	11	11	312	135	47	34	312	312	238	177	137	137	44	44
5	17	3	12	12	12	12	135	711	44	47	47	47	251	176	44	44	137	137
6	1	8	13	10	10	10	136	424	312	44	44	44	44	118	177	356	177	177
7	8	9	10	13	13	13	212	136	89	116	13	13	296	178	176	177	176	176
8	4	11	1	1	1	1	660	617	34	46	187	187	168	168	46	176	46	46
9	2	12	8	8	8	8	523	20	109	109	34	34	23	3	356	244	356	356
10	3	14	4	4	4	4	221	116	46	148	46	46	65	117	3	3	3	3
网络	PDZ_base						Email							AS_200010				
	BC	CC	GIC	GSM	CI	SIR	BC	CC	GIC	GSM	CI	SIR	BC	CC	GIC	GSM	CI	SIR
1	14	14	14	20	14	14	333	333	105	105	105	105	2	2	2	2	2	2
2	114	20	20	14	20	20	105	23	42	42	16	16	7	10	10	10	10	10
3	45	45	19	45	19	19	23	105	16	16	42	42	10	7	7	7	7	7
4	61	62	45	21	26	26	578	42	333	333	333	333	29	3	1	3	1	1
5	144	106	37	106	45	45	76	41	23	23	23	23	8	1	8	1	8	8
6	5	18	21	18	3	3	233	76	196	196	196	196	1	4	3	486	3	3
7	26	99	44	19	32	32	135	233	41	41	41	41	3	6	23	4	4	4
8	20	61	26	99	37	37	41	52	76	3	76	76	394	8	42	8	6	6
9	99	57	32	84	34	34	355	135	3	332	233	3	403	29	29	29	23	23
10	106	8	34	16	44	44	42	378	233	76	3	3	6	22	6	9	29	29

参考文献：

[1] LI X X, HU W, SHANG T, et al. Security analysis of power system network based on complex network[C]//Proceedings of 2020 IEEE 4th Conference on Energy Internet and Energy System Integration. Piscataway: IEEE Press, 2020: 4336-4341.

[2] KHOSROSERESHI F, MOAVENI B. Traffic modeling and validation for intersecting metro lines by considering the effect of transfer stations[J]. IEEE Access, 10: 1946-1955.

[3] GARCIA J O, ASHOURVAN A, MULDOON S, et al. Applications of community detection techniques to brain graphs: algorithmic considerations and implications for neural function[J]. Proceedings of the IEEE, 2018, 106(5): 846-867.

[4] ZHAO J Y, MO H M, DENG Y. An efficient network method for time series forecasting based on the DC algorithm and visibility relation[J]. IEEE Access, 8: 7598-7608.

[5] LI X Y, ZHOU M Y, WU X Y, et al. A novel method to identify multiple influential nodes in complex networks[J]. Scientia Sinica Informationis, 2019, 49(10): 1333-1342.

[6] LIU X, HE D, LIU C, et al. Social network information propagation dynamic analysis based on time-varying damping motion[J]. Scientia Sinica Informationis, 2021, 51(11): 1867.

[7] TRONCOSO F, WEBER R. Integrating relations and criminal background to identifying key individuals in crime networks[J]. Decision Support Systems, 2020, 139: 113405.

[8] HUANG C S, ZHU L. Robust evaluation method of communication network based on the combination of complex network and big data[J]. Neural Computing and Applications, 2021, 33(3): 887-896.

[9] FAN L J, JIANG Z Y, WANG Q. Robustness evaluation for real traffic network from complex network perspective[J]. International Journal of Modern Physics C, 2021, 32(8): 2150102.

[10] ZHUD R, WANGH F, DUAN J D, et al. Study of Improved PageRank algorithm in key node identification of power system[C]// Proceedings of 2020 4th International Conference on HVDC (HVDC). Piscataway: IEEE Press, 2020: 1154-1159.

[11] XU X, ZHU C, WANG Q Y, et al. Identifying vital nodes in complex networks by adjacency information entropy[J]. Scientific Reports, 2020, 10: 2691.

[12] MAJI G, MANDAL S, SEN S. A systematic survey on influential spreaders identification in complex networks with a focus on K-shell based techniques[J]. Expert Systems With Applications, 2020, 161: 113681.

[13] MAJI G, NAMTIRTHA A, DUTTA A, et al. Influential spreaders identification in complex networks with improved K-shell hybrid method[J]. Expert Systems With Applications, 2020, 144: 113092.

[14] ZHAO G H, JIA P, ZHOU A M, et al. InfGCN: identifying influential nodes in complex networks with graph convolutional networks[J]. Neurocomputing, 2020, 414: 18-26.

[15] YANG H H, AN S. Critical nodes identification in complex networks[J]. Symmetry, 2020, 12(1): 123.

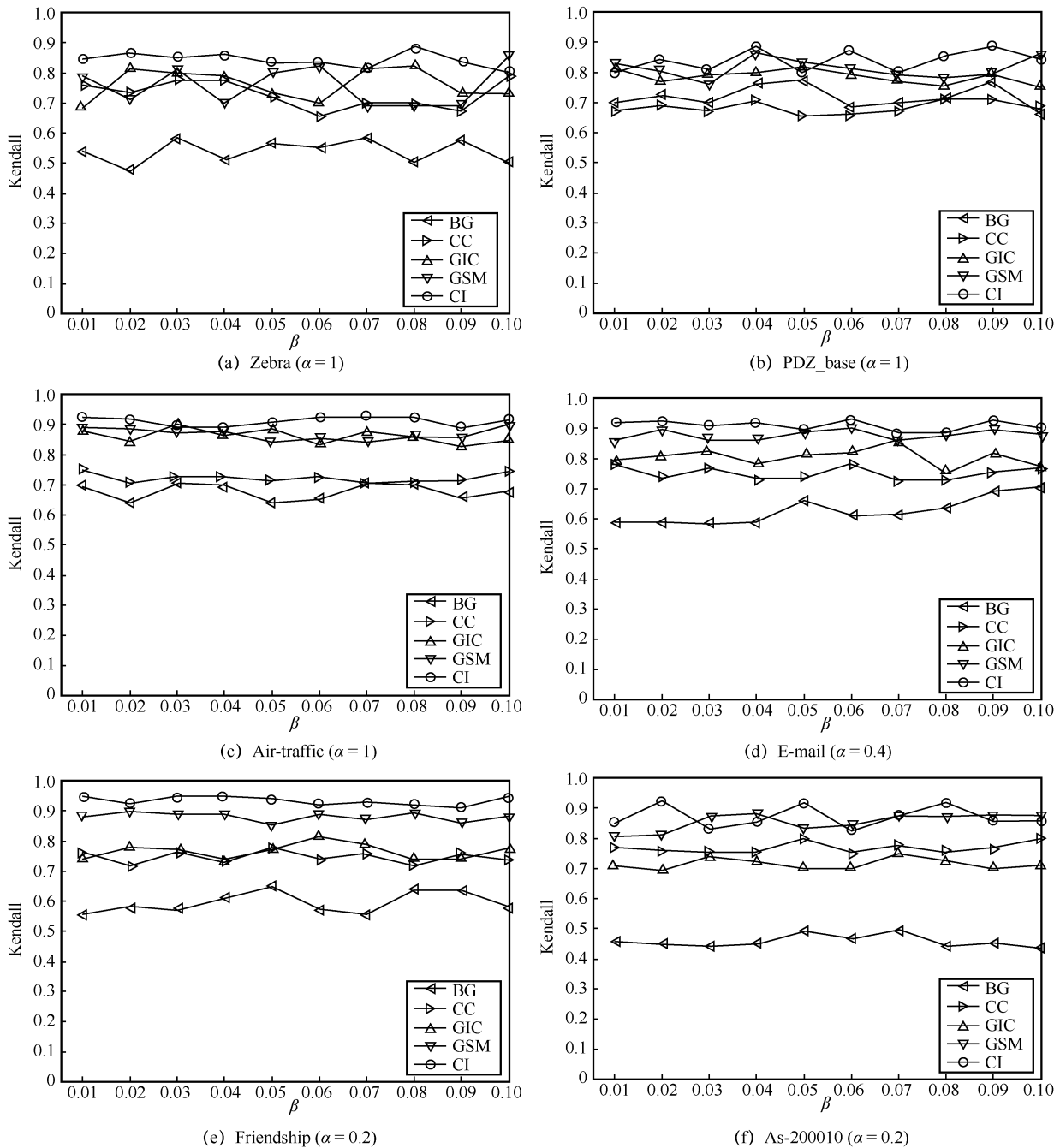


图9 各方法在不同网络下的 Kendall 系数  $\tau$  的比较

[16] GUO C G, YANG L W, CHEN X, et al. Influential nodes identification in complex networks via information entropy[J]. Entropy (Basel, Switzerland), 2020, 22(2): 242.

[17] MAHARANI W, ADIWIJAYA, GOZALI A A. Degree centrality and eigenvector centrality in twitter[C]//Proceedings of 2014 8th International Conference on Telecommunication Systems Services and Applications (TSSA). Piscataway: IEEE Press, 2014: 1-5.

[18] LIU X Y, DING N, LIU C, et al. Novel social network community discovery method combined local distance with node rank optimization function[J]. Applied Intelligence, 2021, 51(7): 4788-4805.

[19] WANG Y F, YAN G H, MA Q Q, et al. Identifying influential spreaders on weighted networks based on Cluster Rank[C]//Proceedings of

2017 10th International Symposium on Computational Intelligence and Design (ISCID). Piscataway: IEEE Press, 2017: 476-479.

[20] KITSAK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11): 888-893.

[21] MALANG K, WANG S L, PHAPHUANGWITTAYAKUL A, et al. Identifying influential nodes of global terrorism network: a comparison for skeleton network extraction[J]. Physica A: Statistical Mechanics and Its Applications, 2020, 545: 123769.

[22] OKAMOTO K, CHEN W, LI X Y. Ranking of closeness centrality for large-scale social networks[C]//Frontiers in Algorithmics, 2008: 186-195.

[23] BARTHLEMY M. Betweenness centrality in large complex net-

- works[J]. The European Physical Journal B - Condensed Matter, 2004, 38(2): 163-168.
- [24] KUMAR P V, BALASUBRAMANIAN C. Trilateral spearman katz-centrality based least angle regression for influential node tracing in social network[J]. Wireless Personal Communications, 2022, 122(3): 2767-2790.
- [25] YANG Y Z, WANG X, CHEN Y, et al. Identifying key nodes in complex networks based on global structure[J]. IEEE Access, 8: 32904-32913.
- [26] IBNOULOAFI A, EL HAZITI M, CHERIFI H. M-Centrality: identifying key nodes based on global position and local degree variation[J]. Journal of Statistical Mechanics: Theory and Experiment, 2018, 2018(7): 073407.
- [27] LI Z, REN T, MA X Q, et al. Identifying influential spreaders by gravity model[J]. Scientific Reports, 2019, 9: 8387.
- [28] ZHUANG Y B, LI Z H, ZHUANG Y J. Identification of influencers in online social networks: measuring influence considering multidimensional factors exploration[J]. Heliyon, 2021, 7(4): e06472.
- [29] CHIANG M F, LIM E P, LEE W C, et al. CO2Vec: embeddings of co-ordered networks based on mutual reinforcement[C]//Proceedings of 2020 IEEE 7th International Conference on Data Science and Advanced Analytics. Piscataway: IEEE Press, 2020: 148-157.
- [30] XU Y, FENG Z D, QI X Q. Signless-Laplacian eigenvector centrality: a novel vital nodes identification method for complex networks[J]. Pattern Recognition Letters, 2021, 148: 7-14.
- [31] WU X M, WU J S, LI Y F, et al. Link prediction of time-evolving network based on node ranking[J]. Knowledge-Based Systems, 2020, 195: 105740.
- [32] PAGE L, BRIN S, MOTWANI R, et al. The Pagerank citation ranking: bringing order to the Web[EB]. 1998.
- [33] ZHANG Y C, YU J X, HOU J Y. HITS and related algorithms[EB]. 2006.
- [34] LI Q, ZHOU T, LÜL Y, et al. Identifying influential spreaders by weighted LeaderRank[J]. Physica A: Statistical Mechanics and Its Applications, 2014, 404: 47-55.
- [35] TONG Q, WEI S, CHANG Z. How to identify the most powerful node in complex networks? A novel entropy centrality approach[J]. Entropy, 2017, 19(11): 614.
- [36] BARBEHENN M. A note on the complexity of Dijkstra's algorithm for graphs with weighted vertices[J]. IEEE Transactions on Computers, 1998, 47(2): 263.
- [37] 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.
- [38] MA LL, 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.
- [39] ULLAH A, WANG B, SHENG J F, et al. Identification of nodes in fluence based on global structure model in complex networks[J]. Scientific Reports, 2021, 11: 6173.

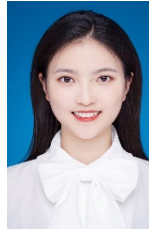
## [作者简介]



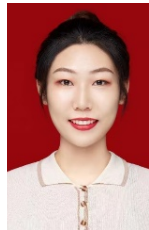
蒋伟进（1964-），男，博士，湖南工商大学计算机学院二级教授，主要研究方向为网络安全、社会计算、区块链技术和群体智能感知。



杨莹（1999-），女，湖南工商大学计算机学院硕士生，主要研究方向为复杂网络、网络安全和区块链技术。



罗田甜（1998-），女，湖南工商大学前交叉学院硕士生，主要研究方向为网络安全、区块链技术和社会计算。



周文颖（1999-），女，湖南工商大学计算机学院硕士生，主要研究方向为网络安全、区块链技术和社会计算。



李恩（1995-），男，湖南工商大学计算机学院硕士生，主要研究方向为网络安全和区块链技术。



张小威（1999-），男，湖南工商大学计算机学院硕士生，主要研究方向为网络安全、社会计算和社交网络。