



基于交叉熵的节点重要性排序算法

龚志豪, 蒋沅*, 代冀阳, 杨智翔

(南昌航空大学信息工程学院 南昌 330063)

【摘要】如何高效地度量节点的重要性一直是复杂网络研究的热点问题。在节点重要性研究中, 目前已有许多算法被提出用于判断关键节点, 然而多数算法局限于时间复杂度过高或评估角度单一。考虑到熵可用于定量描述信息量的大小, 因此, 提出了一种基于交叉熵的节点重要性排序算法, 该算法兼顾了中心节点与其近邻节点之间的整体影响力, 并将节点的邻域拓扑信息有机地融合, 使用交叉熵值来量化节点之间的信息差异性。为验证该算法的性能, 首先采用单调关系、极大连通系数、网络效率以及 SIR 模型作为评价指标, 其次在 8 个不同领域的真实网络上与其他 7 种算法进行比较实验。实验结果表明, 该算法具有有效性和适用性, 此外时间复杂度仅为 $O(n)$, 适用于大型网络。

关键词 攻击策略; 交叉熵; 复杂网络; 节点重要性

中图分类号 TP301 **文献标志码** A **doi**:10.12178/1001-0548.2023058

Node Importance Ranking Algorithm Based on Cross Entropy

GONG Zhihao, JIANG Yuan*, DAI Jiyang, and YANG Zhixiang

(Institute of Information Engineering, Nanchang Hangkong University Nanchang 330063)

Abstract How to efficiently measure the importance of nodes has been a hot issue in the research of complex networks. In the research of node importance, many algorithms have been proposed to judge key nodes, but most of them are limited to high time complexity or single evaluation angle. Considering that entropy can be used to quantitatively describe the amount of information, this paper proposes a node importance ranking algorithm based on cross entropy. This algorithm takes into account the overall influence among the central node and its neighbor nodes, organically fuses the neighborhood topology information of nodes, and uses cross entropy to quantify the information differences between nodes. In order to verify the performance of the algorithm, this paper first uses monotone relation, maximum connectivity coefficient, network efficiency and SIR model as evaluation indicators, and then compares with other seven algorithms on eight real networks in different fields. The experimental results show that the algorithm proposed in this paper is effective and applicable, and the time complexity is only $O(n)$, which is suitable for large networks.

Key words attack strategy; cross entropy; complex network; node importance

随着网络科学的发展与进步, 复杂系统已深入人类社会的各个领域, 复杂网络作为刻画复杂系统的工具, 在生态、社会、经济、交通等诸多系统中有着重要影响^[1-2]。关键节点影响着网络的结构和信息传递功能, 评估关键节点是复杂网络的研究热点^[3-4]。一方面, 快速准确地识别出关键节点并提供保护机制可提升网络的抗毁性^[5]。另一方面, 基于关键节点也可以提出更高效的攻击策略^[6-7]。因此, 设计高效的关键节点评估算法具有重要的理论和实践意义。

近年来, 研究人员关于识别关键节点已有许多研究成果, 经典的评估算法有基于节点邻近信息的度中心性^[8]、基于最短路径数目的介数中心性^[9]、基于平均距离的接近中心性^[10]以及基于网络位置的 K-壳分解法^[11]等。其中度中心性虽然简单直接, 但对邻居节点的重要性区分度较低, 并且考虑的邻近信息有限, 因此评估的精确性不高。介数和接近中心性仅考虑信息在最短路径上传播, 而实际上传播可能基于其他可达路径, 此外基于路径的算法时间复杂度较高, 不适用于大型网络。而具有较

收稿日期: 2023-03-01; 修回日期: 2023-07-14

基金项目: 国家自然科学基金(61663030, 61663032); 南昌航空大学研究生创新专项资金(YC2022-049)

作者简介: 龚志豪(1999-), 男, 主要从事复杂网络分析方面的研究。

*通信作者: 蒋沅, E-mail: jiangyuan@nchu.edu.cn

低时间复杂度的 K-壳分解法认为节点重要性取决于所处网络的位置, 内核层节点的重要性高于边缘的节点, 但对同一壳层的节点却无法进一步区分其重要性差异, 并且节点在剥离时会对网络的整体结构信息造成破坏^[12]。为弥补经典算法评估的局限性, 文献 [13] 考虑节点与其相邻节点之间的相关性, 提出了映射熵来评估网络中节点的重要性。文献 [14] 通过衡量节点的局部拓扑重合度来刻画节点间的相似性, 提出了邻域相似度算法用于评估节点的重要性。文献 [15] 结合节点的 K-壳以及信息熵, 根据其分层大小依次进行迭代, 可区分内层壳与外层壳中的节点重要性。文献 [16] 受引力公式启发, 将节点的度值作为质量, 并将最短路径长度作为距离, 考虑了节点的近邻以及路径信息, 提出了引力模型算法。文献 [17] 考虑节点在网络中所处的位置, 提出了一种基于 K-壳分解法的改进引力模型算法。

熵^[18]可用于定量描述信息量的大小, 当使用熵理论刻画复杂网络时, 信息熵可表征节点的局部重要性, 因此, 可考虑用子网络的熵来表征网络整体结构的特性。如文献 [19] 提出了信息熵来评估复杂系统的结构特征, 取得了较好的成效。文献 [20] 改进了 K-壳值对信息熵的计算, 提出了一种结合节点信息熵与迭代因子的算法。文献 [21] 基于非广延统计力学, 提出了一种局部结构熵来量化复杂网络中的关键节点。文献 [22] 结合网格约束系数以及节点的 K-壳中心性, 基于 Tsallis 熵提出了一种节点重要性识别方法。

受上述研究启发, 本文提出了一种基于交叉熵的节点重要性识别算法 CE+ (cross entropy), 该方法充分考虑了节点自身以及其周围节点信息的整体重要性, CE+ 的值反映了节点与其近邻节点之间的差异性, 并且该算法时间复杂度仅为 $O(n)$, 适用于大型网络。通过在 8 个不同领域的真实网络上进行蓄意攻击实验, 并选用 7 种不同的节点重要性排序算法作为对比, 采用单调性指标^[23]、极大连通系数^[24]、网络效率^[25-26]以及 SIR 模型^[27]等指标验证了本文所提出 CE+ 算法的有效性和适用性。

1 理论基础

假设无向未加权的网络 $G = (V, E)$, 其中 V 是网络 G 的节点集合, $|V| = n$, E 是网络的边集合, $|E| = m$ 。 $A = (a_{ij})_{n \times n}$ 通常表示网络的邻接矩阵, 若节点 i 与节点 j 相连, 则 $a_{ij} = 1$, 否则 $a_{ij} = 0$ 。

1.1 度中心性

度中心性^[9]是度量重要节点性能最基础的指标, 节点的邻居节点数越多, 则自身的影响力越大, 其定义为:

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

式中, $k_i = \sum_{j=1}^n a_{ij}$; n 表示网络的节点数。

1.2 介数中心性

当网络中某一节点 i 存在于其他节点之间的最短路径上并且次数越多时, 则说明节点 i 的关键程度越高, 其定义为:

$$BC_i = \sum_{i \neq s \neq t} \frac{g_{st}^i}{g_{st}} \quad (2)$$

式中, g_{st}^i 表示节点 i 存在于节点 s 到节点 t 之间的最短路径中的次数; g_{st} 表示节点 s 到节点 t 之间的全部路径条数。

1.3 映射熵

文献 [13] 考虑节点中所有邻域相关的局部信息, 提出了映射熵来评估复杂网络中的节点中心性, 定义为:

$$ME_i = -DC_i \sum_{j=1}^M \log DC_j \quad (3)$$

式中, M 是节点 i 的邻居集; DC_j 是其相邻节点之一的度中心性。

1.4 邻域相似度

文献 [14] 综合考虑了节点的度值和邻居节点的拓扑重合度, 提出了一种基于邻域相似度的评估算法, 定义为:

$$LLS_i = \sum_{b, c \in n(i)} (1 - \text{sim}(b, c)) \quad (4)$$

式中, $n(i)$ 为节点 i 的邻居节点; 若节点 b 与 c 无连边, 则 $\text{sim}(b, c) = |n(b) \cap n(c)| / |n(b) \cup n(c)|$; 若节点 b 与 c 有连边, 则 $\text{sim}(b, c) = 1$ 。

1.5 改进的 K-壳分解法

文献 [15] 结合 K-壳分解法以及节点的信息熵, 根据 K-壳的分层大小依次进行迭代, 每层中选择节点信息熵最高的节点, 直到所有节点均被选中为止, 定义为:

$$e_i = - \sum_{j \in \tau(i)} I_j \ln I_j \quad (5)$$

式中, $\tau(i)$ 表示节点*i*的邻居集; $I_i = k_i / \sum_{j=1}^n k_j$ 。

1.6 引力模型

文献 [16] 综合考虑了节点的邻居信息和节点间的路径信息, 分别将节点的度值和最短路径长度类比为物体的质量和距离, 提出了引力模型算法, 计算公式如下:

$$GM_i = \sum_{d_{ij} < R, j \neq i} \frac{k_i k_j}{d_{ij}^2} \quad (6)$$

式中, R 表示截断半径, 通常为平均最短距离的一半。由于仅考虑了截断半径内的引力, 该算法也被称为局部引力模型。

1.7 KSGC 算法模型

文献 [17] 认为节点在网络中所处的位置是一个重要属性, 因此, 在 K-壳分解法的基础上对引力模型进行了改进, 提出了 KSGC 算法用于评估节点的影响力, 计算公式如下:

$$KSGC_i = \sum_{d_{ij} < R, j \neq i} C_{ij} \frac{k_i k_j}{d_{ij}^2} \quad (7)$$

式中, $C_{ij} = e^{\frac{KS_j - k_j}{KS_{\max} - KS_{\min}}}$; KS_{\max} 为节点 K-壳的最大值; KS_{\min} 为节点 K-壳的最小值。

1.8 交叉熵

交叉熵^[28]广泛应用于逻辑回归模型分析中, 其定义为随机分布 p 和 q 之间的自信息差异, 可用来量化两个变量之间的相似度, 通常交叉熵值越大, 则两个变量之间的差异性越大, 其计算公式如下:

$$H(p, q) = E_p[-\ln q(x)] = - \sum_x p(x) \ln q(x) \quad (8)$$

式中, x 表示事件包含的信息量。对式 (8) 同时增减 $\sum_x p(x) \ln p(x)$, 可得:

$$H(p, q) = - \sum_x p(x) \ln q(x) + \sum_x p(x) \ln p(x) - \sum_x p(x) \ln p(x) \quad (9)$$

将式 (9) 中的前两项对数函数合并可得:

$$H(p, q) = - \sum_x p(x) \ln p(x) + \sum_x p(x) \ln \frac{p(x)}{q(x)} \quad (10)$$

则式 (10) 可被定义为:

$$H(p, q) = H(p) + D_{KL}(p||q) \quad (11)$$

式中, $H(p)$ 为信息熵, 表示随机分布 p 的平均信息量; $D_{KL}(p||q)$ 为相对熵, 同样反映随机分布 p 和 q 之

间的差异程度。因此, 交叉熵作为信息熵与相对熵之和, 对于随机变量的信息量及其差异性刻画更加直观。

2 基于交叉熵的节点重要性排序算法

2.1 算法构造

交叉熵可衡量随机变量所包含信息量的差异, 类似地, 复杂网络中不同节点之间的拓扑信息也存在差异, 因此考虑引入交叉熵衡量复杂网络中节点的统计特征。

基于交叉熵的节点重要性排序算法的原理是综合考察节点自身以及其周围节点信息的整体重要性, 结合这两种拓扑信息, 可利用交叉熵值来量化节点分布之间的差异, 若交叉熵值越大, 则节点之间分布差异性越大, 说明该节点代表性越高, 重要性更显著。因此本文提出了基于交叉熵的算法 CE+, 其定义如下:

$$CE_{i+} = - \sum_{j \in \Gamma(i)} I_j \ln I_j \quad (12)$$

式中, $j \in \Gamma(i)$ 表示节点 j 是节点 i 的邻居节点; I_i 表示节点 i 的异构性。考虑到度中心性可在时间复杂度较低的同时较好地反映节点及其近邻节点所构成子网络的拓扑结构, 因此将 I_i 定义为:

$$I_i = \frac{DC_i}{\sum_{j=1}^n DC_j} \quad (13)$$

由此, 节点之间的异构性可得以良好的表征。在基于交叉熵的节点重要性计算方法中, 通过融合中心节点与其局部节点的信息结构并相互交叉考虑, 使得节点影响力的差异刻画更为准确。该算法的伪代码如下。

- 1) 输入: 网络 $G = (V, E)$
- 2) 输出: 节点 i 的重要性评估值 CE_{i+}
 - ① 遍历节点 i 的邻居节点 j
 - ② 计算节点 i 的度中心性 DC_i
 - ③ 计算邻居节点 j 的度中心性之和 $\sum_{j=1}^n DC_j$
 - ④ 计算节点 i 的异构性 I_i
 - ⑤ 计算节点 i 的交叉熵值 CE_{i+}

表 1 列出了 8 种不同评估算法的时间复杂度, 分别包括局部、全局以及混合 3 个类型的网络信息。其中 CE+ 算法的时间复杂度与 DC、LLS、ME 以及 IKS 算法同为最低。

表 1 不同评估算法的时间复杂度

算法	类型	时间复杂度
DC	Local	$O(n)$
BC	Global	$O(n^2)$ or $O(nm)$
LLS	Local	$O(n)$
KSGC	Hybrid	$O(n^2)$
GM	Global	$O(n^2)$
ME	Local	$O(n)$
IKS	Hybrid	$O(n)$
CE+	Local	$O(n)$

2.2 示例分析

为初步验证 CE+算法的量化分辨率与精确性, 构建示例网络如图 1 所示。

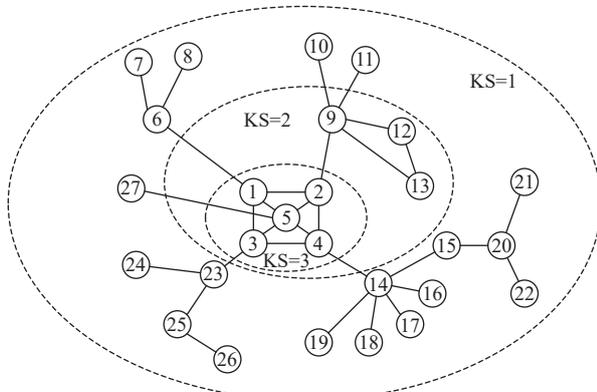


图 1 示例网络

在示例网络中, 以节点 24 为例, 其交叉熵计算过程如下:

$$CE_{24+} = - \sum_{j \in \Gamma(i)} I_i \ln I_j = - \left(\frac{1}{3} \ln \frac{3}{4+2+1} \right) = 0.2824 \quad (14)$$

节点 24 的度值为 1, 其邻居节点 23 的度值为 3, 而节点 23 的其余邻居节点 3、25 的度值分别为 4、2, 故计算节点 24 的交叉熵值为 0.2824。同理, 可以计算示例网络中其他节点的交叉熵值, 其计算结果如表 2 所示。

表 3 记录了不同评估算法对示例网络中节点进行排序的结果, 可以看出, KSGC 和 GM 算法与本文所提出的 CE+算法具有相同的排名广度, 但其对首要节点的排序结果略有不同, 分别移除节点 4、5 和 14, 得出非连通子图的数量分别为 2、2 和 6, 显然移除节点 14 对网络破坏性更大。此外其余算法对网络外围节点的量化区分度较低, 过于粗粒化。因此 CE+算法对重要节点排序的精确性得到了

初步验证。

表 2 各节点的交叉熵值

节点	CE+值	节点	CE+值	节点	CE+值
1	1.201 8	10	0.138 6	19	0.085 1
2	1.080 3	11	0.138 6	20	2.776 0
3	1.253 9	12	0.556 0	21	0.095 9
4	0.973 7	13	0.556 0	22	0.095 9
5	2.189 5	14	6.137 6	23	1.362 0
6	1.791 8	15	0.177 4	24	0.282 4
7	0.231 0	16	0.085 1	25	0.770 2
8	0.231 0	17	0.085 1	26	0.346 6
9	3.614 2	18	0.085 1	27	0.244 8

表 3 不同评估算法排序结果

排名	DC	BC	LLS	KSGC	GM	ME	IKS	CE+值
1	14	14	14	4	5	5	1	14
2	5,9	4	9	2	4	4	4	9
3	1~4	2	5	3	2	2	7	20
4	6,20,23	3	4	1	14	1,3	10,12	5
5	12,13,15,25	9	2	5	3	9	3	6
6	others	1	1,3	9	1	14	2	23
7		23	23	12,13	9	23	6	3
8		15	6	23	23	15	9	1
9		6,20	20	15	6	12,13	5,8	2
10		5	15	14	15	6	11	4
11		25	25	6	12,13	25	13	25
12		others	others	27	20	20	14~17	12,13
13				16~19	25	16~19	18	26
14				20	27	10,11,27	19~21	24
15				25	16~19	7,8	22~26	27
16				10,11	10,11	21,22	27	7,8
17				24	24	24		15
18				7,8	7,8	26		10,11
19				21,22	21,22			21,22
20				26	26			16~19

3 实验数据集与评价指标

3.1 实验数据集

实验选用了 8 个不同领域的真实网络数据集, 分别是: 维基语录编辑网络 Wikiquote^[29]、金融网络 Economic^[30]、约束优化模型网络 BP^[31]、犯罪案件人物关系网络 Crime^[32]、大学生电子邮件网络 Email^[33]、空中交通管制网络 Traffic^[34]、智人细胞中的蛋白质网络 Proteins^[35] 以及青少年朋友关系网络 Adolescent^[36]。表 4 列出了各网络的拓扑统计参数, 其中 N 和 E 分别为网络的节点总数和连边总

数, D 表示网络的密度, k_{\max} 表示网络节点的最大度值, $\langle k \rangle$ 和 $\langle d \rangle$ 分别为网络节点的平均度值与平

均最短路径距离, β_{th} 和 β 分别为 SIR 模型传播阈值以及实际传播率。

表 4 8 个真实网络的拓扑统计特征及传播率

网络	N	E	D	k_{\max}	$\langle k \rangle$	$\langle d \rangle$	β_{th}	β
Wikiquote	25	37	0.123 3	5	1.480 0	1.609 2	0.168 9	0.17
Economic	260	2 942	0.082 7	110	21.697 0	2.532 0	0.044 1	0.05
BP	822	3 276	0.009 7	266	7.623 3	3.354 9	0.125 4	0.13
Crime	829	1 476	0.004 3	25	2.139 1	5.040 0	0.092 8	0.09
Email	1 133	5 451	0.008 5	71	9.622 2	3.606 0	0.053 5	0.05
Traffic	1 226	2 615	0.003 4	37	4.265 9	5.929 0	0.234 4	0.23
Proteins	2 239	6 452	0.002 5	314	5.763 2	3.978 6	0.173 5	0.17
Adolescent	2 539	12 969	0.004 0	36	10.215 8	4.516 4	0.097 8	0.10

3.2 评价指标

为验证 CE+算法的性能, 本文基于网络节点的单调关系、鲁棒性以及 SIR 传播动力学模型对节点重要性进行研究。首先采用单调性指标定量分析不同算法的分辨率, 其次选取极大连通系数以及网络效率指标来评估攻击节点前后网络结构和功能的变化, 最后在 SIR 模型上再进行传播仿真实验分析。

单调性指标^[23]是通过计算节点重要性排序中具有相同排名索引的节点比例来度量节点的评估效果, 定义为:

$$M(R) = \left[1 - \frac{\sum_{r \in R} n_r(n_r - 1)}{n(n - 1)} \right]^2 \quad (15)$$

式中, R 为经由节点重要性排序算法所得到的排名索引; n_r 表示具有相同排名索引的节点数量; $M(R) \in [0, 1]$, 若 $M(R) = 1$, 则该算法完全单调, 所有节点都具有不同的排名索引值。若 $M(R) = 0$, 则该算法无法区分, 每个节点的排名索引值相同。

极大连通系数^[24]常用于评价移除网络中的节点后对极大连通子集的影响, 表示为:

$$C = \frac{N_i^R}{N} \quad (16)$$

式中, N_i^R 为移除节点 i 后网络巨片的规模。网络巨片中包含的节点数若随攻击节点数的增多而急剧减少, 则说明该评估算法更准确。

网络效率^[25-26]常用于度量网络连通性的强弱, 计算公式为:

$$E = \frac{1}{N(N-1)} \sum_{1 \leq i < j \leq N} \frac{1}{d_{ij}} \quad (17)$$

攻击网络中一定比例的节点, 若网络效率下降趋势越明显, 则该算法的排序准确性越高。

SIR 模型^[27]常用于验证节点传播信息与病毒的能力, 在 SIR 模型中, 网络中的所有节点具有 3 种状态, 分别是易感状态 S、感染状态 I 以及免疫状态 R。病毒传播初始时, 除少数节点处于感染状态外, 其他节点均处于易感状态。每个时间步长里, 感染节点以 β 的概率尝试感染易感的邻居节点, 此外, 感染节点还以 μ 的概率尝试恢复为免疫节点, 为不失一般性, 本文设定恢复率 $\mu=1$ 。需要注意的是, 免疫节点不会被再次感染, 同样也不具有感染能力。当网络达到稳定时, 传播过程结束, 此时免疫节点的数量可用于量化初始感染节点的传播能力。

4 实验分析

本文选取了度中心性 (DC)、介数中心性 (BC) 等经典算法作为对比方法, 还选取了邻域相似度 (LLS)、引力模型 (GM)、改进的引力模型 (KSGC)、映射熵 (ME) 以及改进的 K-壳分解法 (IKS) 等近期提出的性能显著的排序算法进行比较, 记录 8 种排序算法在 8 个真实网络数据上各评估指标的实验结果。

4.1 单调性指标

表 5 记录了本文所提出的 CE+算法与其他评估算法的单调性指标, 加粗数值为最优值, 可以看出, CE+算法在大部分网络中均表现出较好的分辨率, 并在一些网络 (如 BP、Adolescent) 中达到了 1, 这说明 CE+算法是完全单调的, 网络中的每个节点都具有不同的排名索引值。此外, KSGC 和 GM 算法也表现出良好的区分度。

表 5 不同评估算法的单调性指标

网络	DC	BC	LLS	KSGC	GM	ME	IKS	CE+值
Wikiquote	0.480 7	0.513 6	0.413 9	0.804 0	0.804 0	0.768 5	0.768 5	0.804 0
Economic	0.830 0	0.977 4	0.975 9	0.999 0	0.998 9	0.994 6	0.994 8	0.999 4
BP	0.851 8	1	0.999 2	1	1	0.999 5	0.999 7	1
Crime	0.699 0	0.866 9	0.865 1	0.999 4	0.999 3	0.983 7	0.984 6	0.998 8
Email	0.887 4	0.940 0	0.940 0	0.999 9	0.999 9	0.998 9	0.999 0	0.999 9
Traffic	0.592 2	0.977 0	0.962 4	0.999 7	0.999 8	0.988 4	0.991 6	0.999 3
Proteins	0.592 8	0.623 3	0.620 1	0.993 7	0.993 8	0.991 2	0.991 3	0.994 2
Adolescent	0.867 7	0.997 5	0.997 4	1	1	0.999 7	0.999 8	1

4.2 极大连通系数

图 2 呈现了不同评估算法模拟蓄意攻击网络所造成极大连通系数变化的趋势, 其中横坐标为各评估算法排序下依次移除节点占节点总数的比例, 纵坐标为极大连通系数, 可以看出在 8 个真实网络中, 本文提出的 CE+ 算法均表现出更好的攻击效

果。在 BP 和 Adolescent 网络中, 各算法的前期破坏效果出现了高度重合, 这是因为网络的连边总数远高于节点总数, 导致节点间紧密连接, 而当攻击节点数达到一定比例时, CE+ 算法表现出了较其他算法更为明显的攻击效果。因此, CE+ 算法度量节点重要性的准确性得到了验证。

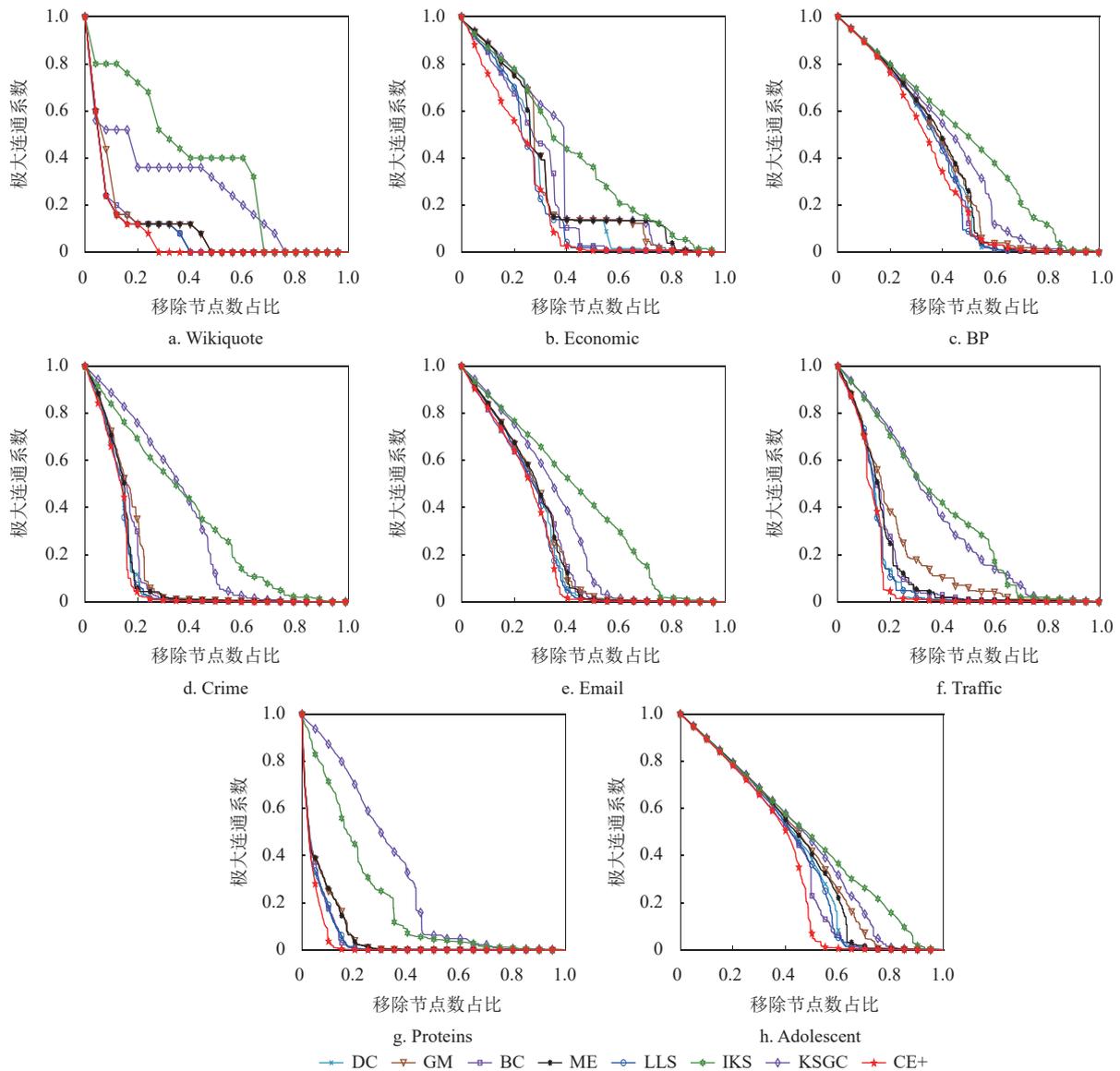


图 2 8 个网络在各类评估算法攻击下极大连通系数变化

4.3 网络效率

图 3 展现了利用不同的评估算法排序依次移除节点后，对网络效率所造成的变化趋势。可以看出，在 Wikiquote 网络、Economic 网络和 Adolescent 网络中，CE+算法相较于其他算法对网络的蓄意攻击效果更为明显，而在 Crime 网络与 Traffic 网络

中，KSGC 和 IKS 算法的攻击效果基本一致，ME 和 DC 算法分别在移除节点比例达到 16% 以及 18% 时，破坏效果最明显。此外，当攻击节点比例相同时，CE+算法的破坏曲线整体下降最快，这说明此时的网络连通性最差。因此，CE+算法度量节点重要性的准确性得到了进一步验证。

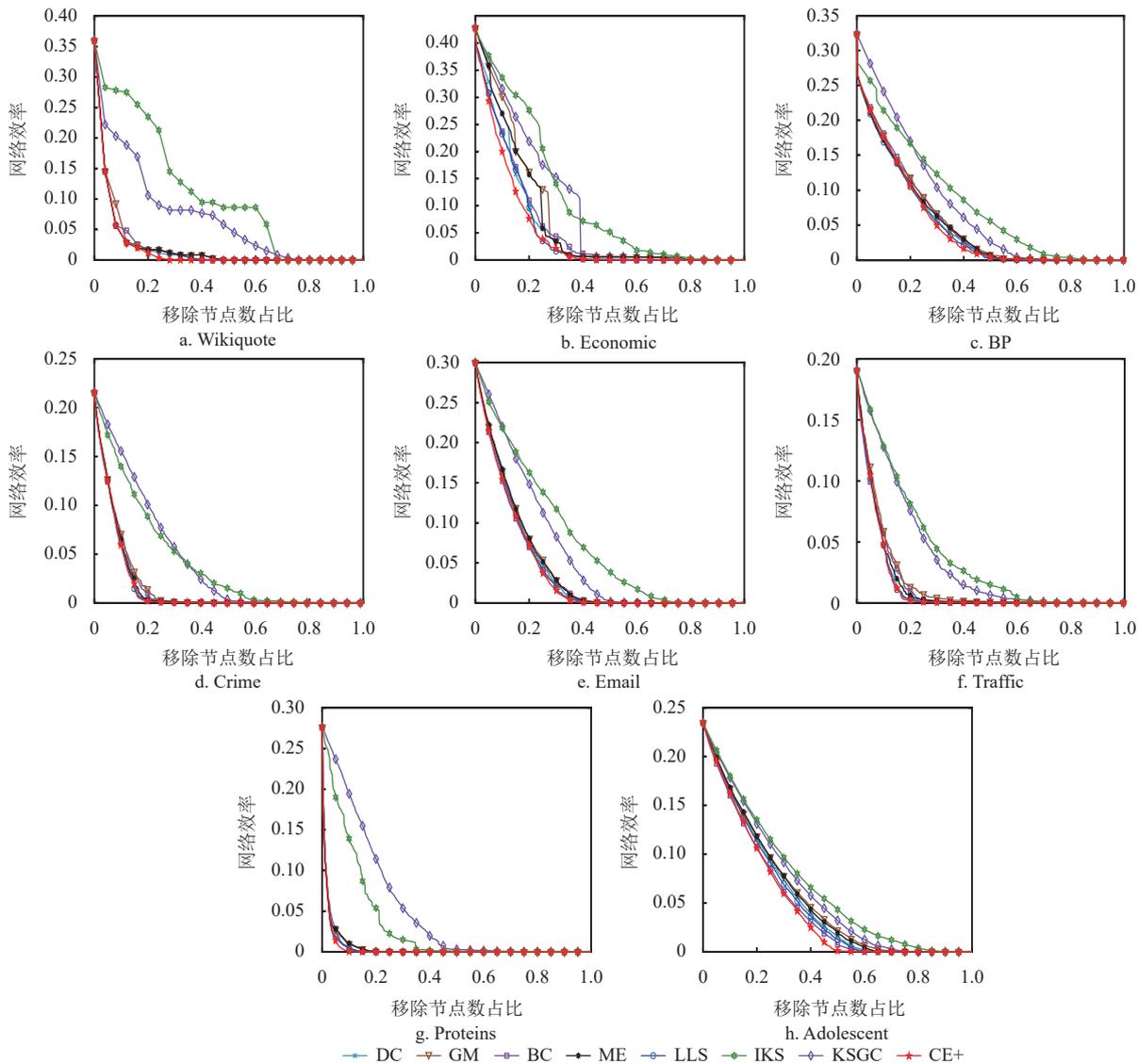


图 3 8 个网络在各类评估算法攻击下网络效率的变化

4.4 SIR 模型传播

为从不同角度评价 CE+算法的有效性和适用性，在 SIR 模型中再进行传播实验分析，利用各种评估算法排序前 10% 的节点作为初始感染节点，传播率阈值设定为 $\beta_{th} = \langle k \rangle / \langle k^2 \rangle$ 。其中 $\langle k \rangle$ 为平均度， $\langle k^2 \rangle$ 为二阶平均度。实际传播率基于阈值四舍五入并精确到百分位，具体取值如表 5 所示。为减小迭代过程随机的影响，本文选择对于 $N < 1000$ 的网络模拟迭代 1000 次，对于 $N > 1000$ 的网络模拟

迭代 100 次。

图 4 反映了各评估算法中重要性排名前 10% 的节点作为感染节点 N 随时间步长 t 的变化趋势，可以看出当时间步长达到 15 时，大部分网络趋于稳定，其中 CE+算法的传播曲线具有处于稳定状态时的最高高度以及最大斜率，这说明本文所提出的算法具有最广泛的传播范围以及最快传播速率。而在 BP 网络中，CE+算法的传播效果与其他算法的差异较小，是因为该网络的节点间分布密集，关联

程度较高, 信息具有易传播、易扩散的特点。整体来看, CE+算法相较于其他算法更能准确快速地挖掘出网络中影响能力较强的节点。

为比较不同比例的初始节点在传播达到稳态时的规模, 本文分别选取模拟迭代 1 000 次的 Crime 网络以及模拟迭代次数 100 次的 Traffic 网络, 并分别选取各种算法排名前 5%、15% 以及 20% 的节点作为感染节点进行传播验证实验, 实验结果分别

如图 5 和图 6 所示。结合图 5 中 Crime 和 Traffic 网络图可知, BC 算法曲线趋于稳态的高度仅在初始感染节点为 5% 时较高, IKS 算法曲线也仅在初始感染节点比例为 20% 时具有较大的斜率, 而 CE+算法在各种比例初始感染节点下均具有显著的传播范围以及感染速率, 可知 CE+算法在这两方面优于其他算法, 由此验证了 CE+算法的有效性以及适用性。

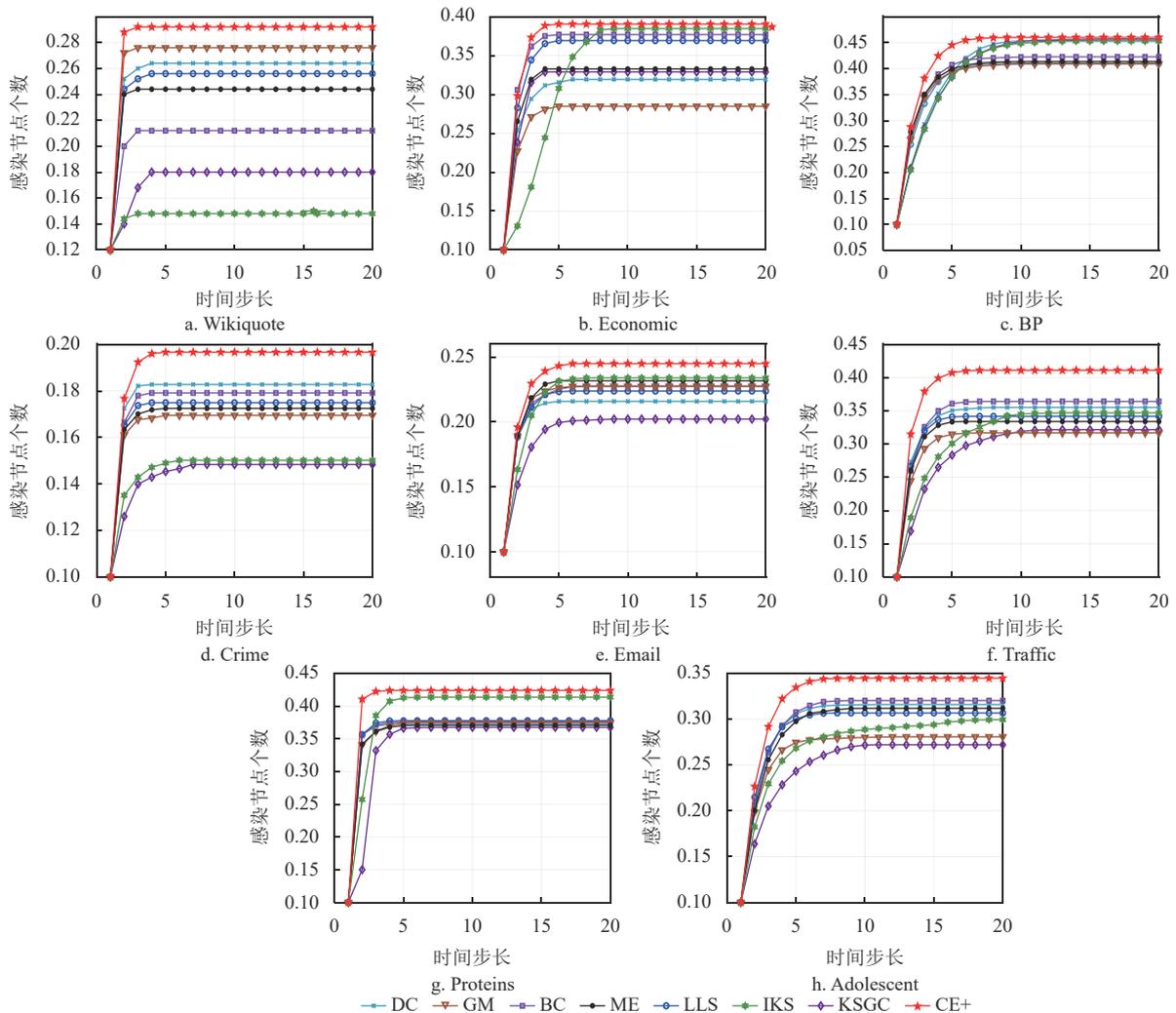


图 4 8 个网络在各类评估算法下感染节点的变化

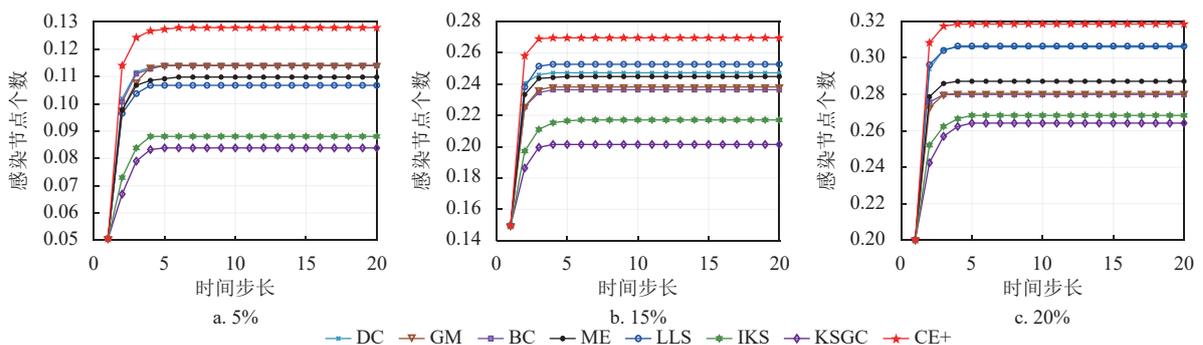


图 5 不同比例初始节点在 Crime 网络上感染节点的变化

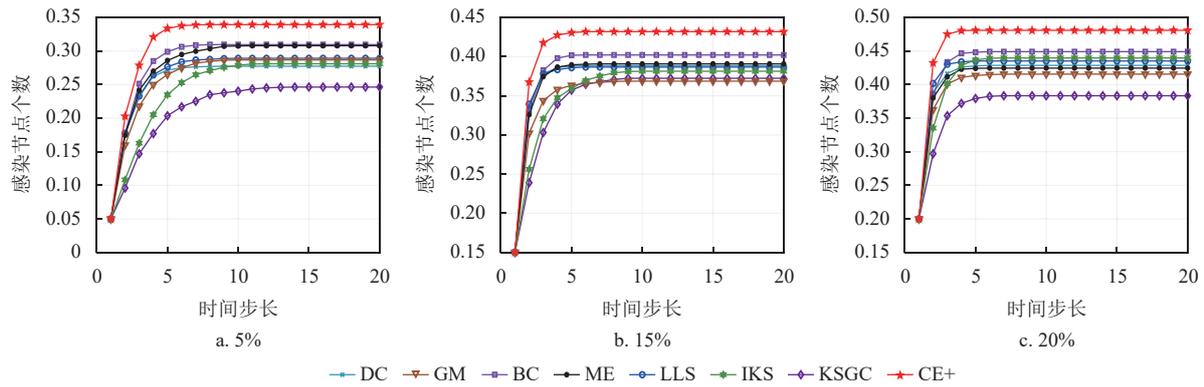


图 6 不同比例初始节点在 Traffic 网络上感染节点的变化

5 结束语

在网络科学研究中, 识别复杂网络中的核心关键节点有助于提高网络的鲁棒性和抗毁性。本文基于交叉熵提出了一种新的节点性能评价排序算法, 算法只需获取中心节点与其近邻节点的信息就可反映节点之间的差异性, 并且时间复杂度仅为 $O(n)$, 适用于刻画大规模复杂网络中节点的重要性。通过在 8 个不同领域的真实网络上进行单调关系、极大连通系数、网络效率以及 SIR 模型传播实验, 结果表明, 本文所提的 CE+算法对比 DC、BC、LLS、KSGC、GM、ME 和 IKS 算法具有更高效的性能。

参 考 文 献

- [1] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393(6684): 440-442.
- [2] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509-512.
- [3] 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.
REN X L, LYU L Y. Review of ranking nodes in complex networks[J]. *Science Bulletin*, 2014, 59(13): 1175-1197.
- [4] 朱军芳, 陈端兵, 周涛, 等. 网络科学中相对重要节点挖掘方法综述[J]. *电子科技大学学报*, 2019, 48(4): 595-603.
ZHU J F, CHEN D B, ZHOU T, et al. A survey on mining relatively important nodes in network science[J]. *Journal of University of Electronic Science and Technology of China*, 2019, 48(4): 595-603.
- [5] GUO Y, GUO C, YANG J. A tri-level optimization model for power systems defense considering cyber-physical interdependence[J]. *IET Generation, Transmission and Distribution*, 2023, 17(7): 1477-1490.
- [6] VIDAL M, CUSICK M E, BARABASI A L. Interactome networks and human disease[J]. *Cell*, 2011, 144(6): 986-998.
- [7] ROGERS T. Assessing node risk and vulnerability in epidemics on networks[J]. *Europhysics Letters*, 2015, 109(2): 28005.
- [8] FREEMAN L C. Segregation in social networks[J]. *Social network*, 1978, 6(4): 411-429.
- [9] FREEMAN L C. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977, 40(1): 35-41.
- [10] SABIDUSSI G. The centrality index of a graph[J]. *Psychometrika*, 1966, 31(4): 581-603.
- [11] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. *Nature Physics*, 2010, 6(11): 888-893.
- [12] BASARAS P, KATSAROS D, TASSIULAS L. Detecting influential spreaders in complex, dynamic networks[J]. *Computer*, 2013, 46(4): 24-29.
- [13] NIE T, 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.
- [14] 阮逸润, 老松杨, 王竣德, 等. 基于领域相似度的复杂网络节点重要度评估算法[J]. *物理学报*, 2017, 66(3): 371-379.
YUAN Y R, LAO S Y, WANG J D, et al. Node importance measurement based on neighborhood similarity in complex network[J]. *Acta Physica Sinica*, 2017, 66(3): 371-379.
- [15] WANG M, LI W, GUO Y, 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.
- [16] LI Z, HUANG X Y. Identifying influential spreaders in complex networks by an improved gravity model[J]. *Scientific Reports*, 2021, 11(1): 1-10.
- [17] YANG X, XIAO F. An improved gravity model to identify influential nodes in complex networks based on k-shell method[J]. *Knowledge-Based Systems*, 2021, 227: 107198.
- [18] SHANNON C E. Communication in the presence of noise[J]. *Proceedings of the IRE*, 1949, 37(1): 10-21.
- [19] FEI L, DENG Y. A new method to identify influential nodes based on relative entropy[J]. *Chaos, Solitons and Fractals*, 2017, 104: 257-267.
- [20] 汪亭亭, 梁宗文, 张若曦. 基于信息熵与迭代因子的复杂网络节点重要性评价方法[J]. *物理学报*, 2023, 72(4): 1-25.
WANG T T, LIANG Z W, ZHANG R X. Identifying

- influential nodes in complex networks using iteration factor and information entropy[J]. *Acta Physica Sinica*, 2023, 72(4): 1-25.
- [21] ZHANG Q, LI M Z, DENG Y. A new structure entropy of complex networks based on nonextensive statistical mechanics[J]. *International Journal of Modern Physics C*, 2016, 27(10): 440-452.
- [22] 杨松青, 蒋沅, 童天驰, 等. 基于 Tsallis 熵的复杂网络节点重要性评估方法[J]. *物理学报*, 2021, 70(21): 273-284.
YANG S Q, JIANG Y, TONG T C, et al. A method of evaluating importance of nodes in complex network based on Tsallis entropy[J]. *Acta Physica Sinica*, 2021, 70(21): 273-284.
- [23] 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.
- [24] DEREICH S, MORTERS P. Random networks with sublinear preferential attachment: The giant component[J]. *The Annals of Probability*, 2013, 41(1): 329-384.
- [25] VRAGOVIC I, LOUIS E, DIAZ-GUILERA A. Efficiency of informational transfer in regular and complex networks[J]. *Physical Review E*, 2005, 71(3): 036122.
- [26] LATORA V, MARCHIORI M. A measure of centrality based on network efficiency[J]. *New Journal of Physics*, 2007, 9(6): 188.
- [27] CASTELLANO C, PASTOR-SATORRAS R. Thresholds for epidemic spreading in networks[J]. *Physical Review Letters*, 2010, 105(21): 218701.
- [28] JAMIN A, HUMEAU-HEURTIER A. (Multiscale) cross-entropy methods: A review[J]. *Entropy*, 2019, 22(1): 45.
- [29] CHANDRA A, GARG H, MAITI A. A general growth model for online emerging user-object bipartite networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2019, 517: 370-384.
- [30] SARKAR D, ANDRIS C, CHAPMAN C A, et al. Metrics for characterizing network structure and node importance in spatial social networks[J]. *International Journal of Geographical Information Science*, 2019, 33(5): 1017-1039.
- [31] ZHANG L, WANG F, SUN T, et al. A constrained optimization method based on BP neural network[J]. *Neural Computing and Applications*, 2018, 29: 413-421.
- [32] TRONCOSO F, WEBER R. A novel approach to detect associations in criminal networks[J]. *Decision Support Systems*, 2020, 128: 113159.
- [33] GUIMERA R, DANON L, DIAZ-GUILERA A, et al. Self-similar community structure in a network of human interactions[J]. *Physical Review E*, 2003, 68(6): 065103.
- [34] PIEN K C, HAN K, SHANG W, et al. Robustness analysis of the European air traffic network[J]. *Transportmetrica A: Transport Science*, 2015, 11(9): 772-792.
- [35] GURSOY A, KESKIN O, NUSSINOV R. Topological properties of protein interaction networks from a structural perspective[J]. *Biochemical Society Transactions*, 2008, 36(6): 1398-1403.
- [36] HUNTER D R, GOODREAU S M, HANDCOCK M S. Goodness of fit of social network models[J]. *Journal of the American Statistical Association*, 2008, 103(481): 248-258.

编辑 张莉