# 基于传播概率熵的重要节点识别方法

# 叶梦梦

上海理工大学管理学院,上海

收稿日期: 2024年6月16日; 录用日期: 2024年7月9日; 发布日期: 2024年7月16日

# 摘要

本研究提出了一种新的方法ESPE (传播概率熵),旨在更有效地评估复杂网络中节点的重要性。该方法通 过结合信息熵同时考虑节点自身的传播影响力及其邻居节点的传播影响,解决了以往研究中忽略邻居节 点影响力的问题。为了验证ESPE算法的性能,采用肯德尔相关系数来衡量其识别节点传播能力的准确性, 通过与SIR(易感 - 感染 - 恢复)传播模型的比较,证明了ESPE在识别关键节点方面的优越性。实验结果 表明,ESPE算法在六个真实网络中平均提高了8.46%,其Kendall相关系数普遍超过现有方法。此外, 在分析ESPE和SPC算法获得的归一化传播影响力分数之间的相关性时,ESPE比SPC表现出更高的准确性 和判别力。总体而言,ESPE算法为复杂网络分析提供了准确可靠的工具,能够更有效地识别和评估网络 中的关键节点。

# 关键词

复杂网络,重要节点,传播概率,信息熵

# Identifying Vital Nodes Based on Epidemic Spreading Probability Entropy

# **Mengmeng Ye**

School of Business, University of Shanghai for Science and Technology, Shanghai

Received: Jun. 16<sup>th</sup>, 2024; accepted: Jul. 9<sup>th</sup>, 2024; published: Jul. 16<sup>th</sup>, 2024

#### Abstract

This study introduces a novel method known as ESPE (Epidemic Spreading Probability Entropy), designed to more effectively evaluate the importance of nodes within complex networks. The method addresses the oversight in previous research, which neglected the influence of neighboring nodes, by incorporating information entropy to simultaneously consider a node's own spreading influence and the spreading impact of its neighbors. To validate the performance of the ESPE algo-

rithm, Kendall's tau rank correlation coefficient was employed to measure its accuracy in identifying the spreading capabilities of nodes. The superiority of ESPE in identifying key nodes was demonstrated through comparison with the SIR (Susceptible-Infected-Recovered) spreading model. Experimental results indicate that the ESPE algorithm achieved an average improvement rate of 8.46% across six real-world networks, with its Kendall correlation coefficients generally exceeding those of existing methods. Furthermore, in the analysis of the correlation between the normalized spreading influence scores obtained by the ESPE and SPC algorithms, ESPE demonstrated greater accuracy and discriminative power compared to SPC. Overall, the ESPE algorithm provides an accurate and reliable tool for complex network analysis, capable of more effectively identifying and assessing key nodes within the network.

#### **Keywords**

Complex Networks, Vital Nodes, Spreading Probability, Information Entropy

Copyright © 2024 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <u>http://creativecommons.org/licenses/by/4.0/</u> © Open Access

# 1. 引言

网络中的关键节点在传播现象的研究中扮演着至关重要的角色[1],无论是在流行病学[2]、疾病传播 [3]还是信息传播[4]等方面。这些关键节点的识别和影响力评估对于理解和控制传播过程具有重要意义。 在社交网络中,关键节点往往是那些连接人数众多或能够影响大量人群的用户。这些节点在信息传播过 程中起到放大器的作用,能够加速信息的扩散速度和范围[5]。在疾病传播或谣言扩散的网络中,识别关 键节点有助于快速定位并采取措施,如隔离或辟谣,从而有效控制传播过程[6][7]。在电力网络中,关键 节点的稳定性对整个系统的可靠性至关重要。通过识别这些节点,可以优先加强其保护措施,提高整个 网络的抗干扰能力[8][9]。

目前,评估网络中节点影响力的方法众多,包括但不限于中心性指标,例如,度中心性衡量一个节 点的连接数[10],而接近中心性则衡量节点到其他所有节点的平均距离[11];介数中心性衡量了一个节点 在网络中所起到"桥梁"作用的大小,作用越强,即它在连接网络中其他节点对之间的路径上扮演着重 要的角色[12]; k-壳分解是一种将网络分解为不同层次的方法,其中 k-壳指的是在移除度最高的节点后, 剩下的网络中度为 k 的节点集合[13];特征向量中心性是基于网络中节点影响力的传递性,即一个节点的 影响力不仅取决于其直接连接的节点,还取决于其邻居节点的影响力[14]; PageRank 中心性是一种基于 随机游走者在网络中的停留时间来评估节点重要性的方法,常用于搜索引擎优化[15]等。

除此之外,还有基于传播动力学模型的方法,例如,Qu 等[16]则将邻接节点的感染风险考虑在内, 并提出了一种基于 SIS 模型的中心性方法。Jun 等[17]通过计算感染节点以一定概率感染其邻居节点的数 量,来量化网络中每个节点的传播能力。此外,一些研究者利用熵的概念来描述网络结构的复杂性。例 如,Chen 等[18]提出结构熵来度量网络的复杂性。张齐[19]提出一种基于非广延统计力学的结构熵来度量 网络的复杂性。胡钢等[20]通过一种邻接信息熵来识别复杂网络中的重要节点。吴英晗等[21]提出了一种 利用节点传播熵识别超网络中重要节点的方法。尽管这些方法在识别关键节点方面取得了一定成果,但 它们往往忽略了节点的局部和全局结构信息的综合考量,且在细微区分不同节点方面存在局限。为了解 决这些问题,本研究提出了一种新的方法,该方法结合了传播动力学模型和信息熵,旨在更全面地评估 网络中节点的传播能力。通过这种方法,我们不仅考虑了节点的局部传播特性,还整合了其邻居节点的 传播能力,从而实现了对网络中关键节点的更准确识别和区分。

#### 2. 相关知识

#### 2.1. 复杂网络定义

在研究复杂网络时,通常采用图论的方法来构建模型。具体来说,一个网络可以通过一个图 *G* 来表示,该图由两部分组成:节点集 *V* 和边集 *E*。在这个模型中,*V* 包含了网络中的所有节点,记为  $V = \{v_1, v_2, v_3, \dots, v_{|V|}\}$ ,而 *E*则包含了网络中所有的连接,记为  $E = \{e_1, e_2, e_3, \dots, e_{|E|}\}$ 。为了便于分析,网络的连接关系可以通过一个邻接矩阵  $A = a_{ij} \in V \times V$  来表示。在这个矩阵中,如果  $a_{ij}$  的值为 1,则表示节点  $v_i$  与节点  $v_j$ 之间存在直接的连接;如果  $a_{ij}$  的值为 0,则表示这两个节点之间没有直接的连接。本研究中,我们假设所有考虑的网络都是无向的,即边是双向的,没有指定的方向。

#### 2.2. 度中心性

度中心性[10]定义了节点的一阶邻居数作为其传播影响的度量,表示为:

$$DC(i) = \frac{k(i)}{N-1} \tag{1}$$

其中 k(i)表示节点 i 的度, N 是网络中节点的总数。

#### 2.3. 介数中心性

介数中心性[12]定义了作为两个不相连群体之间桥梁角色的节点作为传播影响节点,表示为:

$$BC(i) = \sum_{u \neq i \neq v} \frac{g'_{uv}}{g_{uv}}$$
(2)

其中  $g_{uv}$  表示节点 v 和节点 u 之间的最短路径数,  $g_{uv}^i$  表示通过节点 i 的节点 v 和节点 u 之间的最短路径数。

#### 2.4. K-Shell

*k*-核值[13]能够反映节点的位置。具体而言,*k*-核值根据节点的度递归地移除节点,将节点分割成不同的层次来获得。

#### 2.5. SPC 模型

SPC 算法[17] 是一种基于信息传播和感染动态来识别复杂网络中具有影响力的节点的模型。表示为:

$$SPC(i) = \sum_{j \in R_i^3} \alpha^{d_{ij}}$$
(3)

其中,  $R_i^3$ 表示节点 *i* 的三阶邻居节点集合,  $d_{ij}$ 表示节点 *i* 和节点 *j* 之间的最短路径长度,  $\alpha$  是一个可调 参数, 代表信息从节点 *i* 到节点 *j* 的传播概率。

#### 2.6. ESPE 算法的提出

在网络科学和复杂系统的研究中,用户影响通常被定义为一个用户促使其他用户采取特定行为的能力。在本文中,我们依据易感者-感染者(SI)模型,将一个节点作为感染源并衡量其感染网络中其他未感

染节点的能力,以此作为该节点影响力的量化指标。具体而言,节点的传播能力可以通过计算该节点感 染未感染邻居节点的概率来确定。然而,一个节点的影响力并非仅由其自身的属性决定,还受到其邻居 节点影响力的影响。因此,一个节点的总体影响力应该是由其自身的影响力和其邻居节点影响力的综合 效应所共同决定的。为了准确评估这种复合影响力,本文通过结合信息熵,同时考虑节点自身的传播影 响力及其邻居节点的传播影响,提出了一种传播概率熵(Epidemic Spreading Probability Entrop, ESPE)的模 型,该模型不仅考虑了节点的直接传播能力,还考虑了其邻居节点的传播潜力。

ESPE 模型的实施步骤可以概括如下:

1) **节点状态初始化:** 首先,需要确定网络中每个节点的当前状态,这可以是易感者、感染者或传播者。

2) 识别三阶邻居节点集:对于目标节点 i, 识别其所有直接相连的邻居节点, 即三阶邻居节点集合 R<sub>i</sub><sup>3</sup>。

3) 计算单个节点的传播概率:对于目标节点 *i* 的每一个三阶邻居节点 *j*,计算节点 *i* 传播给节点 *j* 的 概率 *SP*(*i*)。这通常基于节点之间的连接强度和网络的拓扑结构。

$$SP(i) = \sum_{j \in R_i^3} \alpha^{d_{ij}} \tag{4}$$

其中, α 是传播概率的衰减因子, d<sub>ii</sub>是节点 i 到其三阶邻居 j 之间的距离。

4) 计算节点的总传播概率:累加节点 i 的所有一阶邻居节点的 SP 值,得到 TSP(i):

$$TSP(i) = \sum_{j \in R_i^1} SP(i)$$
<sup>(5)</sup>

**5) 确定信息传播的概率分布:**使用 *SP*(*i*)和*TSP*(*i*)来确定信息从节点 *i* 传播到其每一个一阶邻居节 点 *j* 的概率 *p*(*j*):

$$p(j) = \frac{SP(j)}{TSP(i)} \tag{6}$$

6) 计算传播影响概率熵: 最后,计算目标节点 *i* 的传播影响概率熵 *ESPE*(*i*),这反映了节点 *i* 对其 邻居节点影响力的不确定性或信息的混乱程度:

$$ESPE(i) = -\sum_{j \in R_i^1} p(j) \times \ln\left[p(j)\right]$$
(7)

# 3. 实验分析

### 3.1. 数据集

为了全面地评估 ESPE 算法的性能,本文使用以下六个真实网络,这些网络相关的统计特征如表 1 所示。1) LastFM [22]: 该网络由亚洲地区的 LastFM 用户组成,用户之间的连接关系基于相互的追随关系。2) Jazz [23]: 此网络展现了爵士音乐家之间的合作关系,节点代表音乐家,边代表他们之间的合作。 3) Email [24]: 该网络基于西班牙塔拉戈纳的 Rovirai Virgili 大学(URV)的内部电子邮件通信数据构建,其中节点代表组织成员,包括员工、研究人员等,边代表他们之间的电子邮件往来。4) Faa [25]: 是 https://networkrepository.com/网站上的一个网络。5) Facebook [26]: 这个网络由 Facebook 用户及其朋友构成,每个节点代表一个用户,节点间的连接表示朋友关系。6) Vote [27]: 该网络由维基百科用户之间的 投票关系构成,节点代表用户,有向边代表用户之间的投票行为。在本研究中,这些有向边被处理为无向边以简化分析。

# 3.2. 评价标准

#### 3.2.1. SIR 模型

SIR 模型[28]是流行病学中用于描述传染病传播的典型模型,它将人群分为三个主要状态:易感者 (Susceptible, S)、感染者(Infected, I)和康复者(Recovered, R)。在该模型中,每个感染节点以概率 *h* 将其感 染状态传播给其易感的直接邻居,同时以概率 *b* 恢复并退出感染状态。恢复后的节点不会再被感染。这 个过程会持续进行,直到网络中没有新的感染发生,达到一个稳定状态。

为了评估节点的感染能力,我们首先将特定节点x设定为初始感染状态,而其他所有节点设为易感状态。通过模拟网络达到稳定状态的过程,我们可以统计出在最终状态下的感染节点数和恢复节点数。节点x的感染能力,即其标签,可以通过 1000 次独立实验的平均结果来确定,以提高评估的准确性并减少随机波动的影响。感染能力的计算公式为:

$$IC(x) = \frac{\sum_{j=1}^{1000} N_j^i + N_j^r}{1000 \times N}$$
(8)

其中, IC(x)表示节点 x 的感染能力,  $N_i^i \oplus N_i^r$ 表示在第 j 个实验中感染节点和恢复节点的数量。

此外,为了量化网络中感染传播的难易程度,我们引入了感染传播阈值  $\beta_{th}$ ,该阈值可以通过平均场 理论来估算。具体计算公式如下:

$$\beta_{th} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \tag{9}$$

其中, k 表示节点的度, 〈·〉表示计算平均值。

通过这种方法,SIR 模型不仅能够模拟传染病的传播过程,还可以用于评估个体节点在网络中的感染能力,以及整个网络对疾病传播的抵抗性。这些信息对于设计有效的疾病控制策略和提高网络的鲁棒 性具有重要意义。

Table	<b>1.</b> Statistical	attributes	of six	real-world	networks
表 1.	六个真实网络	各的统计	属性		

Network	п	т	$\langle k \rangle$	$k_{\rm max}$	с	$G_{_{C_V}}$ %	$G_{\scriptscriptstyle C_E}$ %	d	$eta_{\scriptscriptstyle th}$
LastFM	7624	27806	7.294	216	0.219	100.000	100.000	0.001	0.041
Jazz	198	2742	27.697	100	0.617	100.000	100.000	0.141	0.027
Email	1133	5451	9.622	71	0.220	100.000	100.000	0.009	0.057
Faa	1226	2410	3.931	34	0.068	100.000	100.000	0.003	0.157
Facebook	4039	88234	43.691	1045	0.606	100.000	100.000	0.011	0.009
Vote	889	2914	6.556	102	0.153	100.000	100.000	0.007	0.058

其中  $n \to m$ 分别是网络中的节点总数和边总数,  $\langle k \rangle$ 是网络的平均度,  $k_{max}$ 是网络的最大度, c是网络的平均聚类系数,  $G_{c_v}$ %和 $G_{c_e}$ %分别表示网络中最大组件中节点和边占总节点数和总边数的比例, d是网络的密度,  $\beta_h$ 是网络的传播阈值。若网络的 $G_{c_v}$ % ≈100 且 $G_{c_e}$ % ≈100,则它是一个连通网络,否则它是一个非连通网络。

#### 3.2.2. 肯德尔相关性系数

在本文中,我们采用了肯德尔相关系数(Kendall's tau coefficient) [29]来衡量两组排名之间的关联程

度。肯德尔相关系数的计算公式如下:

$$\tau = \frac{2\left(N_c - N_d\right)}{N_L\left(N_L - 1\right)} \tag{10}$$

在这个公式中,  $N_L$ 是排名列表中的节点数量,  $N_c n N_d$ 分别是一致对和不一致对的数量。具体来说, 对于两个数据对 $(x_i, y_i)$ 和 $(x_j, y_j)$ , 如果  $x_i > x_j \pm y_i > y_j$ , 或  $x_i < x_j \pm y_i < y_j$ , 则这一对数据 $(x_i, y_i)$ 和 $(x_j, y_j)$ 是一致的。如果  $x_i > x_j \pm y_i < y_j$ , 或  $x_i < x_j \pm y_i > y_j$ , 则这一对数据 $(x_i, y_i)$ 和 $(x_j, y_j)$ 是不一致的。如果  $x_i = x_i$ ,  $y_i = y_i$ , 这对既不是一致的也不是不一致的。

肯德尔相关系数的取值范围从-1到1。当 $\tau$ 的值越接近1,表明两个排名列表的一致性越高。如果 $\tau$ =1,则表示两个排名完全一致。当 $\tau$ =0,表明排名列表之间没有显著的相关性。而当 $\tau$ =-1,则意味着两个排名列表完全相反。

在本文的第 3.3.1 节中,我们将利用 τ 来评估由 SIR 模型产生的排名与通过 ESPE 模型以及其他方法 得到的排名之间的关联性。通过这种方法,我们可以定量地分析不同排名方法之间的一致性和差异性, 为进一步的研究提供依据。

#### 3.2.3. 改进率

该性能指标旨在衡量所提出方法相较于其他现有方法在识别性能上的提升幅度。改进率是基于所提 方法的识别准确率与对比方法的识别准确率之间的差异百分比来计算的。具体来说,这个指标反映了所 提方法在识别准确率上相对于其他方法的增长或减少的幅度。当这个百分比值为正时,意味着所提方法 在性能上优于对比方法;若为负,则意味着性能有所下降。此指标是评估新方法在竞争性方法面前性能 优势的重要工具。

$$\eta\% = \frac{ESPE(\tau) - SPC(\tau)}{SPC(\tau)} \times 100 \tag{11}$$

其中, *ESPE*(*i*)表示所提出方法的肯德尔相关性系数,而 *SPC*(*i*)表示 SPC 算法的肯德尔相关性系数。 通过这个公式,我们可以定量地评估所提出方法在识别任务上相对于 SPC 算法的改进情况,进而为选择 更合适的方法提供决策依据。

#### 3.3. 实验结果

#### 3.3.1. 算法的节点排序能力

从图 1 的分析结果可以观察到, ESPE 算法实现了较高的肯德尔相关系数,这揭示了 ESPE 算法生成 的节点排名与实际 SIR 传播模型的一致性较高。本研究进一步通过将 SPC 算法设定为对照组,来评估 ESPE 算法在肯德尔相关系数上相较于 SPC 算法的改进程度。

图 2 的数据揭示了 ESPE 算法在六个真实网络中都展现了正向的改进,意味着其性能超越了 SPC 算法。具体来看,在 Vote 网络中,ESPE 算法的平均改进率达到了 5.98%,在 Jazz 网络中达到了 9.80%,在 LastFM 网络中为 4.96%,在 Email 网络中为 5.88%,在 Facebook 网络中显著达到了 18.77%,在 Faa 网络中则为 5.38%。这些结果表明,ESPE 算法在识别网络中关键节点的准确性方面具有优势,能够更有效地捕捉节点间的复杂关系,从而为关键节点的识别提供了更加精确的方法。

综上所述, ESPE 算法在多个真实网络结构上的表现均优于 SPC 算法, 尤其在 Facebook 网络中的改进表现尤为突出。这证明了 ESPE 算法在网络分析中的潜力和实用性, 尤其是在需要高精度识别网络关键节点的应用场景中。



Figure 1. Comparative analysis of node rankings generated by centrality measures such as ESPC and SIR: A Kendall's tau correlation approach

图 1. 使用 ESPC 等中心性方法产生的节点排名与 SIR 产生的节点排名进行比较得到的肯德尔相关系数



**Figure 2.** Improvement rate of the ESPC algorithm over the SPC algorithm 图 2. ESPC 算法相对于 SPC 算法的改进率

 Table 2. Top 10 vidal nodes identified by the ESPE method and other approaches across various networks

 表 2. 使用 ESPE 方法和不同网络中的其他方法计算出的前 10 个重要节点

LastFM						Email					
SIR	ESPE	SPC	DC	KS	BC	SIR	ESPE	SPC	DC	KS	BC
7237	7237	7237	7237	6105	7199	105	105	105	105	299	333
3240	3530	3530	3530	7237	7237	42	42	333	333	389	105

续表											
378	3450	3450	4785	1334	2854	333	333	23	16	434	23
763	3597	2854	524	951	4356	16	23	42	23	552	578
3597	3240	7199	3450	3165	6101	3	41	41	42	571	76
3544	4900	5127	2510	3544	5454	23	16	76	41	726	233
4900	3544	6101	3597	4900	4338	196	3	233	196	756	135
2083	763	3544	2854	7075	5127	41	196	52	233	788	41
1334	5127	4900	6101	5854	3450	128	21	135	21	885	355
2734	290	4785	5127	1381	4785	49	233	378	76	886	42
Facebook								Ja	ZZ		
SIR	ESPE	SPC	DC	KS	BC	SIR	ESPE	SPC	DC	KS	BC
1912	107	107	107	1912	107	60	136	136	136	35	136
2347	1684	1684	1684	2543	1684	136	60	60	60	60	153
2543	1912	1912	1912	2266	3437	132	132	168	132	98	60
2073	1888	0	3437	2347	1912	168	168	132	168	99	149
2206	1800	3437	0	2542	1085	108	70	70	70	100	168
2507	1663	58	2543	2468	0	70	108	83	99	101	167
1943	1352	428	2347	1917	698	99	83	99	108	108	189
2266	1730	563	1888	1918	567	122	99	158	83	131	115
2233	1431	1577	1800	1929	58	7	194	108	158	132	96
2410	2543	483	1663	1938	428	194	131	194	7	154	83
Vote						Faa					
SIR	ESPE	SPC	DC	KS	BC	SIR	ESPE	SPC	DC	KS	BC
273	431	273	431	447	273	47	52	68	312	1	68
431	273	431	273	273	431	68	68	52	52	2	52
204	204	204	170	536	170	52	113	148	68	3	213
399	536	536	536	482	204	116	47	44	89	4	312
536	399	170	399	204	736	46	44	113	113	6	135
416	416	399	204	431	550	109	116	47	187	10	136
448	550	550	550	562	21	113	34	135	47	14	212
504	170	416	416	132	230	148	148	312	44	16	660
550	448	736	736	392	22	124	109	89	109	18	523
132	736	448	762	399	536	44	110	110	135	20	221

# 3.3.2. 算法在识别头部节点的传播性能

为了更清楚地验证算法在识别网络中关键节点传播能力方面的性能,本研究通过表2对不同网络中 被识别为最重要的前10个节点进行了比较分析。使用 ESPE 算法与其他基线方法进行对比,结果如下。 对于 Facebook 网络, ESPE 算法的识别结果与其他算法相比略逊一筹。然而在其他5个真实网络中, 本文所提出的算法在所有比较的基线方法中与SIR模型匹配的节点数量均是最高的。具体而言,在LastFM 网络中, ESPE 算法与 SIR 模型识别出的前 10 个节点有 6 个是重合的。在 Email 网络中, ESPE 算法与 SIR 模型匹配的节点数量为 8 个。在 Jazz 网络、Vote 网络和 Faa 网络中, ESPE 算法与 SIR 模型一致的 节点数量均为 8 个。

综合考虑,与其它基线方法相比,ESPE 算法在识别网络中最关键的前 10 个节点方面整体表现更为 优秀。这一结果表明 ESPE 算法在捕捉网络中节点的重要性和传播影响力方面具有较高的准确性和可靠 性,尤其是在 Email 网络、Jazz 网络、Vote 网络和 Faa 网络中,ESPE 算法展现出了其在识别头部节点方 面的显著优势。

#### 3.3.3. ESPE 与 SPC 算法相关性对比

为了评估 ESPE 算法与 SPC 算法在区分网络节点传播影响力上的能力,图 3 展示了两种算法得到的 排名分数之间的相关性分析。在此图中,X 轴代表 ESPE 算法计算得到的归一化传播影响力,而 Y 轴代 表 SPC 算法得出的相应数据。每个数据点的颜色基于 SIR 模型模拟得到的节点传播影响力而定,这使得 我们可以通过观察点的分布和颜色来评估两种算法的识别效率。





图 3. ESPC 算法和 SPC 算法对 6 个真实网络的归一化传播影响得分之间的相关性

从图 3 的观察结果可以得出,在所有六个真实世界的网络数据集上,ESPE 与 SPC 算法的排名分数 呈现出正相关性。特别地,在 LastFM 和 Facebook 网络中,ESPE 算法为具有较高传播影响力的红色节点 赋予了更高的评分,而在蓝色节点上,SPC 算法则给出了相对较高的评分。这一现象表明 ESPE 算法在 辨别传播影响力节点的准确性和分辨率方面优于 SPC 算法。

这一发现强调了 ESPE 算法在识别网络中具有显著传播能力节点方面的优越性,尤其是在区分节点 传播影响力的细微差别时。ESPE 算法通过更精确地量化节点的传播潜力,为网络分析提供了一种改进的 方法,这在复杂网络的结构和行为分析中具有重要的应用价值。

#### 4. 结论

本文提出了一种新颖的评估复杂网络中节点重要性的方法,即传播概率熵(ESPE)。与之前的 SPC 算 法相比,ESPE 方法在评估节点传播影响力时,不仅考虑了节点自身的传播潜力,还综合了其邻居节点的 传播影响,通过信息熵的概念提供了一种更全面的分析框架。为了验证 ESPE 方法的有效性,本研究采 用了肯德尔相关系数来衡量算法的准确性。模拟实验的结果表明,ESPE 算法在识别节点传播能力方面得 到的肯德尔相关系数普遍高于其他现有方法。特别是在确定网络中最重要的前 10 个节点时,ESPE 算法 与 SIR 传播模型的一致性更高,六个不同网络中的平均改进率达到了 8.46%,这进一步证实了 ESPE 方 法的优越性。

此外,通过对比 ESPE 算法和 SPC 算法得到的归一化传播影响得分的相关性分析, ESPE 算法展现 了更高的准确性和区分度。这一发现表明, ESPE 算法在识别网络中具有显著传播影响力的节点方面,能 够提供更为精确的评估。综上所述, ESPE 算法在复杂网络节点重要性评估领域提供了一种新的有效工具。 它不仅提高了关键节点识别的准确性,还增强了对这些节点的区分能力,这对于网络分析和相关领域的 研究具有重要的理论和实践意义。

尽管 ESPE 算法在本研究中表现出了显著的优势,未来的工作仍需在更广泛的网络类型和更复杂的 传播场景中进一步验证其性能。此外,算法的计算效率和可扩展性也是未来研究的重要方向。我们期待 ESPE 算法能够在网络科学、信息传播、疾病控制等领域发挥更大的作用,并为相关领域的研究者提供强 有力的分析工具。

### 参考文献

- Dorogovtsev, S.N., Goltsev, A.V. and Mendes, J.F.F. (2008) Critical Phenomena in Complex Networks. Reviews of Modern Physics, 80, 1275-1335. <u>https://doi.org/10.1103/revmodphys.80.1275</u>
- [2] Moreno, Y., Pastor-Satorras, R. and Vespignani, A. (2002) Epidemic Outbreaks in Complex Heterogeneous Networks. *The European Physical Journal B*, **26**, 521-529. <u>https://doi.org/10.1140/epjb/e20020122</u>
- [3] Wang, Z., Andrews, M.A., Wu, Z., Wang, L. and Bauch, C.T. (2015) Coupled Disease-behavior Dynamics on Complex Networks: A Review. *Physics of Life Reviews*, 15, 1-29. <u>https://doi.org/10.1016/j.plrev.2015.07.006</u>
- [4] Davis, J.T., Perra, N., Zhang, Q., Moreno, Y. and Vespignani, A. (2020) Phase Transitions in Information Spreading on Structured Populations. *Nature Physics*, 16, 590-596. <u>https://doi.org/10.1038/s41567-020-0810-3</u>
- [5] Yang, L., Li, Z. and Giua, A. (2020) Containment of Rumor Spread in Complex Social Networks. *Information Sciences*, 506, 113-130. <u>https://doi.org/10.1016/j.ins.2019.07.055</u>
- [6] Velásquez-Rojas, F., Ventura, P.C., Connaughton, C., Moreno, Y., Rodrigues, F.A. and Vazquez, F. (2020) Disease and Information Spreading at Different Speeds in Multiplex Networks. *Physical Review E*, **102**, Article ID: 022312. <u>https://doi.org/10.1103/physreve.102.022312</u>
- [7] Wu, Q. and Chen, S. (2020) Spreading of Two Interacting Diseases in Multiplex Networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **30**, Article ID: 073115. <u>https://doi.org/10.1063/5.0009588</u>
- [8] Kaiser, F., Latora, V. and Witthaut, D. (2021) Network Isolators Inhibit Failure Spreading in Complex Networks. Nature Communications, 12, Article No. 3143. <u>https://doi.org/10.1038/s41467-021-23292-9</u>
- [9] Motter, A.E. (2004) Cascade Control and Defense in Complex Networks. *Physical Review Letters*, 93, Article ID: 098701. <u>https://doi.org/10.1103/physrevlett.93.098701</u>
- [10] Freeman, L.C. (1978) Centrality in Social Networks Conceptual Clarification. Social Networks, 1, 215-239. <u>https://doi.org/10.1016/0378-8733(78)90021-7</u>
- [11] Sabidussi, G. (1966) The Centrality Index of a Graph. Psychometrika, 31, 581-603. https://doi.org/10.1007/bf02289527
- [12] Freeman, L.C. (1977) A Set of Measures of Centrality Based on Betweenness. Sociometry, 40, 35-41. <u>https://doi.org/10.2307/3033543</u>
- [13] Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., et al. (2010) Identification of Influential Spreaders in Complex Networks. *Nature Physics*, 6, 888-893. <u>https://doi.org/10.1038/nphys1746</u>
- [14] Bonacich, P. (1972) Factoring and Weighting Approaches to Status Scores and Clique Identification. The Journal of

Mathematical Sociology, 2, 113-120. https://doi.org/10.1080/0022250x.1972.9989806

- [15] Brin, S. and Page, L. (2012) Reprint Of: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks, 56, 3825-3833. <u>https://doi.org/10.1016/j.comnet.2012.10.007</u>
- [16] Qu, J., Tang, M., Liu, Y. and Guan, S. (2020) Identifying Influential Spreaders in Reversible Process. *Chaos, Solitons & Fractals*, 140, Article ID: 110197. <u>https://doi.org/10.1016/j.chaos.2020.110197</u>
- [17] Ai, J., He, T., Su, Z. and Shang, L. (2022) Identifying Influential Nodes in Complex Networks Based on Spreading Probability. *Chaos, Solitons & Fractals*, 164, Article ID: 112627. <u>https://doi.org/10.1016/j.chaos.2022.112627</u>
- [18] Chen, D., Lü, L., Shang, M., Zhang, Y. and Zhou, T. (2012) Identifying Influential Nodes in Complex Networks. *Physica A: Statistical Mechanics and Its Applications*, **391**, 1777-1787. <u>https://doi.org/10.1016/j.physa.2011.09.017</u>
- [19] 张齐. 基于熵的复杂网络结构特性研究[D]: [硕士学位论文]. 重庆: 西南大学, 2017.
- [20] 胡钢, 徐翔, 高浩, 等. 基于邻接信息熵的网络节点重要性识别算法[J]. 系统工程理论与实践, 2020, 40(3): 714-725.
- [21] 吴英晗,田阔,李明达,等.利用节点传播熵识别超网络重要节点[J].计算机工程与应用,2023,59(19):66-74.
- [22] Rozemberczki, B. and Sarkar, R. (2020) Characteristic Functions on Graphs: Birds of a Feather, from Statistical Descriptors to Parametric Models. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 19-23 October 2020, 1325-1334. <u>https://doi.org/10.1145/3340531.3411866</u>
- [23] Gleiser, P.M. and Danon, L. (2003) Community Structure in Jazz. Advances in Complex Systems, 6, 565-573. https://doi.org/10.1142/s0219525903001067
- [24] Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F. and Arenas, A. (2003) Self-Similar Community Structure in a Network of Human Interactions. *Physical Review E*, 68, Article ID: 065103. https://doi.org/10.1103/physreve.68.065103
- [25] Rossi, R.A. and Ahmed, N.K. (2016) An Interactive Data Repository with Visual Analytics. ACM SIGKDD Explorations Newsletter, 17, 37-41. <u>https://doi.org/10.1145/2897350.2897355</u>
- [26] McAuley, J. and Leskovec, J. (2012) Learning to Discover Social Circles in Ego Networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, 3-8 December 2012, 539-547.
- [27] Leskovec, J., Huttenlocher, D. and Kleinberg, J. (2010). Signed Networks in Social Media. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, 10-15 April 2010, 1361-1370. https://doi.org/10.1145/1753326.1753532
- [28] Hethcote, H.W. (2000) The Mathematics of Infectious Diseases. SIAM Review, 42, 599-653. <u>https://doi.org/10.1137/s0036144500371907</u>
- [29] Knight, W.R. (1966) A Computer Method for Calculating Kendall's Tau with Ungrouped Data. Journal of the American Statistical Association, 61, 436-439. <u>https://doi.org/10.1080/01621459.1966.10480879</u>