

识别高阶网络传播中最有影响力的节点*

李江¹⁾ 刘影^{1)2)†} 王伟³⁾ 周涛⁴⁾

1) (西南石油大学计算机科学学院, 成都 610500)

2) (弗里堡大学物理系, 弗里堡 1700)

3) (重庆医科大学公共卫生学院, 重庆 400016)

4) (电子科技大学大数据研究中心, 成都 611731)

(2023年9月1日收到; 2023年10月31日收到修改稿)

识别网络传播中最有影响力的节点是控制传播速度和范围的重要步骤, 有助于加速有益信息扩散, 抑制流行病、谣言和虚假信息的传播等. 已有研究主要基于描述点对交互的低阶复杂网络. 然而, 现实中个体间的交互不仅发生在点对之间, 也发生在3个及以上节点形成的群体中. 群体交互可利用高阶网络来刻画, 如单纯复形与超图. 本文研究单纯复形上最有影响力的传播者识别方法. 首先, 提出单纯复形上易感-感染-恢复(SIR)微观马尔可夫链方程组, 定量刻画单纯复形上的疾病传播动力学. 接下来利用微观马尔可夫链方程组计算传播动力学中节点被感染的概率. 基于网络结构与传播过程, 定义节点的传播中心性, 用于排序节点传播影响力. 在两类合成单纯复形与4个真实单纯复形上的仿真结果表明, 相比于现有高阶网络中心性和复杂网络中最优的中心性指标, 本文提出的传播中心性能更准确地识别高阶网络中最有影响力的传播者.

关键词: 高阶网络, 单纯复形, 节点影响力排序, 微观马尔可夫链, 复杂网络

PACS: 89.75.Fb, 87.23.Ge, 89.20.Ff

DOI: 10.7498/aps.73.20231416

1 引言

现实世界中, 许多具有扩散效应但仅由局部相互作用驱动的动力学过程可以用复杂网络上的传播来描述, 如社交网络上的流行病与谣言传播^[1,2]、电力网络上电站的级联失效^[3]、交通网络上的道路阻塞^[4]、计算机网络上的病毒传播^[5]等. 失控的传播可能对社会经济及人们生活造成大范围的负面影响, 如 COVID-19 大流行影响经济发展^[6], 谣言传播引起民众恐慌^[7], 电力网络中部分设备故障导致电网瘫痪^[8]. 由于网络结构的异质性, 网络中节点的传播影响力具有显著的差异^[9]. 如在 Twitter 与 Facebook 平台仅有少量账户的信息被广泛分享^[10],

COVID-19 中 10% 最具传染性的宿主导致近 80% 的感染^[11]. 网络传播中影响力最大的节点被称为最有影响力的传播者. 识别最有影响力的传播者是快速、有效地控制传播过程的关键步骤. 节点中心性利用网络拓扑结构度量节点在网络中的重要性. 如基于邻居数目的度中心性^[12]、H-index 中心性^[13]、k 壳中心性^[14]、集体影响中心性^[15]等, 基于路径的接近中心性^[16]、介数中心性^[17]、子图中心性^[18]以及基于特定动力学过程迭代稳态的指标, 如特征向量中心性^[19]、PageRank^[20]、非回溯中心性^[21]等. 随后, 大量基于中心性的方法被提出, 识别网络中最有影响力的传播者的算法性能不断提升, 并从单层网络推广到多层网络^[22-28].

已有识别最有影响力传播者的研究主要针对

* 国家自然科学基金 (批准号: 61802321, 61903266)、四川省科技计划项目重点研发计划 (批准号: 2023YFG0129) 和中国博士后科学基金特别资助 (批准号: 2019T120829) 资助的课题.

† 通信作者. E-mail: shinningliu@163.com

描述节点间成对关系的简单网络. 简单网络用节点表示个体, 边表示两个节点间的交互. 然而现实中许多系统存在超越点对的成组交互, 例如学术论文通常由团队共同完成^[29], 谣言与思想的传播在小范围群体中被加强^[30]. 由 3 个及以上个体组成的群体交互被定义为高阶交互, 单纯复形和超图是两种描述高阶交互的数学工具^[31]. 单纯复形是由一系列单纯形组成的集合. 一个 k 阶单纯形由 $k + 1$ 个节点组成, 且这些节点的任意节点子集也构成单纯形. 如在科学家合作网中, 若 4 个作者组成一个 3 阶单纯形, 自然其中任意 2 个作者, 任意 3 个作者也具有合作关系. 在超图中, 一条超边可包含任意数量的节点, 表示这些节点相互影响. 相比于单纯复形, 超图不要求高阶交互的所有子交互必须出现在超图中. 高阶网络上的传播动力学受到了广泛的关注. 在超图上的社会传播中, 研究发现高阶结构会导致不连续相变与双稳态区域, 双稳态区域中稳态规模依赖初始感染种子数目^[30,32]. 在社会传播模型的基础上, 研究人员提出了高阶网络上的多动力学耦合传播模型, 例如高阶网络上的竞争传播与意识疾病耦合传播等. 相比于简单图上的竞争传播^[33], 引入高阶交互使两种流行病可共存^[34]. 意识疾病耦合传播中, 社交网络中的高阶交互可缓解低流行病传播率下信息传播不足的问题, 有助于抑制流行病传播^[35].

简单网络与高阶网络间的传播现象具有明显差异, 因此识别高阶网络上最有影响力传播者的研究值得关注. 度中心性、接近中心性、介数中心性和特征向量中心性等概念被扩展至单纯复形, 用于识别核心蛋白^[36]. 类似地, 特征向量中心性被推广并应用于识别超图中的重要节点与重要超边, 其主要思想为节点的重要性依赖于节点参与超边的重要性, 而超边的重要性依赖超边所包含节点的重要性^[37]. 还有研究人员提出向量中心性以描述节点在不同尺寸超边中的重要性^[38]. 现有高阶网络中重要节点识别方法主要基于网络结构, 鲜有同时考虑网络结构与传播过程.

已有研究表明节点的传播影响力由网络拓扑结构与传播过程共同决定. 相比于仅使用网络拓扑结构的方法, 同时考虑传播过程与网络拓扑结构的方法能更准确地识别网络传播中最有影响力的传播者^[39,40]. 本文基于网络结构与传播过程, 提出传播中心性. 首先提出单纯复形上易感-感染-恢复

(susceptible-infected-recovered, SIR) 微观马尔可夫链方程组用于描述单纯复形上的 SIR 传播动力学; 接下来计算微观马尔可夫链方程组收敛时节点被感染的概率. 节点被感染的概率由网络结构与传播过程共同决定, 将其定义为“传播中心性” (spreading centrality, SC), 用于排序节点的传播影响力. 在两类合成网络与 4 个真实网络上的大规模仿真实验表明, 传播中心性与节点真实传播影响力高度正相关. 相比于高阶网络的度中心性、特征向量中心性以及简单网络上最优的集体影响中心性、非回溯中心性等基准中心性, 本文提出的传播中心性能更为准确地识别高阶 SIR 传播中最有影响力的节点, 排序准确性更高.

2 单纯复形上的传播模型

本节首先介绍单纯复形的定义, 再说明单纯复形上的易感-感染-恢复传播模型. 单纯复形定义为 $M = (V, C)$, 其中, V 表示节点集合, C 表示单纯形集合^[30]. k 阶单纯形 $\sigma \subset C$ 表示 $k + 1$ 个节点 $\{v_0, v_1, v_2, \dots, v_k\}$ 组成的高阶交互. 单纯复形中, 0 阶单纯形表示单个节点, 1 阶单纯形表示边, 2 阶单纯形表示“满”三角面等. 单纯复形满足以下条件, 若单纯形 $\omega \subset C$, 其所有子单纯形 $v \subset \omega$ 必须被包含在 C 中. 如 2 阶单纯形 $w = \{v_0, v_1, v_2\} \subset C$, 则 $\{v_0\}, \{v_1\}, \{v_2\}, \{v_0, v_1\}, \{v_0, v_2\}$ 和 $\{v_1, v_2\} \subset C$. 为了简便, 使用的单纯复形中仅包含 0 阶单纯形 (节点), 1 阶单纯形 (边) 和 2 阶单纯形 (高阶交互).

在单纯复形上的 SIR 传播^[41] 中, 个体可有 3 种状态: 易感态 (susceptible, S), 感染态 (infected, I), 或恢复态 (recovered, R). 在每个时间步, I 态个体以一定速率将 S 态邻居感染, 并以一定速率自发恢复为 R 态. 单纯复形上, 沿着 1 阶单纯形 (连边) 和 2 阶单纯形 (三角面) 都将发生感染事件. 本模型的传播参数为 (β_1, β_2, μ) , 其中 β_1 表示 1 阶单纯形感染 S 态成员的速率, β_2 表示 2 阶单纯形感染 S 态成员的速率, μ 表示 I 态个体的恢复速率. 图 1 展示单纯复形上个体感染与恢复的过程, 边表示 1 阶单纯形, “满”三角面表示 2 阶单纯形. 图 1(a)–(j) 中橙色边表示 S 态个体通过 1 阶单纯形以速率 β_1 被感染. 图 1(h) 中橙色“满”三角面表示 S 态个体通过 2 阶单纯形以速率 β_2 被感染. 图 1(h) 中, S 态个体不仅通过 1 阶单纯形被感染, 同时也通过

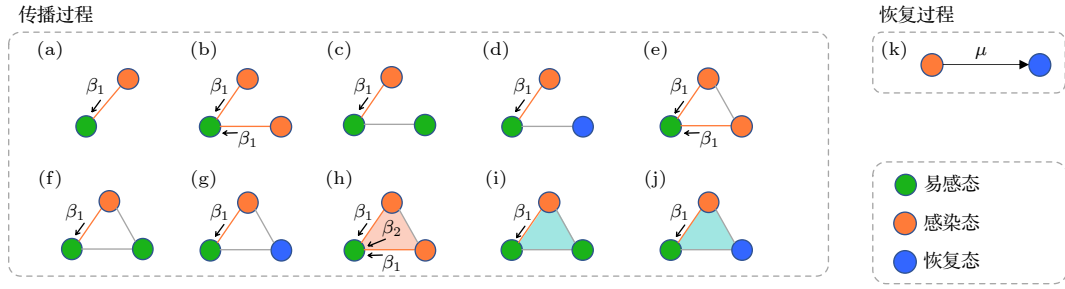


图 1 单纯复形上 SIR 传播过程 (a)–(j) 传播过程; (k) 恢复过程

Fig. 1. SIR spreading process on simplicial complex: (a)–(j) Spreading process; (k) recovery process.

2 阶单纯形被感染. 注意, 当且仅当 2 阶单纯形中的两个节点同时处于感染态时, 2 阶单纯形才能感染其第 3 个成员. 图 1(j) 中, 3 个节点虽形成“满”三角面, 但其中只有 1 个节点感染, 流行病无法通过 2 阶单纯形进行传播.

3 传播中心性

本节提出单纯复形上的微观马尔可夫链方程组, 通过微观马尔可夫链方程组计算动力学稳态时各节点被感染的概率 p_i^R , 并将其定义为节点的传播中心性. 定义 $p_i^X(t)$ 为 t 时刻节点 i 处于 X 状态的概率, $X \in \{S, I, R\}$, 如 $p_i^S(t)$ 表示 t 时刻节点 i 处于 S 态的概率. 定义 $q_i(t)$ 表示 t 时刻节点 i 没有被 1 阶单纯形感染的概率, 有

$$q_i^\Gamma(t) = \prod_{j \in \Gamma_i} (1 - \beta_1 p_j^I(t)), \quad (1)$$

其中, Γ_i 表示包含节点 i 的 1 阶单纯形集合. 定义 $q_i^\Delta(t)$ 表示 t 时刻节点 i 没有被 2 阶单纯形感染的概率, 有

$$q_i^\Delta(t) = \prod_{j,k \in \Delta_i} (1 - \beta_2 p_j^I(t) p_k^I(t)), \quad (2)$$

其中, Δ_i 表示包含节点 i 的 2 阶单纯形集合. 节点 i 在 t 时刻未被感染的概率 $q_i(t)$ 可定义为

$$q_i(t) = q_i^\Gamma(t) q_i^\Delta(t). \quad (3)$$

节点 i 处于不同状态概率随时间 t 的演化方程组可写为

$$p_i^S(t+1) = p_i^S(t) q_i(t), \quad (4)$$

$$p_i^I(t+1) = p_i^S(t)(1 - q_i(t)) + p_i^I(t)(1 - \mu), \quad (5)$$

$$p_i^R(t+1) = 1 - p_i^S(t+1) - p_i^I(t+1), \quad (6)$$

(4) 式表示节点 i 在 t 时刻处于 S 态且未被感染,

在 $t+1$ 时刻仍处于 S 态的概率. (5) 式右边第 1 项表示节点 i 在 t 时刻处于 S 态且被感染, 在 $t+1$ 时刻处于 I 态的概率. 右边第 2 项表示节点 i 在 t 时刻处于 I 态且未恢复, 在 $t+1$ 时刻仍处于 I 态的概率. (6) 式由节点处于 S 态, I 态与 R 态的概率之和为 1 的性质所得.

为了求解单个节点作为初始传播源时, 系统达到稳态后各节点被感染的概率, 初始时设置 $p_i^S(0) = 1 - 1/N$, $p_i^I(0) = 1/N$, $p_i^R(0) = 0$, N 为网络中节点数目. 迭代 (1)–(6) 式至收敛, 获得各节点被感染概率 p_i^R . 将其定义为传播中心性 (spreading centrality, SC). 流行病通过接触传播, 疾病将优先到达拥有着更多接触的个体, 即表现出更高的被感染概率. 一旦这些个体被感染, 他们更有可能将流行病传播出去, 造成大规模感染. 传播中心性属于迭代中心性, 但与特征向量中心性等不考虑传播过程的迭代中心性不同, 传播中心性加入了传播过程, 迭代过程与传播过程一致.

4 仿真实验方法

为了验证传播中心性识别最有影响力传播者的准确性, 我们在两类合成网络 (匀质度分布网络与异质度分布网络) 与 4 个真实网络上进行大量仿真实验. 本节首先介绍合成网络与 4 个真实网络的生成方法, 然后说明 4 个用于比较的基准中心性及评价准确性的指标.

4.1 网络构建方法

在实验中使用了随机单纯复形与无标度单纯复形两种合成网络. 随机单纯复形模型 (random simplicial complex, RSC) 用于生成匀质度分布的单纯复形 [42]. 生成步骤如下: 以概率 p_1 在任意两节点间生成边, 再以概率 p_2 在任意 3 个节点间生成

2 阶单纯形. p_1 , p_2 与节点平均关联的 1 阶单纯形数目 $\langle k_1 \rangle$, 节点平均关联的 2 阶单纯形数目 $\langle k_2 \rangle$ 之间关系如下:

$$p_1 = \frac{\langle k_1 \rangle - 2\langle k_2 \rangle}{(N-1) - 2\langle k_2 \rangle}, \quad (7)$$

$$p_2 = \frac{2\langle k_2 \rangle}{(N-1)(N-2)}. \quad (8)$$

无标度单纯复形模型 (scale-free simplicial complex, SFSC) 用于生成异质度分布网络^[42]. 生成步骤如下: 以 Barabási-Albert 模型生成无标度网络, 即每个新加节点按度偏好选择 m 个旧节点生成连边. 然后以概率 p_2 在任意 3 个节点间生成 2 阶单纯形. $\langle k_2 \rangle$ 可由 (8) 式计算, $\langle k_1 \rangle$ 表示为

$$\langle k_1 \rangle = 2m + 2\langle k_2 \rangle(1 - 2m/N). \quad (9)$$

采用 4 个面对面接触数据集构建真实单纯复形^[43-46]. 这些数据集从不同的人群接触场所中得到, 包括工作场所 (InVS15)、医院 (LH10)、会议 (SFHH) 与高中 (Thiers13). 每个数据集中, 记录了以 20 s 为时间间隔的个体间面对面的交互. 我们以 5 min 作为时间窗口聚合数据. 在一个时间窗口中, 若两个节点有面对面接触则生成一条边, 即 1 阶单纯形. 3 个节点两两接触则生成 2 阶单纯形. 统计边与 2 阶单纯形在所有时间窗口中出现的总次数, 最后保留 50% 出现次数最多的 1 阶单纯形与 20% 出现次数最多的 2 阶单纯形用于构建真实单纯复形. 表 1 展示了使用上述方法构建的合成单纯复形与真实单纯复形的简单统计特征, 其中 N 为网络中节点数目, $\langle k_1 \rangle$ 为节点平均关联的 1 阶单纯形数目, $\langle k_2 \rangle$ 为节点平均关联的 2 阶单纯形数目, β_1^c 是流行病爆发时, 1 阶单纯形传播速率的阈值, 通过计算机模拟所得. 在后文实验中, 将 1 阶单纯形传播速率设置为该模拟阈值的 1—2 倍, 以控制传播规模.

表 1 合成与真实单纯复形属性

Table 1. Properties of the synthetic and real simplicial complexes.

网络	N	$\langle k_1 \rangle$	$\langle k_2 \rangle$	β_1^c
RSC	2000	20	6	0.045
SFSC	5000	16	5	0.049
InVS15	213	20.19	7.94	0.040
LH10	72	15.94	13.04	0.042
SFHH	403	23.73	8.87	0.026
Thiers13	326	18.10	12.15	0.048

4.2 基准中心性

为了与本文提出的传播中心性比较, 我们选择高阶网络的度中心性^[31]与特征向量中心性^[37], 简单网络上当前最优的集体影响中心性^[15]与非回溯中心性^[21]作为比较的基准中心性.

高阶网络上的度中心性定义为节点直接邻居数目或节点参与交互数目. 本文采用节点交互数目的定义方式, 则节点度中心性可写为

$$\text{Deg}(i) = k_i + k_i^\Delta, \quad (10)$$

其中, k_i 表示节点 i 参与 1 阶单纯形的数目, k_i^Δ 表示节点 i 参与 2 阶单纯形的数目.

超图中节点与超边的特征向量中心性 (EVH) 认为节点重要性依赖于节点所在的超边, 超边重要性依赖于超边所包含的节点, 其数学定义为

$$\begin{cases} \mathbf{x} = \lambda g(\mathbf{B}f(\mathbf{y})) \\ \mathbf{y} = \mu \psi(\mathbf{B}^\top \varphi(\mathbf{x})) \end{cases} \quad \mathbf{x}, \mathbf{y} > 0, \lambda, \mu > 0, \quad (11)$$

式中, λ 和 μ 是比例常数. \mathbf{x} , \mathbf{y} 是向量, x_i 表示节点 i 的特征向量中心性值, y_e 表示超边 e 的特征向量中心性值. \mathbf{B} 是超图的关联矩阵, 超边 e 包含节点 i , 则 $B_{ie} = 1$, 反之 $B_{ie} = 0$. g, f, ψ, ϕ 是 4 个映射函数, 用于实现不同的依赖关系. 本文中设定:

$$f(x) = x, \quad g(x) = \sqrt{x}, \quad \psi(x) = e^x, \quad \varphi(x) = \ln x,$$

即节点重要性 $x_i = \lambda \sqrt{\sum_{B_{ie}=1} y_e}$, 超边重要性 $y_e = \mu \prod_{B_{ie}=1} x_i$. 为了计算高阶特征向量中心性, 可将单纯复形转化为超图, 即 1 阶单纯形转化为包含 2 个节点的超边, 2 阶单纯形转化为包含 3 个节点的超边.

集体影响中心性 (collective influence, CI) 与非回溯中心性 (nonbacktracking, NB) 是简单网络上的最优中心性算法. 集体影响中心性基于最优渗流, 推导得出节点 i 的 CI 中心性定义为

$$\text{CI}_l(i) = (k_i - 1) \sum_{j \in \partial \text{Ball}(i, l)} (k_j - 1), \quad (12)$$

式中, k_i 表示节点 i 的度, $\partial \text{Ball}(i, l)$ 表示与节点 i 距离为 l 的节点集合. 本文计算集体影响中心性时设置 $l = 2$.

非回溯中心性与标准特征向量中心性类似, 但通过非回溯矩阵避免标准特征向量中心性的局部化问题. 节点 j 的非回溯中心性定义为

$$\text{NB}(j) = \sum_i \mathbf{A}_{ij} v_{i \rightarrow j}, \quad (13)$$

其中, \mathbf{A} 表示简单图的邻接矩阵, $v_{i \rightarrow j}$ 是简单图的非回溯矩阵的主特征向量中边 $e_{i \rightarrow j}$ 对应的分量. 将简单图上的集体影响中心性与非回溯中心性算法用于单纯复形时, 我们不考虑单纯复形中的高阶交互 (即“满”三角面) 的影响, 即仅考虑网络中的节点与连边.

4.3 评价指标

为了评估不同中心性识别最有影响力传播者的准确性, 我们使用不准确函数^[14]、肯德尔相关系数与 Top- K 准确率作为评价指标. 不准确函数 $\xi(p)$ 度量中心性识别最有影响力传播者的不准确性, 定义为

$$\xi(p) = 1 - \frac{M(p)}{M_{\text{eff}}(p)}, \quad (14)$$

其中 p 是计算节点的比例 ($0 \leq p \leq 1$). $M(p)$ 表示规模为 N 的网络中, 中心性排序在前的 pN 个节点的平均传播影响力, $M_{\text{eff}}(p)$ 表示真实传播影响力排序在前的 pN 个节点的平均传播影响力. $\xi(p)$ 量化了实际最有传播影响力的传播者与中心性最高的传播者间平均传播影响力的接近程度. $\xi(p)$ 越小, 中心性在识别最有影响力的传播者时越准确.

肯德尔相关系数用于度量两个排序序列的相关性, 定义为

$$\tau(r^1, r^2) = \frac{K(r^1, r^2) - K'(r^1, r^2)}{0.5n(n-1)}, \quad (15)$$

其中 n 表示序列长度, $K(r^1, r^2)$ 表示序列 r^1, r^2 中的同序对数目, $K'(r^1, r^2)$ 表示序列 r^1, r^2 中的异序对数目. 用 r_i^1, r_j^1 表示序列 r^1 中第 i 个与第 j 个元素的排序值, r_i^2, r_j^2 表示序列 r^2 中第 i 个与第 j 个元素的排序值. 若 $(r_i^1 - r_j^1)(r_i^2 - r_j^2) > 0$, 则 (i, j) 是同序对. 若 $(r_i^1 - r_j^1)(r_i^2 - r_j^2) < 0$, 则 (i, j) 是异序对. 本文中 r^1 为节点根据中心性排序的序列, r^2 为节点根据真实传播影响力排序的序列. τ 取值在 -1 到 1 之间, τ 越大表明中心性与节点真实传播影响力越相关, 该中心性识别越准确.

Top- K 准确率用于度量中心性对最有传播影响力的前 K 个节点的识别准确性, 定义为

$$\text{Top-}K \text{ 准确率} = \frac{|S_K \cap S'_K|}{K}, \quad (16)$$

其中, S_K 表示中心性值在前 K 的节点集合, S'_K 表示真实传播影响力在前 K 的节点集合, $|S_K|$ 表

示集合 S_K 中的元素数目. Top- K 准确率越高表示中心性识别最有传播影响力的节点越准确.

5 实验结果

我们在合成网络与真实网络上度量传播中心性 SC 排序节点影响力的性能. 首先, 分析了节点传播中心性与传播影响力的相关性, 如图 2 所示. 结果表明, 节点传播中心性与传播影响力高度正相关. 当 $\beta_1 = 1.2\beta_1^c$ 时, RSC 网络上节点传播中心性与传播影响力的皮尔逊相关系数为 0.9691; 当 $\beta_1 = 1.6\beta_1^c, 2.0\beta_1^c$ 时, 皮尔逊相关系数分别为 0.9887 与 0.9910. 其他网络上, 结果与 RSC 网络中类似. 实验中, 为了获得单个节点的传播影响力, 对单纯复形上的 SIR 传播模型进行大规模的蒙特卡罗模拟. 初始时, 将节点 i 设置为感染态, 其余节点设置为易感态, 统计传播达到稳态时网络的感染规模, 即恢复态占比. 独立重复上述模拟 10^3 次, 将感染规模的均值作为对节点 i 的真实传播影响力 S_i 的估计. 当系统初始状态确定时, 微观马尔可夫链方程组的收敛状态是固定的, 不具备随机性. 我们设置 $p_i^S(t=0) = 1 - 1/N, p_i^I(t=0) = 1/N, p_i^R(t=0) = 0$, 通过迭代微观马尔可夫链方程组计算传播中心性. 在实验中, 固定恢复速率 $\mu = 1$, 通过改变传播速率 β_1 与 β_2 设定不同的有效传播速率. μ 的取值仅影响传播的时间演化尺度, 不影响实验结果.

为了考察提出的传播中心性 SC 排序节点传播影响力的准确性, 采用不准确函数作为评价指标. 不准确函数值越低, 该中心性识别最有影响力的传播者越准确. 如图 3 所示, 传播中心性能准确地识别最有影响力的传播者. 在所有网络上, 传播中心性的不准确函数低于 0.02, 传播中心性多数情况下优于其他基准中心性. 在 SFSC 与 Thiers13 网络中, 传播中心性明显优于所有基准中心性. 在 RSC 网络上, 传播中心性略优于集体影响中心性、特征向量中心性与非回溯中心性, 明显优于度中心性. 在 InVS15 网络中, 当 $0 < p \leq 0.05$ 时传播中心性与非回溯中心性接近, 明显优于其他基准中心性; 当 $0.05 < p \leq 0.2$ 时, 传播中心性完全优于基准中心性. 在 LH10 网中, 传播中心性与非回溯中心性接近, 略优于度中心性与特征向量中心性, 明显优于集体影响中心性. 在 SFHH 网络中, 当 $0 < p \leq 0.08$ 时传播中心性略差于非回溯中心性, 与特征向量中

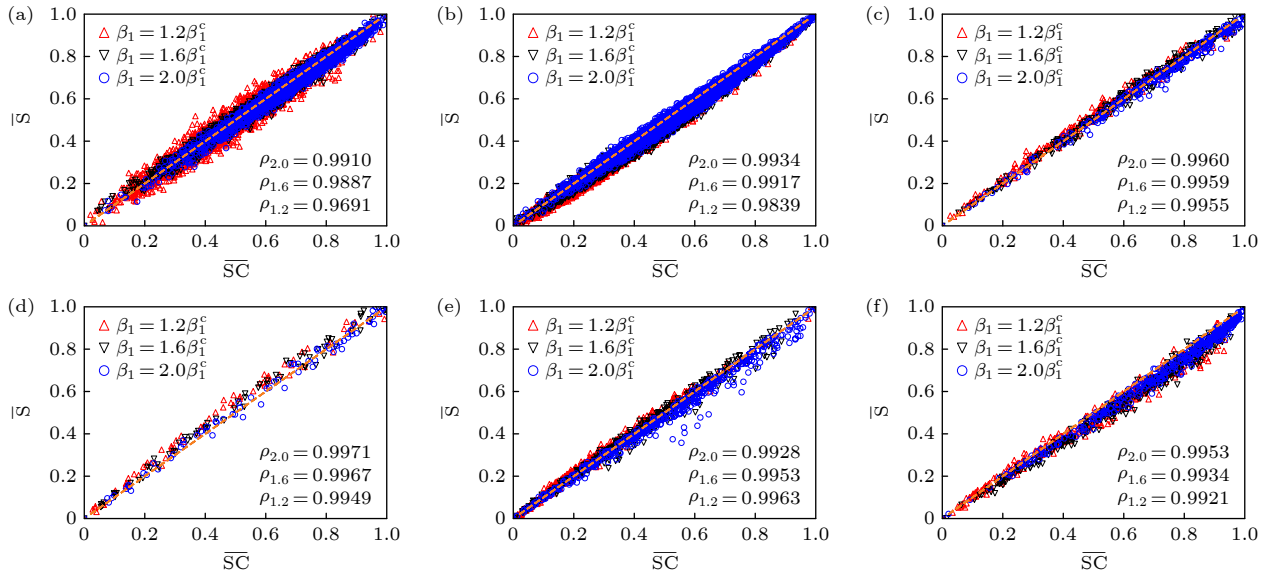


图 2 节点传播中心性与传播影响力的散点图. \overline{SC} 与 \overline{S} 分别表示归一化后的节点传播中心性与传播影响力 (a) RSC; (b) SFSC; (c) InVS15; (d) LH10; (e) SFHH; (f) Thiers13

Fig. 2. Scatter plots of the spreading centrality and spreading influence of nodes. \overline{SC} and \overline{S} represent the normalized spreading centrality and spreading influence of nodes: (a) RSC; (b) SFSC; (c) InVS15; (d) LH10; (e) SFHH; (f) Thiers13.

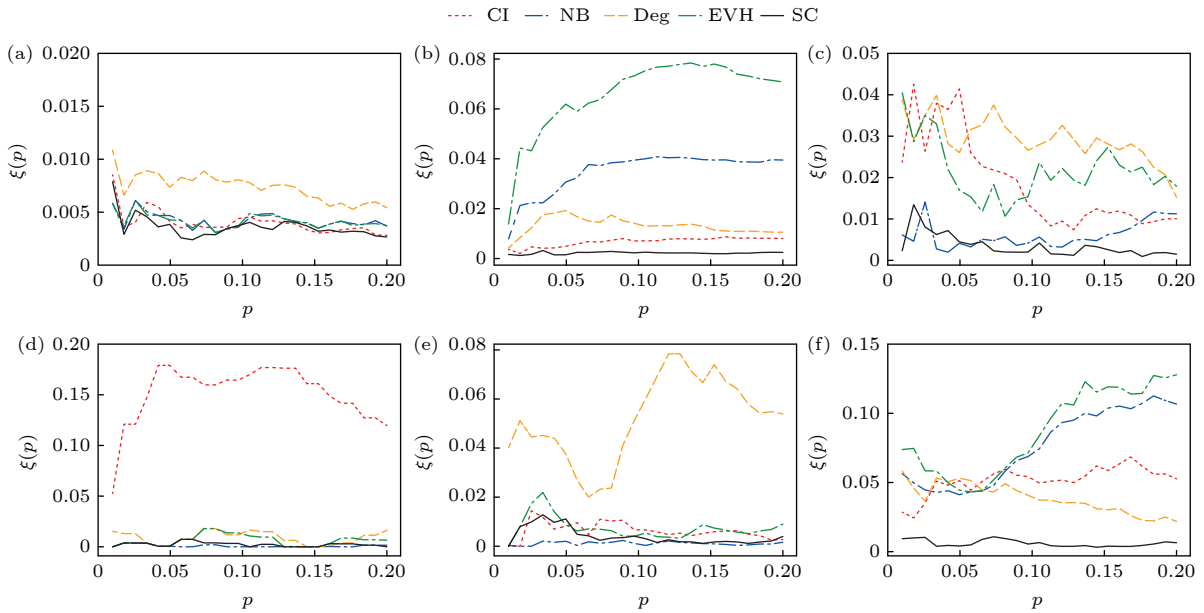


图 3 各中心性的不准确函数 (a) RSC; (b) SFSC; (c) InVS15; (d) LH10; (e) SFHH; (f) Thiers13

Fig. 3. Imprecisions of the centralities: (a) RSC; (b) SFSC; (c) InVS15; (d) LH10; (e) SFHH; (f) Thiers13.

心性和集体影响中心性接近; 当 $0.08 < p \leq 0.2$ 时, 传播中心性与非回溯中心性接近, 优于其他基准中心性. 传播中心性利用节点被感染的概率排序节点影响力. 节点 i 被感染概率越高其参与的交互数目应越多, 即拥有较高的高阶中心性. 同时, 节点 i 的邻居要感染节点 i , 必要条件是自己先被感染, 因此节点 i 的邻居同样具有较高的高阶中心性. 传播参数设置如下: 恢复速率 $\mu = 1$, 1 阶单纯形传

播速率设置为模拟阈值 β_1^c 的 1.6 倍, $\beta_1 = 1.6\beta_1^c$, 2 阶单纯形传播速率 $\beta_2 = 0.1$.

为了考察传播参数改变时, 提出的传播中心性的性能, 对比了不同 1 阶单纯形传播速率下各中心性与节点传播影响力的相关性, 如图 4 所示. 结果表明, 在所有网络中, 所提的传播中心性与传播影响力的肯德尔相关系数大于基准中心性与传播影响力的肯德尔相关系数. 合成网络中传播中心性的

肯德尔相关系数随 1 阶单纯形传播速率的增大而增大. 在 RSC 网络中, 传播中心性在 $1 \leq \alpha \leq 1.4$ 时略优于特征向量中心性、非回溯中心性和集体影响中心性, 明显优于度中心性. 当 $1.4 < \alpha \leq 2$ 时略优于集体影响中心性, 明显优于其余中心性. SFSC 网络中, 传播中心性明显优于其他基准中心性. 真实网络中结果与合成网络结果类似. 在 InVS15 与 Thiers13 网络中, 传播中心性明显优于其他基准中心性. 在 LH10 与 SFHH 网络中, 传播中心性略优于非回溯中心性, 明显优于集体影响中心性, 特征向量中心性与度中心性等基准中心性. 其余传播参数设置如下: 恢复速率 $\mu = 1$, 1 阶单纯形传播速率 $\beta_1 = \alpha\beta_1^0$. 2 阶单纯形传播速率设定为 $\beta_2 = 0.1$. 实验中设置 1 阶单纯形传播速率 $\beta_1 \in [\beta_1^0, 2\beta_1^0]$. 当 $\beta_1 < \beta_1^0$ 时, 流行病不能爆发, 所有节点的传播影响力接近于 0; 当 $\beta_1 > 2\beta_1^0$ 时, 传播速率过大, 所有节点都会导致较高的流行病爆发规模. 传播速率过小或过大都无法充分地体现节点间传播影响力的差异. 基本再生数度量感染节点感染下一代节点的能力, 是传播中的一个重要参量. 复杂网络 SIR 传播模型基本再生数定义为 $R_0 = (k + s^2/k)\beta/\mu$, 其中 β 是传播速率, μ 是恢复速率, k 是平均度, s 是度分布的标准差. 实验中, 设置 1 阶单纯形传播速率 $\beta_1 \in [\beta_1^0, 2\beta_1^0]$, 此时各网络的基本再生数取值范围

为: RSC [0.98, 1.96]; SFSC [1.06, 2.12]; InVS15 [0.97, 1.94]; LH10 [0.97, 1.94]; SFHH [0.99, 1.98]; Thiers13 [1.05, 2.10]. 由于流行病主要通过边进行传播, 高阶作用仅起强化作用, 我们计算 R_0 时未考虑高阶作用.

在高阶网络传播中, 2 阶单纯形传播速率会影响节点的传播影响力. 我们改变 2 阶单纯形传播速率 β_2 , 考察中心性与节点传播影响力的相关性. 如图 5 所示, 在合成网络和真实网络中, 多数情况下传播中心性 SC 与节点传播影响力的相关性最高. 在合成网络 RSC 与 SFSC 中, 传播中心性的肯德尔相关系数基本不随 β_2 变化且优于基准中心性. 在真实网络 InVS15, LH10 与 SFHH 中, 传播中心性的肯德尔相关系数随 β_2 的增大呈现下降趋势. 在 Thiers13 中传播中心性的肯德尔相关系数变化趋势平坦. 与基准中心性相比, InVS15 与 Thiers13 网络中传播中心性完全优于其他基准中心性. 在 LH10 网络中, 当 $0 \leq \beta_2 \leq 0.1$ 时, 传播中心性优于所有基准中心性, 当 $0.1 < \beta_2 \leq 0.2$ 时传播中心性略差于非回溯中心性但优于其他基准中心性. SFHH 网络结果与 LH10 类似, 当 $0 \leq \beta_2 < 0.15$ 时传播中心性优于所有基准中心性, 而当 $0.15 \leq \beta_2 \leq 0.2$ 时传播中心性略差于非回溯中心性但优于其他基准中心性. 当 $\beta_2 = 0$ 时, 2 阶单纯形不再传播流行病,

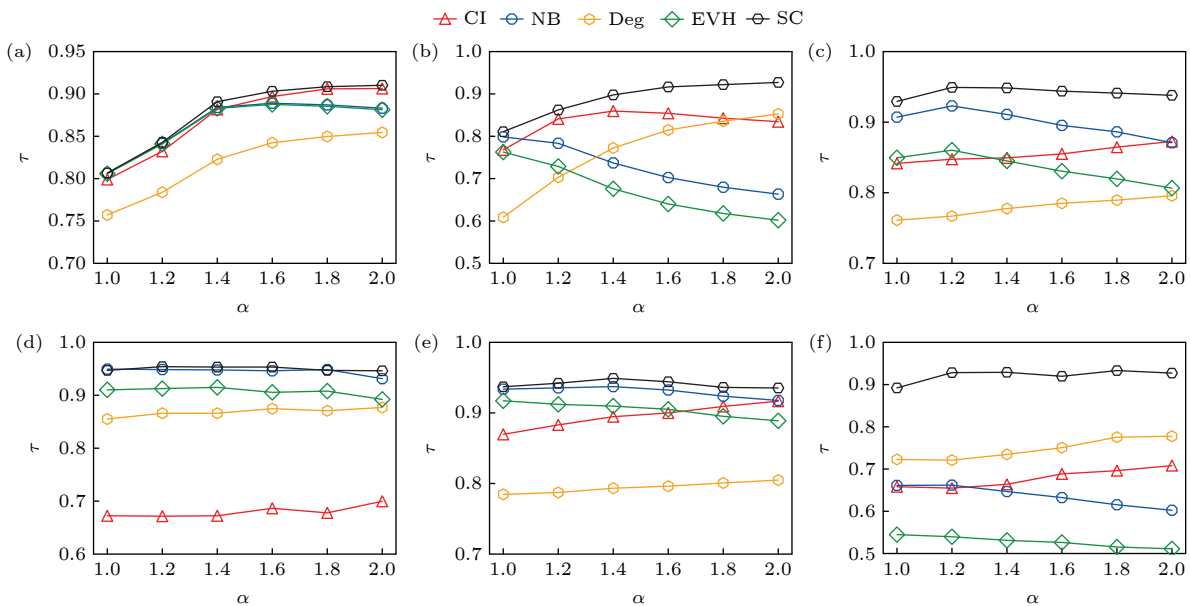


图 4 不同 1 阶单纯形传播速率 $\beta_1 = \alpha\beta_1^0$ 下, 节点各中心性与传播影响力的肯德尔相关系数 (a) RSC; (b) SFSC; (c) InVS15; (d) LH10; (e) SFHH; (f) Thiers13

Fig. 4. Kendall's tau correlation of the centralities and the spreading influence of nodes under different 1-simplex spreading rates $\beta_1 = \alpha\beta_1^0$: (a) RSC; (b) SFSC; (c) InVS15; (d) LH10; (e) SFHH; (f) Thiers13.

高阶网络上的传播退回到普通网络上的 SIR 传播模型. 此时, 节点的传播中心性与传播影响力间的肯德尔系数接近或高于 0.9, 明显优于普通网络上性能最优的集体影响中心性与非回溯中心性. 随着 2 阶单纯形传播速率 β_2 增大, 传播中心性与传播影响力间的相关性呈现出小幅降低的现象. 当传播发生在 2 阶单纯形上时, 节点感染与被感染过程是非对称的. 若 2 阶单纯形中仅有一个节点处于感染态, 该单纯形无法传播流行病, 即一个节点容易通过 2 阶单纯形被感染不代表它容易通过 2 阶单纯形传播流行病. 随着 2 阶传播速率 β_2 增大, 这种非对称性带来的负面效果将会升高, 传播中心性与传播影响力间的相关性将小幅降低. 在图 5 的实验中, 其余传播参数设置为: $\beta_1 = 1.6\beta_1^c$, $\mu = 1$.

最后, 通过 Top- K 准确率表征各中心性对前

K 个最有传播影响力节点的识别准确率. 表 2 展示了各中心性的 Top-10, Top-20 与 Top-30 准确率. 综合来讲, 传播中心性多数情况下能较为准确地识别前 K 个最有影响力的传播者. 当 $K = 10$ 时, 在 RSC, InVS15, SFHH 与 Thiers13 网络中, 传播中心性有着最高的 Top-10 准确率, 即在识别前 10 个影响力最大的节点时, 传播中心性 SC 准确率最高. 当 $K = 20$ 时, 在除 SFSC 网络外的其他网络中, 传播中心性有着最高的 Top-20 识别准确率. 当 $K = 30$ 时, 传播中心性在 SFSC, LH10 与 Thiers13 网络中有着最高的 Top-30 识别准确率. Thiers13 网络中, 传播中心性有着最高的 Top- K 准确率, 且明显高于所有基准中心性的 Top- K 准确率. 传播参数设置如下: $\beta_1 = 1.2\beta_1^c$, $\beta_2 = 0.1$, $\mu = 1$.

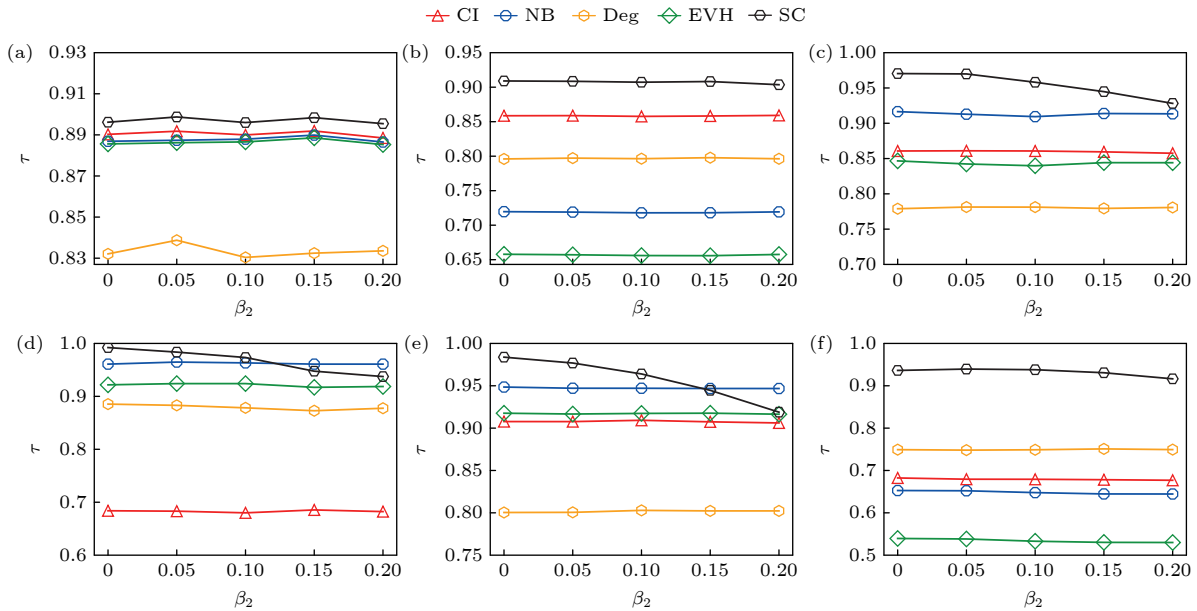


图 5 不同 2 阶单纯形传播速率 β_2 下, 节点各中心性与传播影响力的肯德尔相关系数 (a) RSC; (b) SFSC; (c) InVS15; (d) LH10; (e) SFHH; (f) Thiers13

Fig. 5. Kendall's tau correlation of the centralities and the spreading influence of nodes under different 2-simplex spreading rates β_2 : (a) RSC; (b) SFSC; (c) InVS15; (d) LH10; (e) SFHH; (f) Thiers13.

表 2 各中心性的 Top- K 准确率
Table 2. Top- K accuracy of centralities.

网络	$K = 10$					$K = 20$					$K = 30$				
	CI	NB	Deg	EVH	SC	CI	NB	Deg	EVH	SC	CI	NB	Deg	EVH	SC
RSC	0.60	0.60	0.50	0.60	0.60	0.70	0.75	0.60	0.75	0.75	0.70	0.77	0.57	0.77	0.73
SFSC	0.90	0.90	1.00	0.90	0.90	0.95	1.00	0.90	0.95	0.95	0.93	0.97	0.87	0.93	0.97
InVS15	0.40	0.90	0.50	0.70	0.90	0.65	0.90	0.70	0.90	0.90	0.77	0.97	0.70	0.83	0.90
LH10	0.10	1.00	0.90	0.90	0.90	0.55	0.95	0.90	0.95	0.95	0.87	1.00	0.93	0.97	1.00
SFHH	0.80	0.90	0.60	0.70	0.90	0.85	0.90	0.70	0.90	0.90	0.87	0.93	0.77	0.87	0.90
Thiers13	0.30	0.30	0.50	0.10	0.70	0.40	0.40	0.40	0.40	0.70	0.37	0.50	0.43	0.50	0.83

6 结 论

识别网络传播中最有影响力的传播者是控制传播过程的关键步骤之一. 本文研究高阶网络上 SIR 传播过程中最有影响力传播者的识别方法. 我们提出单纯复形上 SIR 微观马尔可夫链方程组, 用于描述 SIR 传播动力学过程. 利用微观马尔可夫链方程组, 计算单源传播时各节点被感染的概率 p_i^R , 定义度量节点传播影响力的指标, 传播中心性. 大量仿真实验表明, 传播中心性与节点真实传播影响力高度正相关. 相比于 4 种基准中心性, 即高阶网络上的度中心性、特征向量中心性等仅考虑网络结构的中心性与复杂网络上最优的集体影响中心性、非回溯中心性, 本文提出的传播中心性在两类合成网络中识别传播重要节点时比基准中心性更准确, 在真实网络中多数情况下优于基准中心性. 传播中心性方法可以扩展到超图上. 对于超图上的传播, 可以类似地写出描述传播动力学的微观马尔可夫链方程组, 再利用该方程组计算节点的传播中心性.

传播中心性具有一定局限性. 通过节点被感染概率预测节点的真实传播影响力时, 假设越容易被感染的个体越容易感染邻居, 即感染与被感染是对称的. 考虑感染与被感染的方向性, 当传播发生在 1 阶单纯形上时, 感染与被感染是对称的; 当传播发生在 2 阶单纯形上时, 2 阶单纯形中单个感染态节点并不能通过该 2 阶单纯形传播流行病, 感染与被感染是非对称的. 感染与被感染的非对称性会降低传播中心性的排序准确性, 考虑 2 阶单纯形上的非对称性可以进一步提升传播中心性的准确性. 当感染与被感染的对称性受到严重破坏时, 传播中心性可能不再适用. 例如传播在有向网络上发生时, 传播中心性可能不再准确. 因此, 未来需要进一步探索如何在有向网络上识别最有影响力的传播者.

参考文献

[1] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
 [2] Moreno Y, Nekovee M, Pacheco A F 2004 *Phys. Rev. E* **69** 066130
 [3] Motter A E 2004 *Phys. Rev. Lett.* **93** 098701
 [4] Li D, Fu B, Wang Y, Lu G, Berezin Y, Stanley H E, Havlin S 2014 *Proc. Natl. Acad. Sci.* **112** 669
 [5] Kephart J O, Sorkin G B, Chess D M, White S R 1997 *Sci. Am.* **277** 88

[6] Hale T, Angrist N, Goldszmidt R, et al. 2023 *Nat. Hum. Behav.* **5** 529
 [7] Rocha Y M, de Moura G A, Desidério G A, et al. 2023 *J. Public Health* **31** 1007
 [8] Schäfer B, Witthaut D, Timme M, Latora V 2018 *Nat. Commun.* **9** 1975
 [9] Ren X L, Lü L Y 2014 *Sci. Bull.* **59** 1175 (in Chinese) [任晓龙, 吕琳媛 2014 *科学通报* **59** 1175]
 [10] Yang K C, Pierri F, Hui P M, Axelrod D, Torres-Lugo C, Bryden J, Menczer F 2021 *Big Data Soc.* **8** 1
 [11] Nielsen B F, Simonsen L, Sneppen K 2021 *Phys. Rev. Lett.* **126** 118301
 [12] Freeman L C 1978 *Soc. Networks* **1** 215
 [13] Lü L, Zhou T, Zhang Q M, Stanley H E 2016 *Nat. Commun.* **7** 10168
 [14] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
 [15] Morone F, Makse H A 2015 *Nature* **524** 65
 [16] Sabidussi G 1966 *Psychometrika* **31** 581
 [17] Freeman L C 1977 *Sociometry* **40** 35
 [18] Estrada E, Rodríguez-Velázquez J A 2005 *Phys. Rev. E* **71** 056103
 [19] Bonacich P, Lloyd P 2001 *Soc. Networks* **23** 191
 [20] Brin S, Page L 1998 *Comput. Netw. ISDN Syst.* **30** 107
 [21] Martin T, Zhang X, Newman M E J 2014 *Phys. Rev. E* **90** 052808
 [22] Lü L, Chen D, Ren X L, et al. 2016 *Phys. Rep.* **650** 1
 [23] Wang T T, Liang Z W, Zhang R X 2023 *Acta Phys. Sin.* **72** 048901 (in Chinese) [汪亭亭, 梁宗文, 张若曦 2023 *物理学报* **72** 048901]
 [24] Maji G, Namtirtha A, Dutta A, Malta M C 2020 *Exp. Syst. Appl.* **144** 113092
 [25] Liu J Q, Li X R, Dong J C 2021 *Sci. China Technol. Sci.* **64** 451
 [26] Liu Y, Zeng Q, Pan L, Tang M 2023 *IEEE Trans. Netw. Sci. Eng.* **10** 2201
 [27] Fan T, Lü L, Shi D, Zhou T 2021 *Commun. Phys.* **4** 272
 [28] Ruan Y R, Lao S Y, Tang J, Bai L, Guo Y M 2022 *Acta Phys. Sin.* **71** 176401 (in Chinese) [阮逸润, 老松杨, 汤俊, 白亮, 郭延明 2022 *物理学报* **71** 176401]
 [29] Lung R I, Gaskó N, Suciú M A 2018 *Scientometrics* **117** 1361
 [30] Iacopini I, Petri G, Barrat A, Latora V 2019 *Nat. Commun.* **10** 2485
 [31] Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J G, Petri G 2020 *Phys. Rep.* **874** 1
 [32] de Arruda G F, Petri G, Moreno Y 2020 *Phys. Rev. Res.* **2** 023032
 [33] Wang W, Liu Q H, Liang J, Hu Y, Zhou T 2019 *Phys. Rep.* **820** 1
 [34] Li W Y, Xue X, Pan L, Lin T, Wang W 2022 *Appl. Math. Comput.* **412** 126595
 [35] Fan J, Yin Q, Xia C, Perc M 2022 *Proc. R. Soc. A.* **478** 20220059
 [36] Estrada E, Ross G J 2018 *J. Theor. Biol.* **438** 46
 [37] Tudisco F, Higham D J 2021 *Commun. Phys.* **4** 201
 [38] Kovalenko K, Romance M, Vasilyeva E, et al. 2022 *Chaos Solitons Fractals* **162** 112397
 [39] Liu J G, Lin J H, Guo Q, Zhou T 2016 *Sci. Rep.* **6** 21380
 [40] Zeng Q, Liu Y, Tang M, Gong J 2021 *Knowledge-Based Syst.* **229** 107365
 [41] Li W, Nie Y, Li W, Chen X, Su S, Wang W 2022 *Chaos* **32** 093135
 [42] Wang H, Ma C, Chen H S, Lai Y C, Zhang H F 2022 *Nat.*

Commun. **13** 3043[43] Génois M, Barrat A 2018 *Epl Data Sci.* **7** 11[44] Isella L, Stehlé J, Barrat A, Cattuto C, Pinton J F, Van den Broeck W 2011 *J. Theor. Biol.* **271** 166[45] Vanhems P, Barrat A, Cattuto C, Pinton J F, Khanafer N, Régis C, Kim B, Comte B, Voirin N 2013 *PloS One* **8** e73970[46] Mastrandrea R, Fournet J, Barrat A 2015 *PloS One* **10** e0136497

Identifying influential nodes in spreading process in higher-order networks*

Li Jiang¹⁾ Liu Ying^{1)2)†} Wang Wei³⁾ Zhou Tao⁴⁾1) (*School of Computer Science, Southwest Petroleum University, Chengdu 610500, China*)2) (*Department of Physics, University of Fribourg, Fribourg 1700, Switzerland*)3) (*College of Public Health, Chongqing Medical University, Chongqing 400016, China*)4) (*Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China*)

(Received 1 September 2023; revised manuscript received 31 October 2023)

Abstract

Identifying influential nodes in spreading process in the network is an important step to control the speed and range of spreading, which can be used to accelerate the spread of beneficial information such as healthy behaviors, innovations and suppress the spread of epidemics, rumors and fake news. Existing researches on identification of influential spreaders are mostly based on low-order complex networks with pairwise interactions. However, interactions between individuals occur not only between pairwise nodes but also in groups of three or more nodes, which introduces complex mechanism of reinforcement and indirect influence. The higher-order networks such as simplicial complexes and hypergraphs, can describe features of interactions that go beyond the limitation of pairwise interactions. Currently, there are relatively few researches of identifying influential spreaders in higher-order networks. Some centralities of nodes such as higher-order degree centrality and eigenvector centrality are proposed, but they mostly consider only the network structure. As for identification of influential spreaders, the spreading influence of a node is closely related to the spreading process. In this paper, we work on identification of influential spreaders on simplicial complexes by taking both network structure and dynamical process into consideration. Firstly, we quantitatively describe the dynamics of disease spreading on simplicial complexes by using the Susceptible-Infected-Recovered microscopic Markov equations. Next, we use the microscopic Markov equations to calculate the probability that a node is infected in the spreading process, which is defined as the spreading centrality (SC) of nodes. This spreading centrality involves both the structure of simplicial complex and the dynamical process on it, and is then used to rank the spreading influence of nodes. Simulation results on two types of synthetic simplicial complexes and four real simplicial complexes show that compared with the existing centralities on higher-order networks and the optimal centralities of collective influence and nonbacktracking centrality in complex networks, the proposed spreading centrality can more accurately identify the most influential spreaders in simplicial complexes. In addition, we find that the probability of nodes infected is highly positively correlated with its influence, which is because disease preferentially reaches nodes with many contacts, who can in turn infect their many neighbors and become influential spreaders.

Keywords: higher-order network, simplicial complexes, rank the spreading influence of nodes, microscopic Markov chain, complex networks

PACS: 89.75.Fb, 87.23.Ge, 89.20.Ff

DOI: 10.7498/aps.73.20231416

* Project supported by the National Natural Science Foundation of China (Grant Nos. 61802321, 61903266), the Science and Technology Program of Sichuan Province, China (Grant No. 2023YFG0129), and the China Postdoctoral Science Foundation-Special Fund (Grant No. 2019T120829).

† Corresponding author. E-mail: shinningliu@163.com



识别高阶网络传播中最有影响力的节点

李江 刘影 王伟 周涛

Identifying influential nodes in spreading process in higher-order networks

Li Jiang Liu Ying Wang Wei Zhou Tao

引用信息 Citation: *Acta Physica Sinica*, 73, 048901 (2024) DOI: 10.7498/aps.73.20231416

在线阅读 View online: <https://doi.org/10.7498/aps.73.20231416>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于区域密度曲线识别网络上的多影响力节点

Identifying multiple influential nodes based on region density curve in complex networks

物理学报. 2018, 67(19): 198901 <https://doi.org/10.7498/aps.67.20181000>

动态复杂网络中节点影响力的研究进展

Node influence of the dynamic networks

物理学报. 2020, 69(4): 048901 <https://doi.org/10.7498/aps.69.20190830>

基于引力方法的复杂网络节点重要度评估方法

Node importance ranking method in complex network based on gravity method

物理学报. 2022, 71(17): 176401 <https://doi.org/10.7498/aps.71.20220565>

有限步传播范围期望指标判别节点传播影响力

Distinguishing node propagation influence by expected index of finite step propagation range

物理学报. 2020, 69(2): 028901 <https://doi.org/10.7498/aps.69.20191313>

基于复杂网络动力学模型的无向加权网络节点重要性评估

Evaluation methods of node importance in undirected weighted networks based on complex network dynamics models

物理学报. 2018, 67(9): 098901 <https://doi.org/10.7498/aps.67.20172295>

复杂网络牵制控制优化选点算法及节点组重要性排序

Node-set importance and optimization algorithm of nodes selection in complex networks based on pinning control

物理学报. 2021, 70(5): 056401 <https://doi.org/10.7498/aps.70.20200872>