

基于传播方向冗余抑制的复杂网络多影响力节点识别方法 *

叶超然 张俊丽 张萌萌 阮逸润[†] 白亮

(国防科技大学系统工程学院, 长沙 410073)

在舆情导控、病毒营销及疫情阻断等复杂网络应用中, 多影响力节点识别需要兼顾种子节点的传播能力和种子集合内部的覆盖重合。本文从传播方向的异质性出发, 提出一种基于传播方向冗余抑制 (propagation-direction redundancy suppression, PDRS) 的多影响力节点识别算法。PDRS 利用传播方向权重分布的熵定义有效方向指数, 以衡量支撑节点传播潜能的有效方向数。每当选出一个种子节点后, 算法根据相邻候选节点与该种子的传播方向重合程度动态削减候选得分, 使后续种子分布到尚未充分覆盖的网络区域。本文在 9 个真实无向网络和 9 个真实有向网络上采用易感、感染和恢复模型 (susceptible-infected-recovered, SIR) 开展实验。PDRS 取得较高的平均最终感染比例, 并减少种子节点之间的局部覆盖重合。

关键词: 复杂网络, 多影响力节点, 传播方向冗余, 有效方向指数

PACS: 89.75.Hc, 89.70.Cf, 89.75.Fb

1 引言

复杂网络为刻画现实系统中个体及其相互作用提供了有效工具, 社交、交通、生物和通信 [1-3] 等众多系统均可抽象为由节点和连边构成的网络。在信息扩散、舆情调控、疾病防控和病毒式营销 [4,5] 等场景中, 少数节点往往主导着传播的范围与速度。因此, 识别网络中具传播影响力的节点, 并据此选取一组初始传播源以扩大传播范围, 成为复杂网络传播动力学研究中的重要问题 [6-9]。

[†] 通信作者. E-mail: ruanyirun@nudt.edu.cn

* 国家自然科学基金 (批准号: 72401286) 资助的课题。

贪心算法及其改进通过反复估计候选节点对当前种子集合的边际增益来选种，在独立级联和线性阈值等经典模型下具有较好的理论保证 [4,5,10]。这类方法依赖大量传播模拟，在大规模网络中计算代价较高 [11–13]。度中心性、介数中心性、接近中心性 [14]、Katz 中心性 [15]、特征向量中心性 [16]、PageRank [17]、H 指数 [18] 和 K-shell [19] 等拓扑指标分别从邻居数量、核层位置、最短路径或随机游走评价单节点的传播潜力。部分方法进一步结合节点度、邻居度、公共邻居、聚类系数和网络位置 [20–25]，或在重力中心性基础上引入递归排序 [26]。熵类方法则利用香农熵（Shannon entropy）刻画局部结构 [27]，利用 Tsallis 熵（Tsallis entropy）构造节点排序 [28,29]，或结合社区结构与结构熵识别影响力节点 [30]。

上述方法主要评价单节点的重要性。直接选取排序前 K 名时，高分节点可能集中于同一局部密集区域，增加种子集合内部的覆盖重合。

为提升种子分散性，另一类方法在迭代选种中动态调整候选节点得分。度折扣算法（Degree Discount）每选出一个种子后，折扣其邻域内候选节点的有效度 [12]。基于信息熵更新的 EnRenew 算法在每轮选种后按距离递减地衰减可达节点的传播能力 [31]。VoteRank [32] 及其改进通过迭代投票选择种子，并削弱已选种子邻域的投票能力。

WVoteRank 将投票机制推广到加权网络 [33]，NCVoteRank 以邻域核数刻画投票能力 [34]，VoteRank++ 计入节点间联系的紧密程度 [35]。边加权 VoteRank（Edge Weighted VoteRank, EWV）融合边权、区域影响与节点相似性 [36]，DILVoteRank 借助连边重要性区分邻居贡献 [37]，IMVoteRank 引入相似性投票和群组隔离 [38]，基于社区的 VoteRank（Community-based VoteRank, CVoteRank）利用社区结构刻画跨社区传播能力 [39]。NOIRX 先组合多种中心性识别候选节点，再综合信息熵等指标抑制种子重叠 [40]。

近期研究分别从空间分散、环结构和社区尺度处理多节点传播范围的重合。Ramrakhiani 等利用最近优先遍历划分网络区域，并从不同区域选择局部传播能力较强的节点 [57]。Shi 等提出 CycRank，利用环结构从已有中心性排序中选择节点集合 [58]。Tudu 等综合社区内部和社区之间的影响得分，并采用惩罚机制降低已选节点传播范围的重合 [59]。Maji 从无向边的两个传播方向构造方向节点强度熵，用于评价单节点传播能力 [60]。这些工作共同涉及节点自身传播能力和种子集合内部覆盖重合的评价。

本文将节点沿各邻接边向外传播视为不同传播方向，并据此提出 PDRS。与以路径多样性 [41] 刻画单节点、以冗余影响力 [42] 调整种子的工作相比，PDRS 分别在初始评分和迭代选种中处理传播方向冗余。有效方向指数利用方向权重分布的熵衡量节点的有效传播方向数。传播重叠抑制策略在每轮选种后，根据

相邻候选节点与新种子的传播方向重合程度更新候选得分。本文还将该方法扩展到有向网络，并通过传播实验、参数实验和消融实验分析其传播表现、计算开销及适用范围。

2 问题定义与算法模型

设无向网络为 $G = (V, E)$ ，其中 V 和 E 分别表示节点集合和边集合， $|V| = N$ ， $|E| = M$ 。对任意节点 $u \in V$ ，记其邻居集合为 $\Gamma(u) = \{v \in V \mid (u, v) \in E\}$ ，节点度为 $d_u = |\Gamma(u)|$ 。给定种子节点数量 K ，算法目标是选取种子集合 $S \subseteq V$ ， $|S| = K$ ，使其在后续传播中获得较大覆盖范围。PDRS 从两个环节抑制传播方向冗余。有效方向指数使初始得分偏向传播由多个低冗余方向支撑的节点，传播重叠抑制策略则在迭代选中降低种子集合内部的局部覆盖重合。

2.1 初始传播得分

节点 u 的影响力经由不同邻居向外传播，故边 (u, v) 是 u 经由邻居 v 的一条传播方向，记为 $u \rightarrow v$ 。若 u 与 v 具有公共邻居，则边 (u, v) 与每个公共邻居共同构成一个三角形。节点 u 沿方向 $u \rightarrow v$ 传播后，再由 v 进入这些公共邻居时，会与 u 经由其它邻接方向形成的局部覆盖重合。公共邻居越多，该传播方向在三角闭合范围内能够增加的覆盖越少。为刻画这种由三角闭合形成的局部传播冗余，本文定义边 (u, v) 的冗余程度为

$$r_{uv} = \frac{|\Gamma(u) \cap \Gamma(v)|}{\max\{1, \min(d_u, d_v) - 1\}} \quad (1)$$

式 (1) 的分子等于包含边 (u, v) 的三角形数，分母以度较小端除去另一端后的邻居数为参照。当度较小端的其余邻居均为另一端的邻居时，该边所涉及的一阶邻域达到最大重合，此时 $r_{uv} = 1$ 。当一端为叶节点时，边 (u, v) 无法参与三角闭合，式中分母取 1 并得到 $r_{uv} = 0$ 。因此， r_{uv} 刻画该边实际参与三角闭合的程度，其计算范围限于一跳邻接信息。

综合目标邻居的连接能力和边冗余程度，本文定义节点 u 经由邻居 v 的传播方向权重为

$$w_{u \rightarrow v} = d_v \exp(-r_{uv}), \quad v \in \Gamma(u) \quad (2)$$

其中 d_v 刻画传播到达 v 后的后续连接能力， $\exp(-r_{uv})$ 按边冗余程度削减方向权重。由于目标邻居的度可能不同，无向边两端对应的传播方向权重一般不同。

对节点 u ，本文将传播方向权重换算为方向权重分布

$$p_{u \rightarrow v} = \frac{w_{u \rightarrow v}}{\sum_{x \in \Gamma(u)} w_{u \rightarrow x}}, \quad v \in \Gamma(u) \quad (3)$$

$p_{u \rightarrow v}$ 表示节点 u 的传播潜力分配到方向 $u \rightarrow v$ 的比例。基于香农信息熵 [49]，本文定义节点 u 的传播方向熵为

$$H_u = - \sum_{v \in \Gamma(u)} p_{u \rightarrow v} \ln p_{u \rightarrow v} \quad (4)$$

根据 Hill 提出的有效数表示 [50]，本文将传播方向熵转化为有效方向指数 $\Omega_u = \exp(H_u)$ 。该指数把方向权重分布的熵换算为具有相同熵的等权方向数。由于式 (2) 已经降低冗余方向的权重， Ω_u 越大，表明节点 u 的传播潜力分布在越多低冗余方向上。传播权重集中于少数方向时， Ω_u 随之减小。

节点度是度中心性 [14] 的基础量。本文综合节点基础连接能力和有效方向指数，定义节点 u 的初始传播得分为

$$F_u^{(0)} = \ln(d_u + e) \Omega_u \quad (5)$$

其中 $\ln(d_u + e)$ 刻画节点的基础连接能力。对度取对数可以减小大度节点与其余节点之间的量级差异，使有效方向指数 Ω_u 参与初始排序。对孤立节点，规定 $F_u^{(0)} = 0$ 。

2.2 动态得分更新

仅按初始传播得分排序，易在局部密集区选出传播范围重合的种子。为降低多种子之间的传播冗余，本文设计传播重叠抑制策略。每轮选出新种子后，算法根据相邻候选节点与该种子的传播方向重合程度调整候选得分。

设第 t 轮选种前的种子集合为 S_{t-1} ，初始时 $S_0 = \emptyset$ 。第 t 轮选择当前得分最高的未选节点作为新种子

$$s_t = \arg \max_{u \in V \setminus S_{t-1}} F_u^{(t-1)}, \quad S_t = S_{t-1} \cup \{s_t\} \quad (6)$$

已选种子不再参与后续竞争，令 $F_{s_t}^{(t)} = 0$ 。对新种子 s_t 的邻居候选节点 $u \in \Gamma(s_t) \setminus S_t$ ，本文定义其与 s_t 的传播方向重合比例为

$$\phi_{u, s_t} = \frac{w_{u \rightarrow s_t} + \sum_{x \in \Gamma(u) \cap \Gamma(s_t)} w_{u \rightarrow x}}{\sum_{x \in \Gamma(u)} w_{u \rightarrow x}} \quad (7)$$

式 (7) 的分子为候选节点 u 指向新种子 s_t 及二者公共邻居的传播方向权重之和，分母为 u 的全部传播方向权重。因此， ϕ_{u,s_t} 表示候选节点 u 的传播方向中，落入 s_t 局部覆盖范围的权重比例。 ϕ_{u,s_t} 越大， u 与 s_t 的局部传播冗余越高。

局部聚类系数 C_u 刻画节点 u 邻域内三角闭合的程度 [6]。 C_u 较大时，候选节点的邻居连接更密集，从候选节点沿不同邻居展开的传播更容易在局部邻域内到达相同节点。本文据此定义聚类系数调节指数和折扣因子

$$\alpha_u = \frac{1}{1 + C_u}, \quad P_{u,s_t} = 1 - \phi_{u,s_t}^{\alpha_u} \quad (8)$$

式 (8) 满足 $0 \leq P_{u,s_t} \leq 1$ 。当 $\phi_{u,s_t} = 0$ 时，候选节点的传播方向与新种子的局部覆盖范围没有重合， $P_{u,s_t} = 1$ ，候选得分保持不变。当 $\phi_{u,s_t} = 1$ 时，候选节点的传播方向全部落入新种子的局部覆盖范围， $P_{u,s_t} = 0$ 。在 $0 < \phi_{u,s_t} < 1$ 时，局部聚类系数增大使 α_u 减小， $\phi_{u,s_t}^{\alpha_u}$ 随之增大，候选节点保留的得分进一步减少。当 $C_u = 0$ 时， $\alpha_u = 1$ ，折扣因子化为 $P_{u,s_t} = 1 - \phi_{u,s_t}$ 。

本文将候选节点 u 的得分更新为

$$F_u^{(t)} = F_u^{(t-1)} P_{u,s_t}, \quad u \in \Gamma(s_t) \setminus S_t \quad (9)$$

其他未选候选节点得分不变。由式 (6) 至式 (9) 可知，候选节点与新种子的传播方向重合比例越大，其得分下降越多。传播方向没有重合时，候选得分保持不变。

PDRS 的算法流程如表1所示。

表 1 PDRS 算法流程
Table 1. Procedure of the PDRS algorithm

Input	$G = (V, E), K$
Output	S
1	// 初始化节点邻域与度值
2	$S \leftarrow \emptyset$
3	foreach $u \in V$ do
4	$\Gamma(u) \leftarrow \{v \in V \mid (u, v) \in E\}; \quad d_u \leftarrow \Gamma(u) $
5	end foreach
6	// 构造边冗余程度与传播方向权重
7	foreach $(u, v) \in E$ do
8	$r_{uv} \leftarrow \Gamma(u) \cap \Gamma(v) / \max\{1, \min(d_u, d_v) - 1\}$
9	$w_{u \rightarrow v} \leftarrow d_v \exp(-r_{uv})$
10	$w_{v \rightarrow u} \leftarrow d_u \exp(-r_{uv})$
11	end foreach
12	// 由方向权重分布计算初始传播得分
13	foreach $u \in V$ do
14	if $d_u = 0$ then
15	$F_u \leftarrow 0$
16	else
17	$W_u \leftarrow \sum_{x \in \Gamma(u)} w_{u \rightarrow x}$
18	foreach $v \in \Gamma(u)$ do
19	$p_{u \rightarrow v} \leftarrow w_{u \rightarrow v} / W_u$
20	end foreach
21	$H_u \leftarrow - \sum_{v \in \Gamma(u)} p_{u \rightarrow v} \ln p_{u \rightarrow v}$
22	$\Omega_u \leftarrow \exp(H_u)$
23	$F_u \leftarrow \ln(d_u + e) \Omega_u$
24	end if
25	end foreach
26	// 迭代选择种子节点
27	for $t = 1$ to K do
28	$s_t \leftarrow \arg \max_{u \in V \setminus S} F_u$ // 选择当前得分最高的未选节点
29	$S \leftarrow S \cup \{s_t\}$
30	$F_{s_t} \leftarrow 0$ // 已选节点不再参与后续竞争
31	// 按传播方向重合程度更新候选得分
32	foreach $u \in \Gamma(s_t) \setminus S$ do // 仅更新新种子的邻居候选节点
33	$\phi_{u, s_t} \leftarrow [w_{u \rightarrow s_t} + \sum_{x \in \Gamma(u) \cap \Gamma(s_t)} w_{u \rightarrow x}] / W_u$
34	$C_u \leftarrow 0$
35	if $d_u \geq 2$ then
36	$e_u \leftarrow \{(a, b) \in E \mid a, b \in \Gamma(u)\} $ // 统计邻居间连边数
37	$C_u \leftarrow 2e_u / [d_u(d_u - 1)]$
38	end if
39	$\alpha_u \leftarrow 1 / (1 + C_u)$
40	$P_{u, s_t} \leftarrow 1 - \phi_{u, s_t}^{\alpha_u}$
41	$F_u \leftarrow F_u \cdot P_{u, s_t}$
42	end foreach
43	end for
44	return S

图1给出 PDRS 在 DD、CA-GrQc、CA-HepPh 和 hep-th 网络上选出的种子节点。

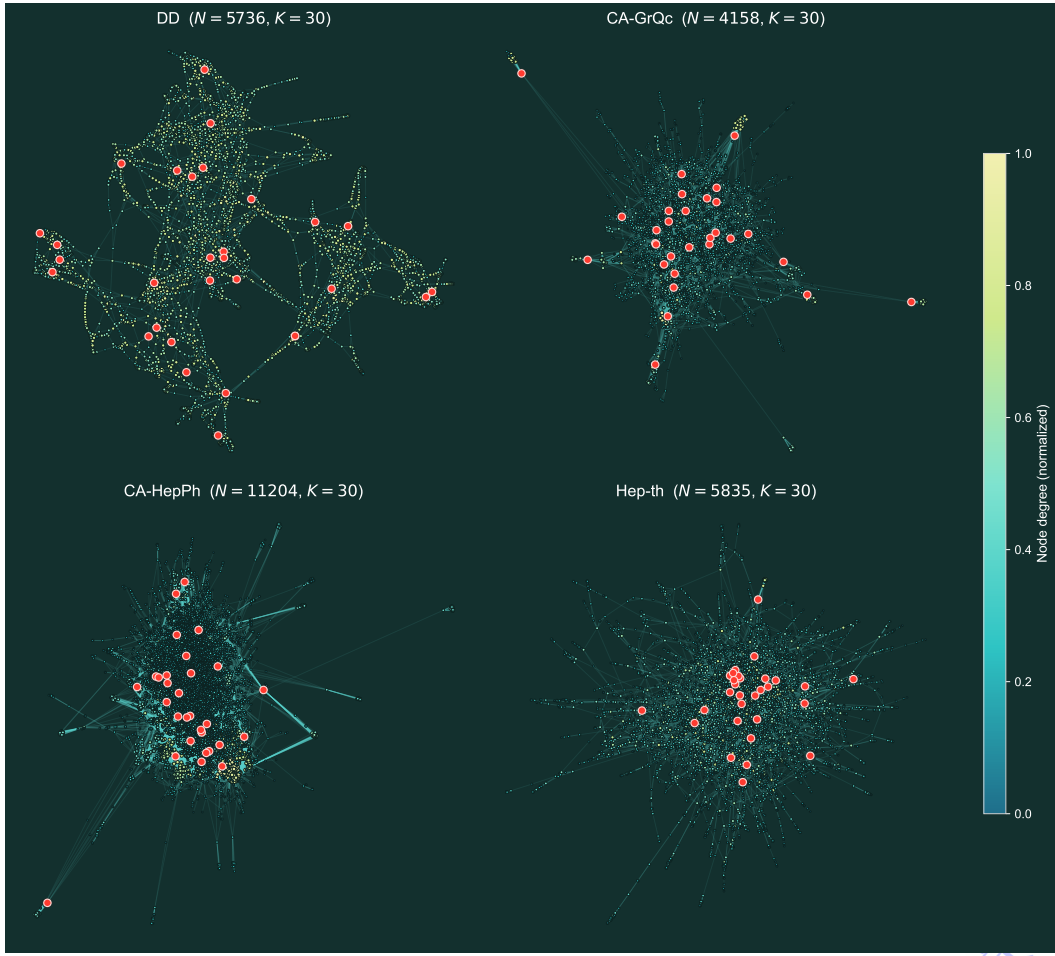


图 1 PDRS 在 4 个真实网络上选出的种子节点分布，红色节点为所选种子

Fig. 1. Distribution of seeds selected by PDRS on four real-world networks; red nodes denote the selected seeds.

图中 30 个种子节点分布于多个结构区域。第 4.2 节和第 4.6 节进一步使用种子节点平均最短路径长度及其它结构指标进行定量分析。

3 实验设计

主实验在 SIR 模型下比较 PDRS 与 13 种基线算法。补充实验用于分析传播模型、参数设置和 PDRS 两个组成部分对结果的影响。各项实验均采用表3所列的种子节点数量设置。

3.1 实验网络与对比算法

本文选用 9 个来自不同领域的真实网络进行实验，分别是社会关系网络 Physicians、在线社交网络 Hamsterster 和 Hamster、接触传播网络 Infect-Dublin、邮件通信网络 Email-DNC、投票关系网络 Wiki-Vote、消息通信网络 CollegeMsg 以及科研合作网络 CA-GrQc 和 CA-HepPh。各网络均转化为无向无权简单图，删除自环和重复边后取最大连通分量。各网络的基本统计特征如表2所列，其中 N 和 M 分别为节点数和边数， $\langle k \rangle$ 为平均度， $k_{\max}$ 为最大度， $\langle C \rangle$ 为平均聚类系数， D 为网络直径， $\langle L \rangle$ 为平均最短路径长度。

表 2 真实网络基本统计特征
Table 2. Basic statistics of real-world networks

网络	类型	N	M	$\langle k \rangle$	$k_{\max}$	$\langle C \rangle$	D	$\langle L \rangle$
Physicians	社交	117	465	7.9487	26	0.2192	5	2.5871
Hamsterster	社交	2000	16097	16.0970	273	0.5400	10	3.5890
Infect-Dublin	接触	410	2765	13.4878	50	0.4558	9	3.6309
Email-DNC	通信	849	10384	24.4617	368	0.5072	8	2.7595
Wiki-Vote	投票	889	2914	6.5557	102	0.1528	13	4.0962
CollegeMsg	通信	1893	13835	14.6170	255	0.1097	8	3.0552
Hamster	社交	1788	12476	13.9553	272	0.1433	14	3.4526
CA-GrQc	合作	4158	13422	6.4560	81	0.5569	17	6.0494
CA-HepPh	合作	11204	117619	20.9959	491	0.6216	13	4.6727

本文选取度中心性 (Degree) [14]、K-shell [19]、PageRank [17]、Degree Discount [12]、VoteRank++ [35]、EWV [36]、DILVoteRank [37]、IMVoteRank [38]、CVoteRank [39]、EnRenew [31]、TsallisRank [28]、DomiRank [43] 和 NOIRX [40] 作为对比方法。

3.2 传播模型与参数设置

主对比实验采用 SIR 模型评价不同算法所得种子集合的传播能力。初始时， K 个种子节点处于感染态，其余节点处于易感态。在每个离散时间步内，感染节点分别以概率 β 尝试感染每个易感邻居，并以概率 μ 转为恢复态。恢复节点不再参与后续传播。当网络中不存在感染节点时，传播过程结束。

本文取 $\mu = 1$ [29,38,52]，每个感染节点在一个时间步内完成对易感邻居的传播尝试，随后转为恢复态。该设置使所有感染节点具有相同的一个时间步感染期， β 由此直接表示该时间步内沿给定邻接边传播成功的概率。已有单节点和多节点传播能力评价研究也采用这一离散时间设置 [29,38,52]。

Newman 基于网络 SIR 传播与键渗流的对应关系指出，对于由给定度分布生成、无度相关且局部树状的无限随机网络，沿边传播概率的理论阈值为 [51]

$$\tau_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (10)$$

其中 $\langle k \rangle$ 和 $\langle k^2 \rangle$ 分别为节点度的一阶矩和二阶矩。本文将 τ_c 称为 SIR 理论传播阈值，并用它为不同度分布的网络设置感染率。真实网络具有有限规模、聚类和度相关性，其实际传播转变位置可以偏离该理论值 [47]。主传播实验的基准感染率取

$$\beta_0 = \max\{0.03, 1.2\tau_c\} \quad (11)$$

系数 1.2 使感染率略高于 SIR 理论传播阈值。感染率下限 0.03 用于避免度异质性较强的网络得到过小的基准感染率。第 4.3 节通过阈值倍数和绝对感染率的多取值比较，分析两个参数对应的非种子节点覆盖及不同种子集合之间的传播差异。

每个网络在其种子节点数量范围内由小到大取 20 个值，由 9 个网络和各 20 个种子节点数量构成 180 组实验条件。各网络的具体范围见表3。主对比实验在相同网络、种子节点数量和传播参数下比较全部算法，每组实验条件独立运行 $R = 1000$ 次 SIR 传播，取最终感染比例的平均值作为评价值。

感染率敏感性实验在上述 180 组条件上令 $m = \beta/\beta_0$ ，并取 $m = 0.5, 0.625, 0.75, 0.875, 1, 1.125, 1.25, 1.375$ 和 1.5。每个条件进行 1000 次 SIR 传播。

为考察 PDRS 所得种子集合在不同传播机制下的表现，本文进一步采用独立级联（independent cascade, IC）模型和线性阈值（linear threshold, LT）模型。IC 模型中，新激活节点以概率 β_0 对每个尚未激活的邻居进行一次独立激活尝试。LT 模型中，节点 v 从每个邻居接收的影响权重取 $1/d_v$ ，节点阈值在区间 $[0, 1]$ 上独立服从均匀分布。当已激活邻居的权重之和达到节点阈值时，该节点被激活。实验沿用 9 个无向网络及表3所列的种子节点数量设置，每个条件分别进行 1000 次独立传播。

表 3 不同网络的种子节点数量范围
Table 3. Seed-size ranges for different networks

网络	种子节点数量 K 范围
Physicians	3 至 22
Hamsterster	10 至 200
Infect-Dublin	10 至 62
Email-DNC	5 至 85
Wiki-Vote	5 至 89
CollegeMsg	10 至 189
Hamster	9 至 179
CA-GrQc	42 至 416
CA-HepPh	112 至 1120

3.3 评价指标

本文从传播效果、种子分布结构与计算效率三方面评价各算法。平均最终感染比例用于衡量种子集合在 SIR 模型下的传播覆盖，种子集合相似性、种子节点平均最短路径长度和社区覆盖率用于刻画种子选

取差异与空间分布，选种时间用于评价计算开销。

3.3.1 传播范围评价

设算法在网络 $G = (V, E)$ 上选出的种子集合为 S ， $|S| = K$ 。第 r 次 SIR 模拟的终止时刻记为 T_r ，即网络中不再存在感染节点的时刻。若 $R_r(T_r)$ 表示第 r 次模拟终止时的恢复节点集合，则种子集合 S 的平均最终感染比例 [32,38] 定义为

$$\sigma(S) = \frac{1}{R} \sum_{r=1}^R \frac{|R_r(T_r)|}{N} \quad (12)$$

其中 R 为独立模拟次数， $N = |V|$ 为网络节点数。本文以 $\sigma(S)$ 作为评价种子集合传播效果的主要指标。同一网络和同一种子节点数量下， $\sigma(S)$ 越大，种子集合在 SIR 传播中覆盖的节点比例越高。

本文统计各算法在全部网络与种子节点数量组合中取得最大平均最终感染比例的次数。PDRS 与各基线的统计差异在相同网络、种子节点数量和传播参数下逐组计算。同一条件下的第 i 次 SIR 传播对所有算法使用相同随机种子。设 PDRS 与基线算法 B 在第 i 次传播中的最终感染比例分别为 x_i^{PDRS} 和 x_i^B ，差值为 $d_i = x_i^{\text{PDRS}} - x_i^B$ 。若 d_i 均值的 95% 置信区间下界大于 0，则 PDRS 显著优于该基线。若置信区间上界小于 0，则 PDRS 显著差于该基线。其余情况记为差异不显著。

3.3.2 选种差异与结构分布

本文首先计算算法间种子集合的相似性 [48]。设算法 A 和 B 在网络 G 上、种子节点数量为 K 时选出的种子集合分别为 $S_A(G, K)$ 和 $S_B(G, K)$ 。同一网络和同一种子节点数量下各算法的种子集合规模相同，定义二者的种子集合相似性为

$$\theta(A, B; G, K) = \frac{|S_A(G, K) \cap S_B(G, K)|}{K} \quad (13)$$

$\theta(A, B; G, K) = 1$ 时两算法的种子集合完全相同， $\theta(A, B; G, K) = 0$ 时无共同种子。该指标衡量相同实验条件下各算法选种结果的相似程度。

为避免网络规模和种子节点数量范围的差异改变总体均值，本文先在每个网络内对不同 K 的种子集合相似性取平均，再对全部网络取平均，使每个网络在总体统计中具有相同权重。设真实网络集合为 $\mathcal{G}$ ，网络 G 上的种子节点数量集合为 $\mathcal{K}_G$ ，则算法 A 和 B 的平均种子集合相似性定义为

$$\bar{\theta}(A, B) = \frac{1}{|\mathcal{G}|} \sum_{G \in \mathcal{G}} \frac{1}{|\mathcal{K}_G|} \sum_{K \in \mathcal{K}_G} \theta(A, B; G, K) \quad (14)$$

本文还从空间分布和社区覆盖两个角度分析种子集合的结构特征。设 $d(u, v)$ 为节点 u 与 v 的最短路径长度，则种子节点平均最短路径长度 [32,38] 定义为

$$L_S = \frac{2}{K(K-1)} \sum_{\substack{u, v \in S \\ u < v}} d(u, v) \quad (15)$$

由于所有网络均取最大连通分量，任意两个种子节点之间均连通， L_S 刻画种子在网络中的空间分散程度。同一网络和同一种子节点数量下， L_S 越大，种子集合分布越分散。

设网络经 Louvain 方法 [44] 划分为 N_c 个社区，节点 u 所属社区为 $c(u)$ ，则种子集合的社区覆盖率定义为

$$Q_S = \frac{|\{c(u) \mid u \in S\}|}{N_c} \quad (16)$$

Q_S 刻画种子集合覆盖的社区范围， Q_S 越大，种子分布于越多社区。

3.3.3 计算效率指标

为比较各算法的计算开销，本文记录其在相同环境下的种子选择时间，记为 T_{sel} ，仅统计节点得分计算、排序和迭代选种的耗时。

3.4 消融实验设置

为区分有效方向指数和传播重叠抑制策略对算法结果的作用，本文按照 PDRS 的构造顺序设置 Base、Static-PDRS 和 PDRS 三个版本。Base 采用基础连接得分 $B_u = \ln(d_u + e)$ 并按得分静态排序。Static-PDRS 采用 $B_u \Omega_u$ 作为初始得分并静态排序，用于考察加入有效方向指数后节点排序和种子集合的变化。PDRS 在 Static-PDRS 的初始得分上继续执行式 (7) 至式 (9) 的动态更新。三个版本的组成如表4所示。

表 4 三种消融版本的组成
Table 4. Components of the three ablation variants

版本	初始得分	选种与更新
Base	$B_u = \ln(d_u + e)$	静态排序
Static-PDRS	$B_u \Omega_u$	静态排序
PDRS	$B_u \Omega_u$	传播重叠抑制策略

消融实验沿用 9 个无向网络及表3所列的种子节点数量设置，每个条件进行 1000 次 SIR 传播。传播性能采用平均最终感染比例评价，种子集合结构采用相邻种子对比例、闭邻域平均 Jaccard 系数、一阶邻域覆盖效率、种子节点平均最短路径长度和社区覆盖率评价。记 $E[S]$ 为种子集合诱导子图中的边集合，

$\Gamma[u] = \Gamma(u) \cup \{u\}$ 为节点 u 的闭邻域。前三项指标定义为

$$\begin{aligned} A_S &= \frac{|E[S]|}{\binom{K}{2}}, \\ J_S &= \frac{1}{\binom{K}{2}} \sum_{\substack{u,v \in S \\ u < v}} \frac{|\Gamma[u] \cap \Gamma[v]|}{|\Gamma[u] \cup \Gamma[v]|}, \\ U_S &= \frac{|\bigcup_{u \in S} \Gamma[u]|}{\sum_{u \in S} |\Gamma[u]|}. \end{aligned} \quad (17)$$

A_S 表示全部种子节点对中直接相邻的比例。 J_S 表示全部种子节点对闭邻域 Jaccard 系数的平均值。 U_S 表示不同种子闭邻域合并后的节点数占各闭邻域节点数之和的比例。 A_S 和 J_S 减小、 U_S 增大均表示种子节点之间的一阶邻域覆盖重合减少。种子节点平均最短路径长度 L_S 和社区覆盖率 Q_S 沿用式 (15) 和式 (16)。

4 无向网络上的实验结果与分析

4.1 传播范围比较

为评价各算法所选种子集合的传播能力，图2给出不同算法在 9 个真实网络上的 SIR 传播范围比较结果，横轴为种子节点数量 K ，纵轴为平均最终感染比例 $\sigma(S)$ 。

由图2可见，PDRS 在 Physicians、Infect-Dublin、Email-DNC、Hamster、Hamsterster 和 CA-HepPh 等网络上处于前列。在 CollegeMsg、Wiki-Vote 和 CA-GrQc 等网络上，VoteRank++、IMVoteRank、Degree Discount 或 DILVoteRank 与 PDRS 的传播曲线较为接近。静态排序方法与动态迭代方法的总体差异由表5和表6进一步比较。

为整体比较各算法的传播效果，本文统计了各算法在 180 组实验条件下的平均最终感染比例、平均排名和最优次数，如表5所示。最优次数表示算法在某一网络和某一种子节点数量下取得最大平均最终感染比例的次数。

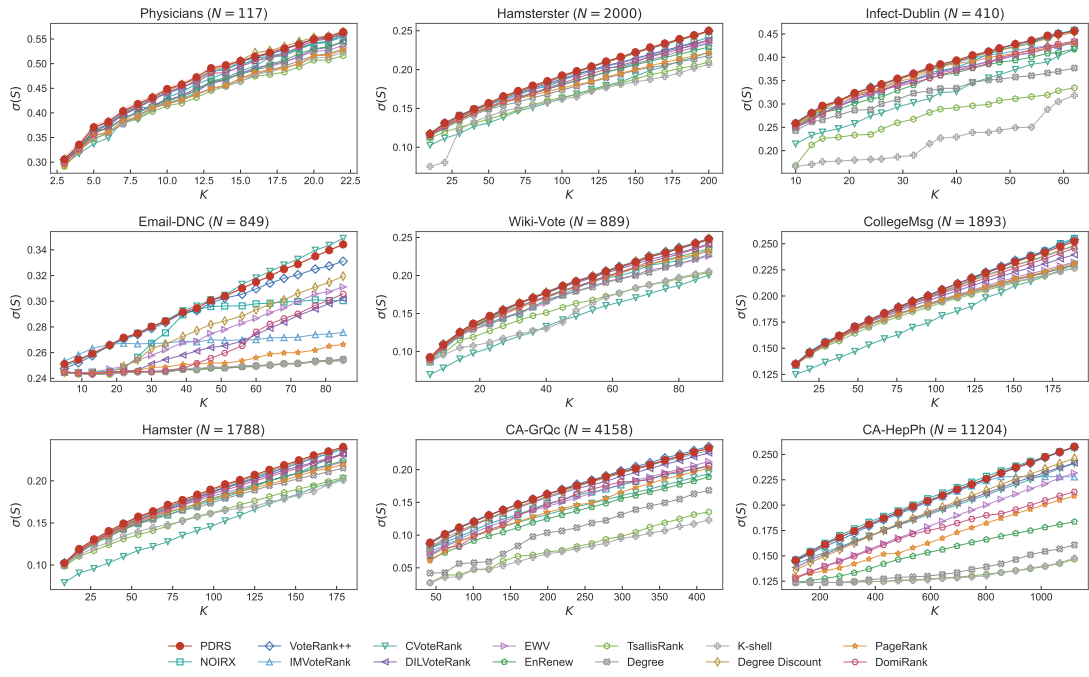


图 2 不同算法在 9 个真实网络上的 SIR 传播范围比较

Fig. 2. Comparison of SIR spreading ranges of different algorithms on nine real-world networks.

表 5 不同算法传播范围总体统计结果

Table 5. Overall statistics of spreading range for different algorithms

算法	平均最终感染比例 $\sigma(S)$	平均排名	最优次数
PDRS	0.2520	1.89	70
VoteRank++	0.2505	2.80	33
Degree Discount	0.2472	3.77	11
IMVoteRank	0.2454	3.84	27
NOIRX	0.2432	5.63	24
DILVoteRank	0.2402	6.11	0
DomiRank	0.2370	7.09	0
EWV	0.2352	8.17	0
PageRank	0.2302	8.78	0
EnRenew	0.2273	9.49	1
CVoteRank	0.2241	10.27	14
Degree	0.2166	11.06	0
K-shell	0.1958	12.77	0
TsallisRank	0.2030	12.92	0

由表5可见，PDRS 的平均最终感染比例、平均排名和最优次数三项指标均居首位，其中最优次数约为 VoteRank++ 的两倍。与传统静态方法相比，PDRS 相对 PageRank、Degree 和 K-shell 的平均最终感染比例提高约 9.47% 至 28.70%，相对 TsallisRank 和 EnRenew 也提高约 10.87% 至 24.14%。这些结果表明，单节点静态排序得到的种子集合在当前 SIR 实验中的传播范围整体低于 PDRS。

VoteRank++、IMVoteRank、Degree Discount、DILVoteRank 和 NOIRX 等动态迭代方法的平均最终感染比例普遍高于静态方法。PDRS 相对 NOIRX 的平均最终感染比例提高约 3.62%，平均排名和最优次数也更高。PDRS 相对 VoteRank++ 和 IMVoteRank 的平均最终感染比例分别提高约 0.60% 和 2.69%，

三种方法的平均最终感染比例较为接近。由于平均值不能反映逐个网络与种子节点数量条件下的差异，本文进一步对 PDRS 与各基线的 1000 次 SIR 结果进行逐组比较，统计显著更优、显著更差和差异不显著的次数，如表6所示。

表 6 PDRS 与基线算法传播范围差异的显著性统计

Table 6. Significance statistics of spreading-range differences between PDRS and baseline algorithms

基线算法	PDRS 显著更优	PDRS 显著更差	差异不显著
TsallisRank	180	0	0
K-shell	178	0	2
EnRenew	178	0	2
Degree	177	0	3
PageRank	177	0	3
EWV	177	0	3
DomiRank	176	0	4
DILVoteRank	172	0	8
CVoteRank	162	10	8
NOIRX	143	24	13
Degree Discount	141	13	26
IMVoteRank	123	18	39
VoteRank++	92	27	61

由表6可见，PDRS 相对 TsallisRank、K-shell、EnRenew、PageRank、Degree、EWV、DomiRank 和 DILVoteRank 显著更优的次数均不少于 172 次，并且没有出现显著更差的条件。相对 Degree Discount、NOIRX、IMVoteRank 和 VoteRank++，PDRS 显著更优、显著更差和差异不显著的条件均有出现，说明这些动态方法与 PDRS 的传播结果更为接近。

分网络观察时，PDRS 在 Hamster、Hamsterster、Infect-Dublin 和 Physicians 上分别取得 19、18、14 和 9 次最优。在 CA-GrQc 上，VoteRank++ 取得 12 次最优，PDRS 取得 1 次最优。在 CA-HepPh 上，NOIRX 取得 17 次最优。PDRS 相对 NOIRX 出现的 24 次显著更差主要来自 CA-HepPh。这些结果说明 PDRS 与强动态方法之间的差异随网络改变。网络结构对 PDRS 相对表现的影响将在第 6 节结合平均度、节点核数和传递性进一步讨论。

4.2 选种相似性与结构分布分析

本节从种子集合相似性、种子节点平均最短路径长度和社区覆盖率三方面分析各算法的种子集合。图3给出不同算法之间的种子集合相似性。

由图3可见，Degree、EnRenew、TsallisRank 和 K-shell 等方法之间的种子集合相似性较高，说明不同静态评分也可能产生相近的种子集合。PDRS 与这些方法的相似性较低，表明 PDRS 的迭代得分更新改变了部分后续种子的选择。

PDRS 与 VoteRank++、IMVoteRank、Degree Discount 和 DILVoteRank 具有中等程度的种子集合

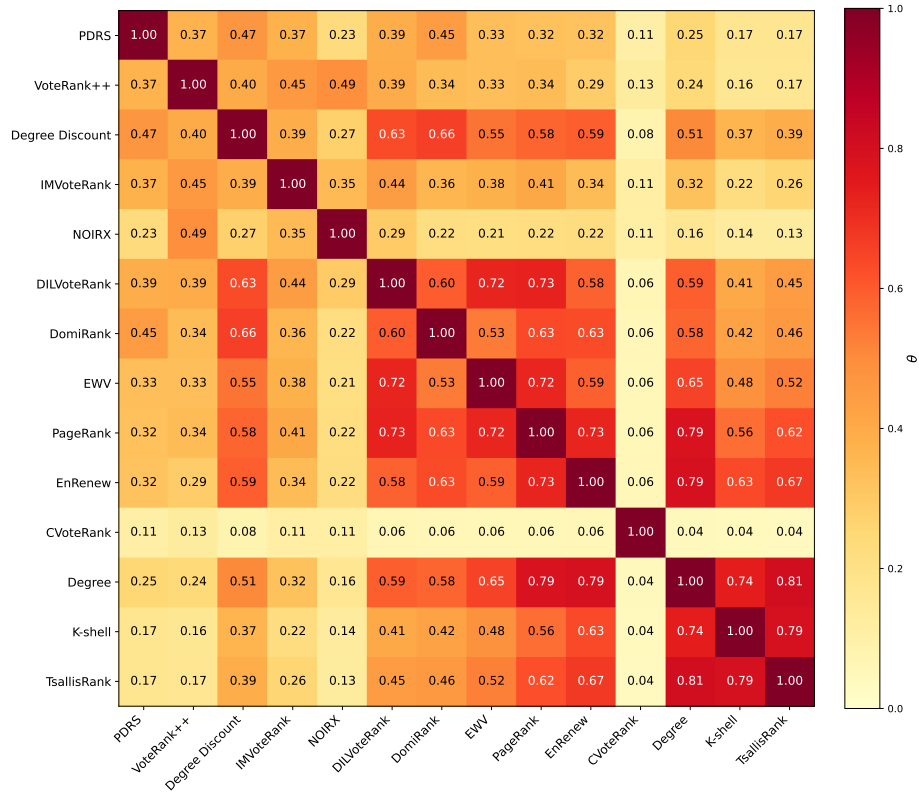


图 3 不同算法的种子集合相似性热力图

Fig. 3. Heatmap of seed-set similarity between different algorithms.

相似性。NOIRX 与 PDRS 的相似性约为 0.23，NOIRX 与 VoteRank++ 的相似性约为 0.49。种子集合相似性只比较选中节点是否相同，不能据此判定算法机制相同。结合图2可见，不同迭代规则可以得到组成不同但传播范围接近的种子集合。

图4给出不同算法的种子节点平均最短路径长度 L_S 随种子节点数量 K 的变化结果。

PDRS 的 L_S 整体高于 Degree、K-shell 和 TsallisRank 等静态或熵类方法。180 组结果中，PDRS 的平均 L_S 相比 Degree、K-shell 和 TsallisRank 分别提高约 46.7%、52.6% 和 56.3%，表明 PDRS 所得种子集合在网络中分布得更为分散。传播重叠抑制策略对该结构差异的作用由第 4.6 节消融实验进一步分析。

L_S 需要与平均最终感染比例联合评价。CVoteRank 的种子节点平均最短路径长度最高，其平均最终感染比例低于 PDRS、VoteRank++ 和 IMVoteRank。NOIRX 的平均最短路径长度约为 3.24，高于 PDRS 的约 2.96，其平均最终感染比例也低于 PDRS。较大的 L_S 表明种子分布更分散，传播结果还取决于种子自身传播能力及其所处的网络位置。因此， L_S 用于描述种子分散程度，不单独作为传播性能的判据。

图5以箱线图形式给出各算法社区覆盖率 Q_S 的分布，每个箱体汇总对应算法在 9 个网络、20 个种子节点数量共 180 组实验下的取值，按 Q_S 中位数由高到低自左向右排列。

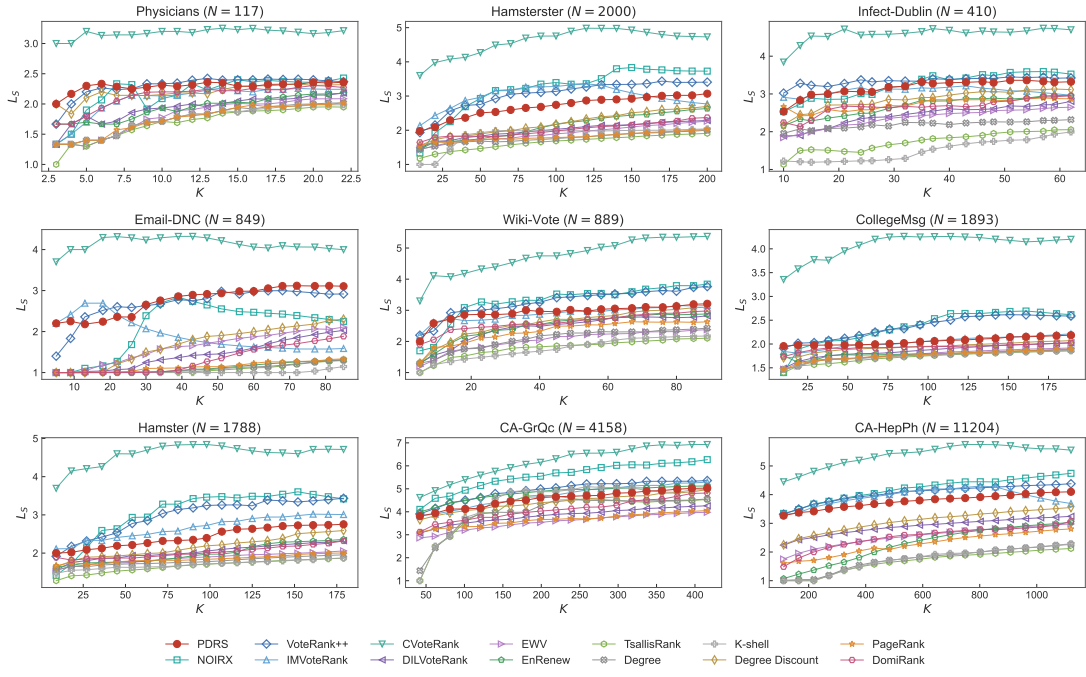


图 4 不同算法的种子节点平均最短路径长度比较

Fig. 4. Comparison of average shortest path length among seed nodes selected by different algorithms.

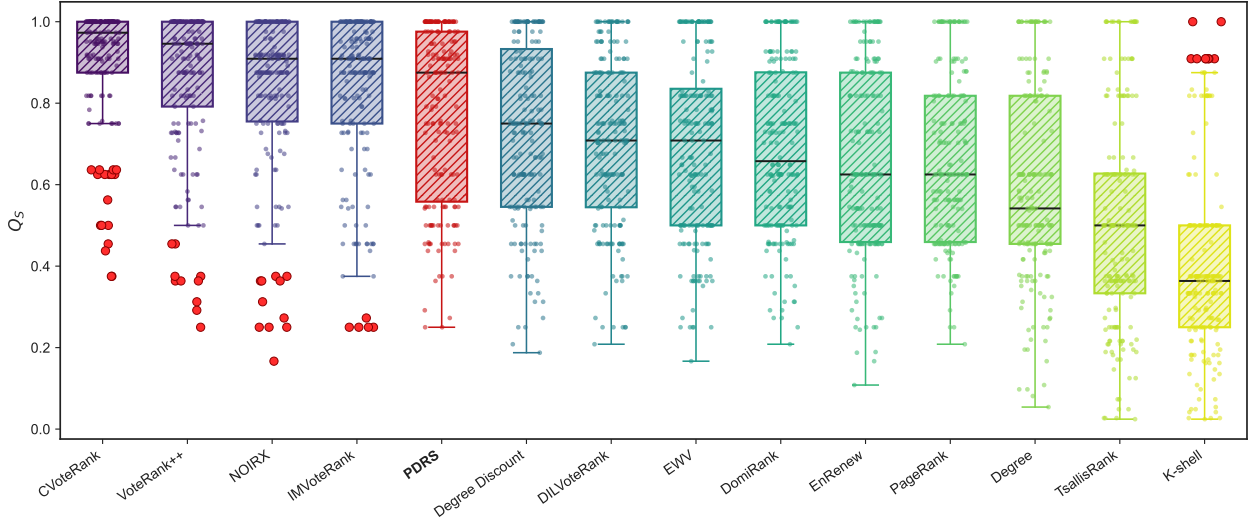


图 5 不同算法社区覆盖率 Q_s 的分布比较

Fig. 5. Distribution of community coverage Q_s of seed sets selected by different algorithms.

由图5可见, CVoteRank、VoteRank++、IMVoteRank 和 NOIRX 的社区覆盖率整体较高, K-shell 和 TsallisRank 的社区覆盖率较低, PDRS 位于中部且 Q_s 均值约为 0.78。各算法的 Q_s 随网络和种子节点数量改变。CVoteRank 的 Q_s 均值约为 0.91, NOIRX 的 Q_s 均值约为 0.84, 二者均高于 PDRS, 但平均

最终感染比例低于 PDRS。较高的社区覆盖率表示种子集合涉及更多社区，传播结果还取决于所选节点在各社区内的传播能力及其局部覆盖重合。

PDRS 所得种子集合与静态排序和熵类方法的相似性较低，与多种动态方法具有中等相似性，其种子节点平均最短路径长度高于静态排序方法。部分社区导向方法取得更高的社区覆盖率和种子节点平均最短路径长度，但其平均最终感染比例低于 PDRS。种子集合的空间分散程度和社区覆盖范围需要与种子节点自身的传播能力共同评价。

4.3 传播参数依据与敏感性分析

式 (11) 中阈值倍数 1.2 和感染率下限 0.03 的取值依据由两组多取值实验分析。对基准感染率由 $a\tau_c$ 决定的 Physicians、Infect-Dublin、Wiki-Vote 和 CA-GrQc，令 a 从 0 增加到 5，取值间隔为 0.05。对基准感染率由下限 0.03 决定的 Hamsterster、Email-DNC、CollegeMsg、Hamster 和 CA-HepPh，令 β 从 0 增加到 0.10，取值间隔为 0.001。每组实验均在表3所列的种子节点数量设置下比较 14 种算法，每个条件进行 1000 次 SIR 传播。

对前 4 个网络，传播差异 D 定义为同一网络、感染率和种子节点数量下 14 种算法新增覆盖比例的极差，再对该网络不同种子节点数量的结果取平均。以各网络全部倍数取值中的最大值 $D_{\max}$ 为参照，4 个网络同时满足 $D/D_{\max} \geq 0.5$ 的范围为 $0.95 \leq a \leq 2.75$ 。 $a = 1.2$ 位于该共同范围内，并接近 SIR 理论传播阈值对应的 $a = 1$ 。在 $a = 1.2$ 时，4 个网络保留各自最大算法差异的 65.71% 至 99.41%，平均新增覆盖比例为 9.01% 至 38.12%。该取值使传播覆盖一定比例的非种子节点，同时保留不同种子集合传播能力的差异。

后 5 个网络的平均新增覆盖比例首次达到 10% 时，对应的感染率依次为 0.025、0.014、0.023、0.026 和 0.023。 $\beta = 0.03$ 高于这 5 个取值，并使平均新增覆盖比例达到 12.48% 至 22.75%。感染率下限 0.03 由此避免理论阈值较小的网络只形成接近种子集合规模的传播结果。图6汇总两组实验结果。

为考察感染率变化对 PDRS 传播范围和相对表现的影响，本文令 $m = \beta/\beta_0$ ，在 $0.5 \leq m \leq 1.5$ 内取 9 个等间隔值。在 $0.5\beta_0$ 下，9 个网络的平均新增覆盖比例为 3.57% 至 11.93%。在 $1.5\beta_0$ 下，该比例增至 14.00% 至 57.56%。这一取值范围覆盖传播相对有限、基准传播和传播明显扩展的条件。

图7给出 PDRS 在不同相对感染率下的平均最终感染比例，以及其与同一网络和感染率下最优基线算法的平均最终感染比例之差。每条曲线均由各网络不同种子节点数量下的结果取算术平均得到。PDRS 的

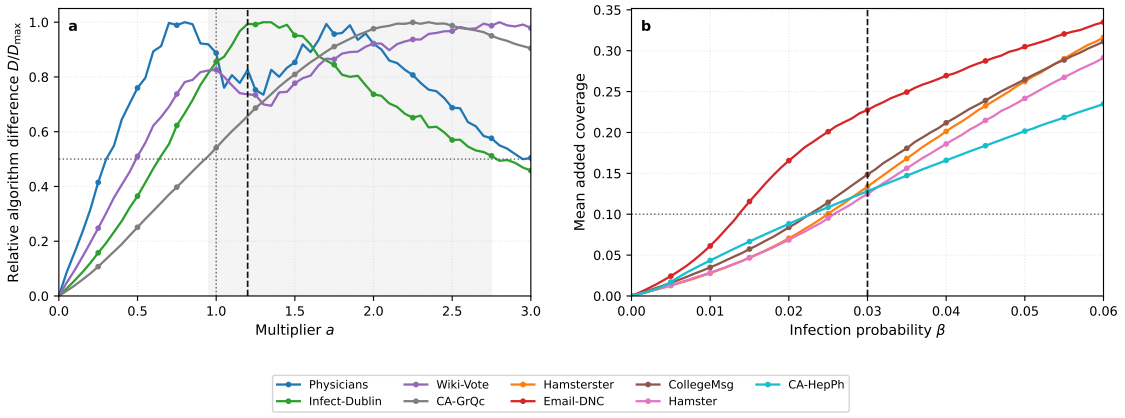


图 6 SIR 理论传播阈值倍数与感染率下限的实验结果

Fig. 6. Results for the SIR theoretical-threshold multiplier and the lower bound of infection probability. 平均最终感染比例随感染率升高而增加，其与最优基线的差值随网络和感染率变化。PDRS 在基准感染率附近的相对表现较强，感染率向两侧变化时，部分基线在相应网络上取得更高的传播范围。PDRS 的相对性能同时受网络结构和传播强度影响。

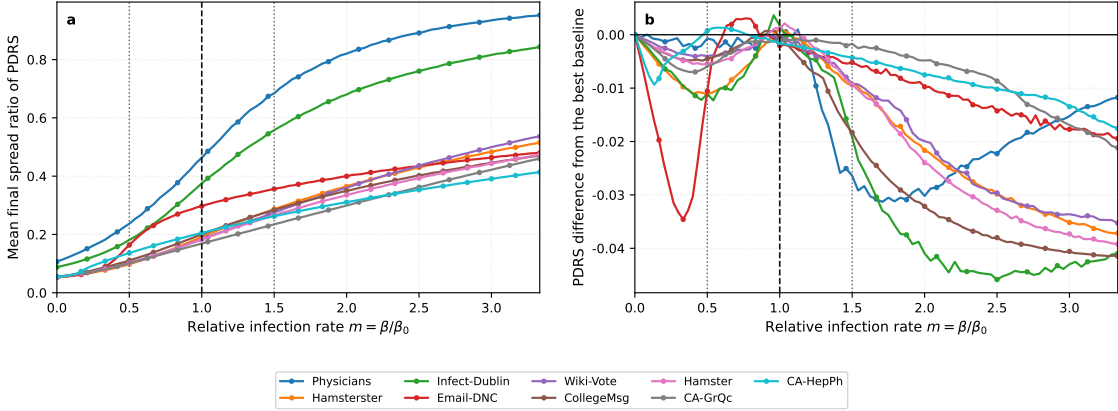


图 7 不同相对感染率下 PDRS 的平均最终感染比例及其与最优基线算法的差值

Fig. 7. Mean final infection proportion of PDRS and its difference from the best baseline at different relative infection probabilities.

4.4 三种传播模型下的传播结果

图8给出 PDRS 种子集合在 SIR、IC 和 LT 三种传播模型下的平均传播范围比例。每幅子图对应一个网络，横轴为种子节点数量，纵轴为 1000 次独立传播所得传播范围比例的平均值。

9 个网络上的 SIR 曲线与 IC 曲线接近。本文的 SIR 设置中，每个感染节点只在感染后的一个时间

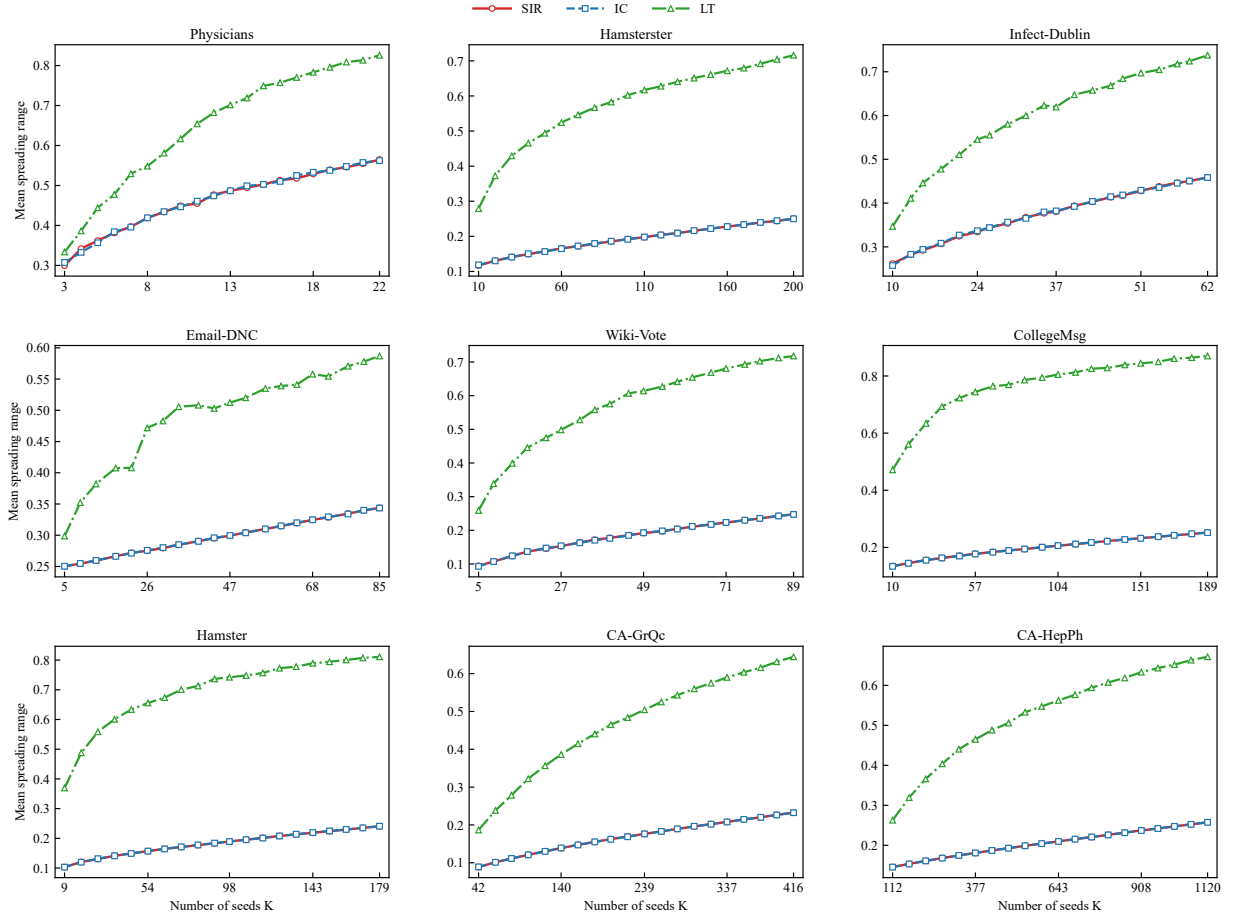


图 8 PDRS 在三种传播模型下的平均传播范围比例

Fig. 8. Mean spreading range of PDRS under three propagation models.

步内对每个易感邻居进行一次独立传播尝试。IC 模型中的新激活节点也对每个尚未激活的邻居进行一次独立激活尝试。当两种模型采用相同的边传播概率 β_0 时，若节点 v 已有 m 个感染或激活邻居完成传播尝试，则 v 至少接收一次成功传播的概率均为 $1 - (1 - \beta_0)^m$ ，两种模型具有相同的最终传播范围分布 [53]。180 组实验条件下，二者平均传播范围比例之差的绝对值平均为 0.000892，最大值为 0.009222，数值差异来自有限模拟次数下的估计波动。

LT 模型由多个已激活邻居的影响累积决定节点是否激活。若节点 v 的度为 d_v ，其中已有 m 个邻居被激活，则 v 接收的总影响为 m/d_v 。由于节点阈值在区间 $[0, 1]$ 上服从均匀分布，给定 m 个已激活邻居时的激活概率为 m/d_v 。LT 参数未与 SIR 和 IC 进行等传播强度换算，因此图 8 中的 LT 曲线同时反映传播机制和参数尺度的差异。在当前 LT 设置下，9 个网络的最终激活范围均高于 SIR 和 IC 模型的传播范围。

三种传播模型给出了同一组 PDRS 种子集合在不同传播机制下的传播范围。PDRS 相对其它算法的

性能结论以第 4.1 节具有完整基线比较的 SIR 实验为依据。本文实验对象为静态无权网络，加权网络、时序网络、多层网络、异质传播概率和真实干预场景仍需进一步验证。

4.5 选种时间分析

表7给出各算法在 180 组实验条件下的选种时间。该时间包括节点得分计算、排序和迭代选种，不含 SIR 传播。

表 7 不同算法的选种时间统计
Table 7. Seed-selection time statistics

算法	平均选种时间/s	最大选种时间/s
Degree	0.0030	0.0169
K-shell	0.0133	0.1150
PageRank	0.0187	0.1267
Degree Discount	0.4038	5.2984
NOIRX	0.6203	5.1231
DomiRank	0.9053	6.6225
DILVoteRank	1.0214	12.8956
VoteRank++	1.8162	24.8751
PDRS	2.1732	37.0180
IMVoteRank	3.0648	24.3576
CVoteRank	3.3210	29.7282
EWV	3.8328	44.7956
EnRenew	3.9869	67.4102
TsallisRank	19.9228	197.9433

由表7可见，Degree、K-shell 和 PageRank 只需一次性计算指标并排序，平均选种时间均低于 0.02 s。Degree Discount、NOIRX、DILVoteRank、DomiRank 和 VoteRank++ 需要在迭代选种中更新候选得分、支配关系或投票状态，平均选种时间为 0.40 至 1.82 s。

PDRS 的平均选种时间为 2.1732 s，高于上述方法，低于 IMVoteRank、CVoteRank、EWV、EnRenew 和 TsallisRank。其额外耗时来自公共邻居统计、方向权重构造和迭代得分更新。结合表5可见，PDRS 在当前数据规模和实现下取得最高的平均最终感染比例，平均选种时间低于 5 种对比方法。式 (1) 使用一跳邻接信息计算局部传播冗余，使结构计算保持在局部邻域内。

4.6 消融实验结果与分析

图9给出 Base、Static-PDRS 和 PDRS 在 9 个网络上的传播曲线。Base 与 Static-PDRS 的曲线在多数网络上接近，PDRS 的传播曲线在 9 个网络上均位于 Static-PDRS 之上。

表8将各网络不同种子节点数量下的平均最终感染比例差值取平均，数值单位为百分点。Static-PDRS 减 Base 表示加入有效方向指数后的变化。PDRS 减 Static-PDRS 表示继续加入传播重叠抑制策略后的变化。PDRS 减 Base 表示完整算法相对于基础版本的总变化。

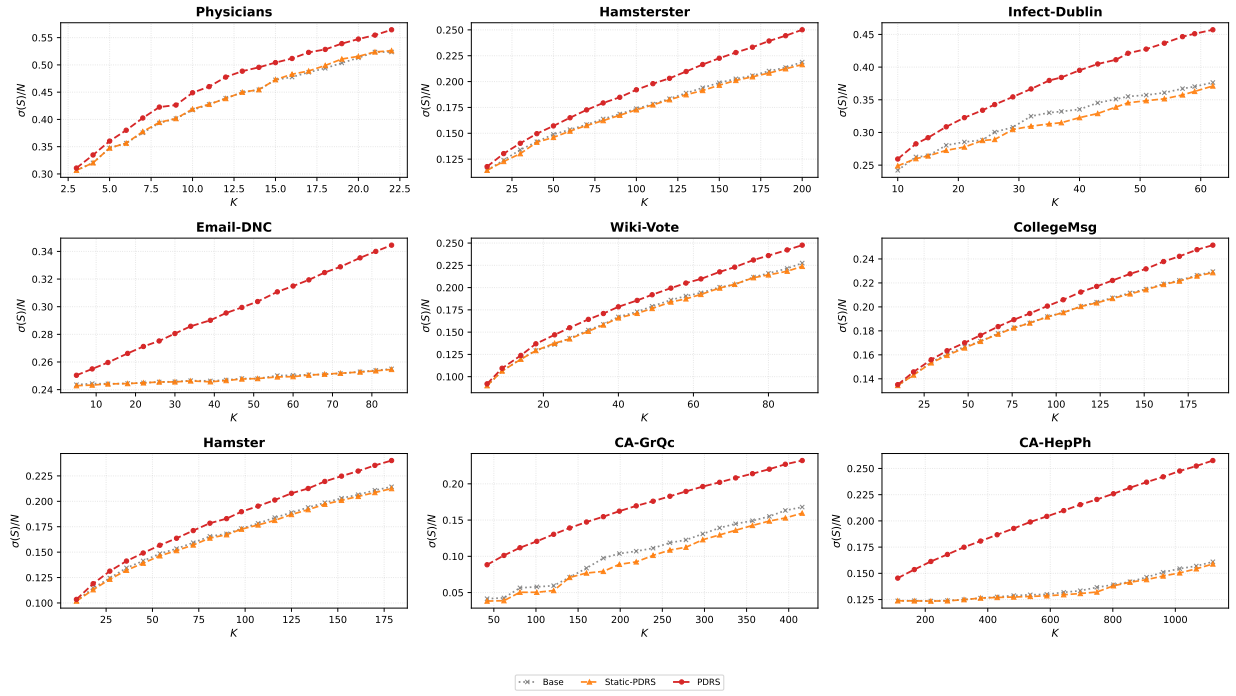


图 9 三种消融版本在 9 个网络上的传播结果

Fig. 9. Spreading results of the three ablation variants on nine networks.

表 8 三种消融版本的平均最终感染比例差值

Table 8. Differences in mean final infection proportion among the three ablation variants

网络	有效方向指数	传播重叠抑制	PDRS 相对 Base
Physicians	0.1530	2.8354	2.9885
Hamsterster	-0.1567	1.9481	1.7914
Infect-Dublin	-0.8386	6.0419	5.2033
Email-DNC	-0.0335	4.9670	4.9335
Wiki-Vote	-0.1206	1.4257	1.3051
CollegeMsg	-0.0462	1.0992	1.0530
Hamster	-0.1735	1.6236	1.4501
CA-GrQc	-0.8544	7.1025	6.2481
CA-HepPh	-0.1553	7.1125	6.9572
均值	-0.2473	3.7951	3.5478

加入有效方向指数后, 平均最终感染比例在 Physicians 上提高, 在其余 8 个网络上下降, 9 个网络的平均变化为 -0.2473 个百分点。Base 与 Static-PDRS 的节点初始得分 Spearman 秩相关系数位于 0.983264 至 0.997483 之间, 种子集合 Jaccard 系数平均为 0.867008。有效方向指数改变了部分节点的初始排序和种子选择, 其对传播结果的影响方向随网络而变。

继续加入传播重叠抑制策略后, 9 个网络的平均最终感染比例均提高, 9 个网络的平均增量为 3.7951 个百分点。完整 PDRS 相对于 Base 在 9 个网络上也均取得更高的平均最终感染比例, 平均增量为 3.5478 个百分点。PDRS 相对于 Base 的传播性能提高主要出现在传播重叠抑制策略加入以后。

图10给出三个版本在 9 个网络上的种子节点平均最短路径长度随种子节点数量的变化。Static-PDRS

相对于 Base 的变化随网络和种子节点数量改变, 全部结果的平均变化为 0.024139。PDRS 相对于 Static-PDRS 的种子节点平均最短路径长度在 9 个网络上均增加, 9 个网络的平均增量为 0.918770。

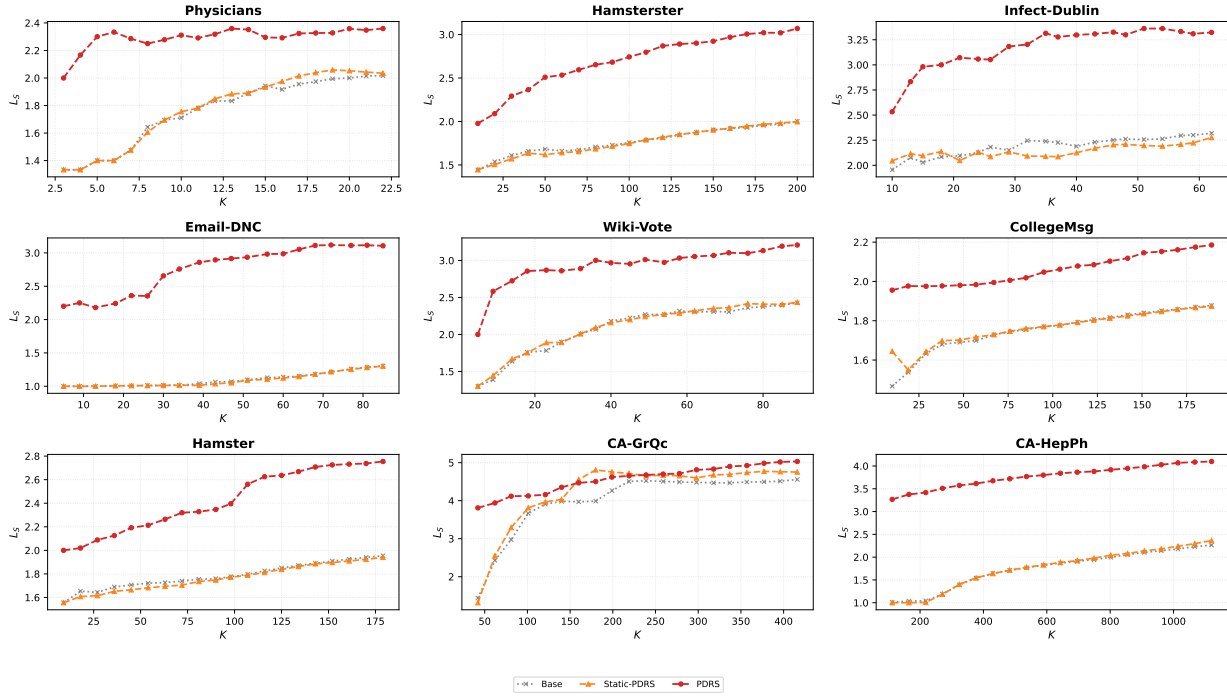


图 10 三种消融版本的种子节点平均最短路径长度

Fig. 10. Mean shortest path length among seeds for the three ablation variants.

加入传播重叠抑制策略后, 9 个网络的相邻种子对比例和闭邻域平均 Jaccard 系数均下降, 一阶邻域覆盖效率均提高。社区覆盖率在 7 个网络上提高, 在 Infect-Dublin 上保持不变, 在 CollegeMsg 上略有下降。动态更新减少了种子节点之间的直接邻接和一阶邻域覆盖重合, 并提高了种子集合的空间分散程度。PDRS 在 9 个网络上的平均最终感染比例均高于 Static-PDRS。消融结果表明, 有效方向指数对传播结果的影响随网络而变, 传播重叠抑制策略构成当前实验中传播性能提高的主要来源。

5 有向网络上的扩展与实验

PDRS 将节点影响力分解到传播方向, 并根据传播方向重合程度更新候选得分。有向网络的连边方向直接限定传播能够到达的邻居, 因此本节将 PDRS 扩展到有向无权网络。Esfandiari 和 Fakhrahmad 分别利用入边、出边及二者组合构造有向 K-shell, 为不同方向邻接信息的结构评价提供了近期参照 [61]。本文在 9 个有向网络上评价有向 PDRS, 并将结果与无向网络实验进行对照。

5.1 有向 PDRS

有向网络中的传播沿出边进行，节点 u 的出邻集合 $\Gamma_{\text{out}}(u)$ 构成其传播方向集合。有向 PDRS 以二跳可达集合规模 cap_u 表示邻居的后续传播能力，并根据局部互惠率调节折扣强度。初始得分仍写为 $F_u = \ln(\text{cap}_u + e) \Omega_u$ ，迭代选种继续使用有效方向指数和传播方向重合比例。

在传播能力方面，式 (2) 中目标 v 的连接能力 d_v 刻画传播到达 v 后的继续扩散能力。单独使用有向网络的出度只能反映一跳可达规模。例如，位于有向长链起点的节点出度为 1，其后续仍可到达链上的其它节点。为计入两跳范围内的出邻节点，本文采用二跳可达集合规模

$$\text{cap}_v = R_2(v) = \left| \left(\Gamma_{\text{out}}(v) \cup \bigcup_{w \in \Gamma_{\text{out}}(v)} \Gamma_{\text{out}}(w) \right) \setminus \{v\} \right| \quad (18)$$

作为有向传播能力，其中 $\Gamma_{\text{out}}(\cdot)$ 为出邻集合。方向权重相应改写为 $w_{u \rightarrow v} = (1 + \text{cap}_v) \exp(-r_{uv})$ ，其中 r_{uv} 由式 (1) 将邻域 Γ 与度 d 替换为出邻域 Γ_{out} 与出度 d_{out} 后算得。 R_2 只使用节点的两跳出邻信息。无向版本以节点度刻画邻居继续扩散能力，有向版本以二跳可达集合补充出边方向受限时的一跳信息，两者分别对应各自的邻接结构。

在折扣强度方面，定义节点 u 的局部互惠率

$$\rho_u = \frac{|\{v \in \Gamma_{\text{out}}(u) \mid u \in \Gamma_{\text{out}}(v)\}|}{d_{\text{out}}(u)}, \quad (19)$$

该比例表示 u 的出边中同时存在反向边 $u \rightarrow v$ 和 $v \rightarrow u$ 的比例。本文将式 (8) 的折扣因子改写为 $P_{u,s_t} = 1 - \rho_u \phi_{u,s_t}^{\alpha_u}$ 。重合比例 ϕ_{u,s_t} 与聚类调节指数 α_u 沿用式 (7) 和式 (8) 的定义，并将邻域替换为出邻域。当 u 的出边几乎没有反向边时， $\rho_u \rightarrow 0$ 且 $P_{u,s_t} \rightarrow 1$ ，候选节点得分变化较小。当大部分出边存在反向边时， $\rho_u \rightarrow 1$ ，折扣因子趋近无向形式 $P_{u,s_t} = 1 - \phi_{u,s_t}^{\alpha_u}$ ，传播方向重合对候选得分产生更强影响。

5.2 实验设置

本文选取 9 个不同规模和领域的有向网络作为实验对象，涵盖社交、通信和交通三类，数据取自公开网络数据集市 [45,46]，基本统计如表 9 所示。其中互惠率 r 为双向连接的边占总边数的比例。最大强连通分量 (largest strongly connected component, LSCC) 占节点总数的比例反映能够相互到达的核心规模。

表 9 有向实验网络的基本统计特征
Table 9. Basic statistics of the directed experimental networks

网络	类型	N	M	r	LSCC
HighSchool	社交	70	366	0.50	96%
ResidenceHall	社交	217	2672	0.62	99%
Email-EU	通信	986	24929	0.71	81%
USAirport	交通	1572	28235	0.78	89%
CollegeMsg	通信	1893	20292	0.64	68%
AddHealth	社交	2539	12969	0.39	85%
OpenFlights	交通	2905	30442	0.97	99%
CiaoDVD	社交	4562	40073	0.35	23%
Advogato	社交	5042	47067	0.33	62%

有向实验仍采用 SIR 模型，感染只沿出边发生，基准感染率按各网络出度分布设置。种子节点平均最短路径长度 L_S 和社区覆盖率 Q_S 在对应的底层无向图上计算。出度（Out-degree）、有向 PageRank（Directed PageRank）和有向度折扣（Directed Degree Discount）采用各自的有向定义。VoteRank 系列、熵类方法和 NOIRX 沿用原有无向定义，在对应的底层无向图上选种。所有方法所得种子集合均在同一个有向 SIR 模型上评价。

5.3 有向网络上的传播结果

图11给出各算法在 9 个有向网络上的 SIR 传播范围。PDRS 的传播曲线在多数网络和种子节点数量取值下位于最高曲线附近。种子节点数量较少时，PDRS 与其它算法的差异通常较大。随着种子节点数量增加，各算法的传播范围逐渐增加，部分网络中的曲线差异随之缩小。

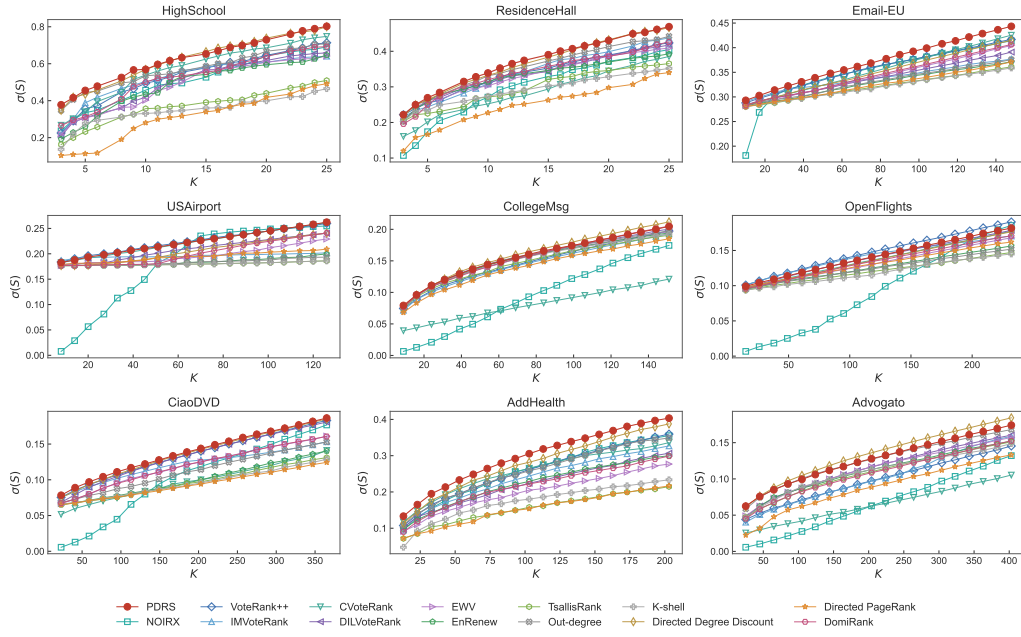


图 11 不同算法在 9 个有向网络上的 SIR 传播范围比较

Fig. 11. Comparison of SIR spreading ranges of different algorithms on nine directed networks.

表10给出各算法在由 9 个有向网络和各 20 个种子节点数量构成的 180 组实验条件下的总体统计。PDRS 的平均最终感染比例为 0.2726，平均排名为 1.72，最优次数为 93，三项指标均居首位。Directed Degree Discount 和 VoteRank++ 的平均排名分别为 3.06 和 4.87。Directed PageRank 的平均排名为 11.84。EWV、K-shell 和 TsallisRank 在底层无向图上选种，其平均排名分别为 7.77、11.34 和 11.58。上述排名反映各方法所得种子集合在当前有向 SIR 设置下的传播结果。

表 10 有向网络上不同算法传播范围的总体统计结果

Table 10. Overall spreading-range statistics of different algorithms on directed networks

算法	平均最终感染比例 $\sigma(S)$	平均排名	最优次数
PDRS	0.2726	1.72	93
Directed Degree Discount	0.2656	3.06	50
VoteRank++	0.2491	4.87	27
DILVoteRank	0.2384	5.97	0
Out-degree	0.2453	6.52	3
IMVoteRank	0.2388	6.96	4
DomiRank	0.2381	7.09	0
EWV	0.2306	7.77	0
CVoteRank	0.2301	7.83	0
NOIRX	0.2168	9.10	7
EnRenew	0.2262	9.11	0
K-shell	0.1987	11.34	0
TsallisRank	0.1995	11.58	0
Directed PageRank	0.1870	11.84	0

表11给出 PDRS 与各基线逐组配对的显著性统计。相对 Directed PageRank、EWV、K-shell 和 TsallisRank，PDRS 在全部 180 组中均显著更优。相对 Directed Degree Discount，PDRS 在 125 组中显著更

优，在 48 组中显著更差，其余 7 组差异不显著。不同基线对应的三类结果数量存在明显差异，说明 PDRS 与各方法的传播差距随算法及实验条件改变。

表 11 有向网络上 PDRS 与各基线算法传播范围差异的显著性统计

Table 11. Significance statistics of spreading-range differences between PDRS and baselines on directed networks

基线算法	PDRS 显著更优	PDRS 显著更差	差异不显著
Directed PageRank	180	0	0
EWV	180	0	0
K-shell	180	0	0
TsallisRank	180	0	0
EnRenew	179	0	1
DomiRank	172	5	3
NOIRX	171	9	0
DILVoteRank	171	2	7
CVoteRank	170	0	10
IMVoteRank	165	14	1
Out-degree	164	8	8
VoteRank++	142	34	4
Directed Degree Discount	125	48	7

图12给出各算法社区覆盖率 Q_S 的分布。社区覆盖率排序与平均最终感染比例排序存在差异。PDRS 的社区覆盖率处于各算法中部，其平均最终感染比例和平均排名均居首位。社区覆盖范围需要与种子节点自身传播能力共同评价。

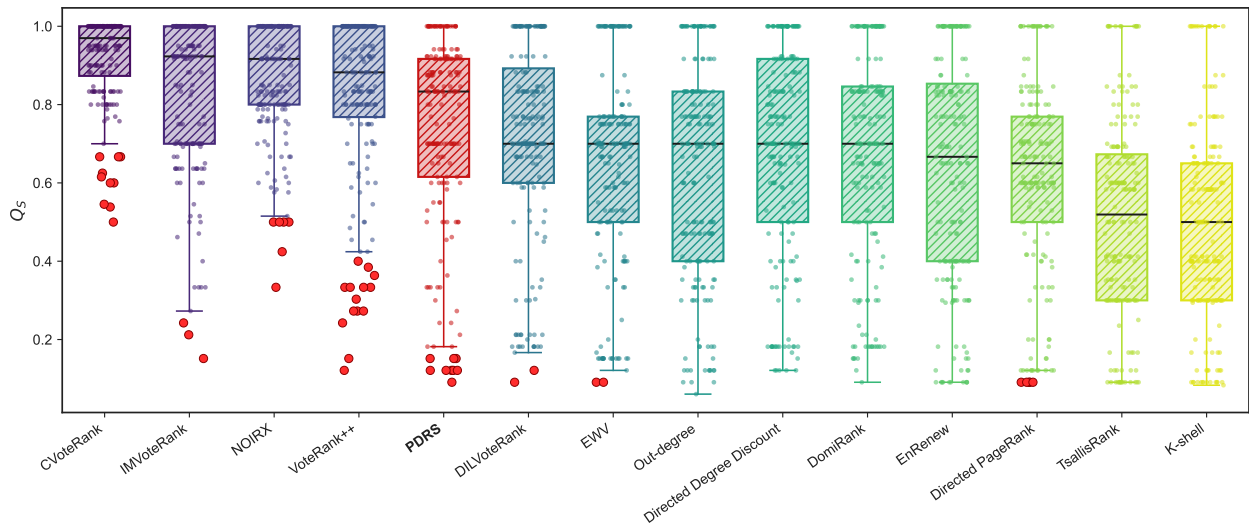


图 12 有向网络上不同算法社区覆盖率 Q_S 的分布比较

Fig. 12. Distribution of community coverage Q_S on directed networks.

5.4 有向与无向结果对照

图13汇总了各算法在无向网络和有向网络实验中的平均排名，用于比较各算法在两类网络上的整体表现。PDRS 在无向和有向实验中的平均排名分别为 1.89 和 1.72。无向网络上与 PDRS 最接近的是

VoteRank++, 平均排名为 2.80。有向网络上与 PDRS 最接近的是 Directed Degree Discount, 平均排名为 3.06。不同基线的平均排名随网络方向性发生不同程度的变化, PDRS 在两类实验中均保持较高排名。

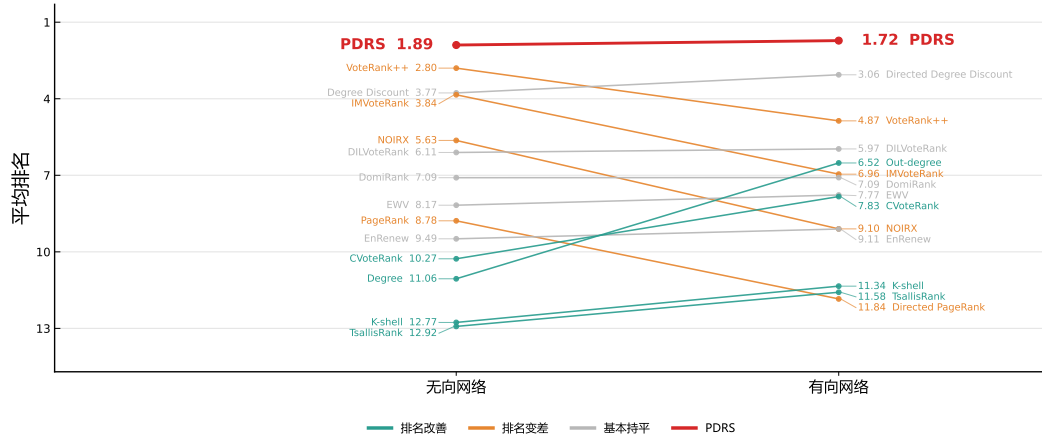


图 13 各算法在无向与有向两种设置下平均排名的迁移

Fig. 13. Mean-rank shift of each algorithm between the undirected and directed settings.

6 讨 论

6.1 式 (1) 所描述的局部传播冗余

式 (1) 利用公共邻居刻画包含边 (u, v) 的三角闭合。传播沿方向 $u \rightarrow v$ 到达 v 后, 若继续进入二者的公共邻居, 便会与 u 经由其它邻接方向形成局部覆盖重合。 r_{uv} 表示该边实际参与三角闭合的程度。

该定义限于一跳邻接信息。两个没有公共邻居的节点仍可能通过更长路径到达同一节点, 社区内部或社区边界上的传播也可能在更远位置汇合。

PDRS 需要计算相邻节点之间的边冗余程度, 并在每轮选种后更新新种子邻域内的候选得分。公共邻居可通过邻接集合求交得到, 计算和更新均限于局部邻域。若将冗余范围扩展到二阶或三阶路径, 需要进一步计算相邻节点在给定步数内的可达节点集合。社区尺度的度量还需完成社区划分并比较节点所属社区。式 (1) 的设计体现了传播冗余描述范围与计算开销之间的权衡。

6.2 网络结构对 PDRS 作用的影响

为考察网络结构与 PDRS 相对性能的关系, 本文在 20 个真实网络上, 根据同一网络和同一种子节点数量下 14 种算法的平均最终感染比例计算 PDRS 的名次, 并进一步得到每个网络的平均名次。PDRS 平

均名次与平均度、节点核数平均值和传递性的 Spearman 秩相关系数分别为 -0.445 、 -0.496 和 -0.426 。三项系数均为负,表明在本组真实网络中,平均度较大、节点核数整体较高且传递性较高时,PDRS 更容易取得靠前名次。

平均度增大为候选节点提供更多邻接方向,较高的节点核数使这些方向通向连接更充分的核心区域,较高的传递性则使公共邻居描述的局部传播冗余更频繁地出现在相邻节点之间。三种结构共同增强方向权重分布与局部传播冗余在候选节点之间的差异,使有效方向指数和传播重叠抑制策略能够更充分地区分候选节点,并最终体现为更靠前的传播名次。

平均聚类系数和模块度与 PDRS 平均名次的相关系数分别为 -0.071 和 -0.060 ,单项指标升高未对应稳定的名次改善。在 LFR 合成网络中,模块度由 0.268 提高到 0.754 时,PDRS 平均名次由 3.65 前移至 1.20 ,表明清晰的社区结构可以改善 PDRS 的相对表现。在 WS 合成网络中,平均聚类系数依次为 0.0098 、 0.4987 和 0.6487 ,PDRS 平均名次依次为 5.05 、 3.05 和 4.20 ,未呈现随平均聚类系数增加而持续前移的趋势。这些结果说明,PDRS 更容易在邻接方向较多、核心连接较充分且三角闭合较多的网络中取得传播优势,社区结构和平均聚类系数的作用还需结合上述结构共同判断。

6.3 PDRS 与 VoteRank 系列的机制区别

VoteRank 中每个节点既是候选节点,也可以为相邻候选节点提供投票贡献^[32]。候选节点的得分由相邻节点提供的投票贡献汇总得到。选出种子后,算法降低种子及其邻域节点的投票能力,后续轮次再根据更新后的投票能力汇总候选得分。VoteRank++、DILVoteRank 和 IMVoteRank 分别调整初始投票能力、投票贡献及邻域削弱范围,但仍沿用投票能力更新和候选得分重新汇总的流程。

PDRS 维护的是每个候选节点的当前传播得分。选出新种子后,算法计算该种子与相邻候选节点的传播方向重合比例,并根据候选节点的局部聚类系数确定折扣强度,随后直接更新候选节点自身的得分。VoteRank 系列中,一个节点的投票能力变化会同时改变其所有相邻候选节点获得的投票贡献,因此已选种子的影响可以经由投票能力发生变化的邻近节点继续作用于其它候选节点。PDRS 的单轮更新限于新种子的一跳候选节点,同一邻域内各候选节点的下降幅度由其与新种子的传播方向重合程度分别确定。

两类方法在首轮均使用由网络结构预先计算的得分,其机制区别体现在候选得分的形成方式、迭代过程中维护的状态和选种后的更新范围。VoteRank 系列通过投票者状态的变化重新汇总候选得分,PDRS 通过已选种子与候选节点之间的传播方向关系直接折扣当前得分。上述区别对传播性能的影响取决于给定

网络与传播过程。节点影响力的评价会随传播动力学、网络类型和模型参数改变^[54,55]，最优节点集合还会随集合规模改变^[56]。因此，方向熵与投票机制之间的性能关系需要依据完整种子集合在给定网络和传播条件下的结果判断。

7 结 论

本文针对多影响力节点选择中的局部覆盖重合问题提出 PDRS。该方法利用传播方向权重分布构造有效方向指数，并在迭代选种中根据新种子与相邻候选节点的传播方向重合程度更新候选得分。

在 9 个真实无向网络和 9 个真实有向网络上的 SIR 实验中，PDRS 取得较高的平均最终感染比例和总体排名。消融实验中，有效方向指数对传播结果的作用随网络而变化。传播重叠抑制策略在 9 个无向网络上均降低相邻种子对比比例和闭邻域平均 Jaccard 系数，提高一阶邻域覆盖效率和种子节点平均最短路径长度，并使 PDRS 的平均最终感染比例高于 Static-PDRS。三种传播模型实验表明，SIR 与 IC 在相同边传播概率下得到接近的传播范围。LT 模型采用邻居影响累积激活机制，且参数尺度与前两种模型不同，其传播范围因此与 SIR 和 IC 存在差异。

式 (1) 使用公共邻居描述三角闭合形成的局部传播冗余，多阶路径冗余和社区级别重合位于其结构范围之外。当前证据适用于本文考察的静态无权网络及相应的 SIR、IC 和 LT 设置，PDRS 相对基线的性能结论由具有完整算法比较的 SIR 实验支持。后续研究可扩展边冗余程度所描述的结构范围，并在加权网络、时序网络、多层网络、异质传播概率和真实干预场景中检验算法表现，同时研究增量计算以控制高阶结构信息带来的选种开销。

参考文献

- [1] Jeong H, Tombor B, Albert R, Oltvai Z N, Barabási A L 2000 *Nature* **407** 651
- [2] Jeong H, Mason S P, Barabási A L, Oltvai Z N 2001 *Nature* **411** 41
- [3] Barabási A L, Gulbahce N, Loscalzo J 2011 *Nat. Rev. Genet.* **12** 56
- [4] Domingos P, Richardson M 2001 *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* San Francisco, USA, August 26–29, 2001 p57
- [5] Richardson M, Domingos P 2002 *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Edmonton, Canada, July 23–26, 2002 p61

- [6] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [7] Barabási A L, Albert R 1999 *Science* **286** 509
- [8] Newman M E J 2003 *SIAM Rev.* **45** 167
- [9] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
- [10] Kempe D, Kleinberg J, Tardos É 2003 *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Washington DC, USA, August 24–27, 2003 p137
- [11] Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N 2007 *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* San Jose, USA, August 12–15, 2007 p420
- [12] Chen W, Wang Y, Yang S 2009 *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Paris, France, June 28–July 1, 2009 p199
- [13] Chen W, Wang C, Wang Y 2010 *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Washington DC, USA, July 25–28, 2010 p1029
- [14] Freeman L C 1978 *Soc. Networks* **1** 215
- [15] Katz L 1953 *Psychometrika* **18** 39
- [16] Bonacich P 1987 *Am. J. Sociol.* **92** 1170
- [17] Brin S, Page L 1998 *Comput. Netw. ISDN Syst.* **30** 107
- [18] Hirsch J E 2005 *Proc. Natl. Acad. Sci. USA* **102** 16569
- [19] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
- [20] Bae J, Kim S 2014 *Physica A* **395** 549
- [21] Zeng A, Zhang C J 2013 *Phys. Lett. A* **377** 1031
- [22] Gao S, Ma J, Chen Z, Wang G, Xing C 2014 *Physica A* **403** 130
- [23] Chen D B, Lü L, Shang M S, Zhang Y C, Zhou T 2012 *Physica A* **391** 1777
- [24] Berahmand K, Bouyer A, Samadi N 2018 *Chaos Solitons Fractals* **110** 41
- [25] Salavati C, Abdollahpouri A, Manbari Z 2019 *Neurocomputing* **336** 36

- [26] Yang P L, Zhao L J, Dong C, Xu G Q, Zhou L X 2023 *Chin. Phys. B* **32** 058901
- [27] Wen T, Deng Y 2020 *Inf. Sci.* **512** 549
- [28] Chen X, Zhou J, Liao Z, Liu S, Zhang Y 2020 *Entropy* **22** 848
- [29] Yang S Q, Jiang Y, Tong T C, Yan Y W, Gan G S 2021 *Acta Phys. Sin.* **70** 216401 (in Chinese) [杨松青, 江沛, 童廷程, 颜亦文, 甘国胜 2021 物理学报 **70** 216401]
- [30] Vafaei N, Sheikhi F, Ghasemi A 2025 *Soc. Netw. Anal. Min.* **15** 23
- [31] Guo C, Yang L, Chen X, Chen D, Gao H, Ma J 2020 *Entropy* **22** 242
- [32] Zhang J X, Chen D B, Dong Q, Zhao Z D 2016 *Sci. Rep.* **6** 27823
- [33] Sun H L, Chen D B, He J L, Ch'ng E 2019 *Physica A* **519** 303
- [34] Kumar S, Panda B S 2020 *Physica A* **553** 124215
- [35] Liu P, Li L J, Fang S Y, Yao Y K 2021 *Chaos Solitons Fractals* **152** 111309
- [36] Li H Y, Wang X, Chen Y, Cheng S Y, Lu D J 2025 *Sci. Rep.* **15** 1693
- [37] Li Y X, Yang X Z, Zhang X W, Xi M Y, Lai X C 2022 *Front. Phys.* **10** 955727
- [38] Li S J, Lei H T, Zhang M M, Zhu C, Ruan Y R 2025 *Acta Phys. Sin.* **74** 186401 (in Chinese) [李尚杰, 雷洪涛, 张萌萌, 朱承, 阮逸润 2025 物理学报 **74** 186401]
- [39] Nandi S, Malta M C, Maji G, Dutta A 2025 *J. Comput. Sci.* **86** 102540
- [40] Zhao Z L, Sun Y, Liu X P, Yan R Y, Wan L, Zhang N N 2025 *Expert Syst. Appl.* **281** 127568
- [41] Chen D B, Xiao R, Zeng A, Zhang Y C 2013 *Europhys. Lett.* **104** 68006
- [42] Wang N, Wang Z Y, Liu J G, Han J T 2019 *arXiv:1903.00248*
- [43] Engsig M, Tejedor A, Moreno Y, Foufoula-Georgiou E, Kasmi C 2024 *Nat. Commun.* **15** 56
- [44] Blondel V D, Guillaume J L, Lambiotte R, Lefebvre E 2008 *J. Stat. Mech. Theory Exp.* **2008** P10008
- [45] Kunegis J 2013 *Proceedings of the 22nd International Conference on World Wide Web Rio de Janeiro, Brazil, May 13–17, 2013* p1343
- [46] Leskovec J, Krevl A 2014 *SNAP Datasets: Stanford Large Network Dataset Collection* <http://snap.stanford.edu/data>
- [47] Moreno Y, Pastor-Satorras R, Vespignani A 2002 *Eur. Phys. J. B* **26** 521

- [48] Erkol Ş, Castellano C, Radicchi F 2019 *Sci. Rep.* **9** 15095
- [49] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379
- [50] Hill M O 1973 *Ecology* **54** 427
- [51] Newman M E J 2002 *Phys. Rev. E* **66** 016128
- [52] Liu J G, Lin J H, Guo Q, Zhou T 2016 *Sci. Rep.* **6** 21380
- [53] Liu P F, Qiu G L, Tao B S, Yang K 2025 *Proc. AAAI Conf. Artif. Intell.* **39** 487
- [54] Lü L Y, Chen D B, Ren X L, Zhang Q M, Zhang Y C, Zhou T 2016 *Phys. Rep.* **650** 1
- [55] de Arruda G F, Barbieri A L, Rodríguez P M, Rodrigues F A, Moreno Y, Costa L F 2014 *Phys. Rev. E* **90** 032812
- [56] Gu J, Lee S, Saramäki J, Holme P 2017 *Europhys. Lett.* **118** 68002
- [57] Ramrakhiyani M, Tiwari M, Sunitha V 2024 *Appl. Netw. Sci.* **9** 70
- [58] Shi W F, Fan T L, Xu S Q, Yang R M, Lü L Y 2025 *New J. Phys.* **27** 054602
- [59] Tudu A, Mandal A, Barman D 2026 *Physica A* **686** 131367
- [60] Maji G 2025 *J. Comput. Math. Data Sci.* **14** 100112
- [61] Esfandiari S, Fakhrahmad M 2026 *Inf. Process. Manag.* **63** 104719

录用稿件，非最终出版稿

Identifying multiple influential nodes in complex networks based on propagation-direction redundancy suppression*

YE Chaoran¹⁾ ZHANG Junli¹⁾ ZHANG Mengmeng¹⁾ RUAN Yirun^{1)†} BAI Liang¹⁾

1) (*College of Systems Engineering, National University of Defense Technology, Changsha 410073, China*)

Abstract

Many applications involving complex networks, including public opinion guidance and control, viral marketing, and epidemic containment, require the identification of multiple influential nodes. An effective seed set should combine strong spreading ability with limited overlap in coverage. Methods based mainly on individual-node influence may select several seeds from the same densely connected region; this concentration limits the additional coverage of later selections. To address this problem, we propose a propagation-direction redundancy suppression (PDRS) algorithm that accounts for heterogeneity among a node's propagation directions. PDRS assigns each direction a weight based on the connectivity of the neighboring node and the redundancy indicated by common neighbors. The entropy of the normalized weight distribution defines an effective direction index that quantifies the effective number of low-redundancy directions supporting propagation. After selecting a seed, PDRS reduces each adjacent candidate's score according to the share of its directional weight assigned to directions that overlap the seed's local coverage. The candidate's local clustering coefficient adjusts the reduction. Subsequent seeds are therefore selected from less-covered regions. For directed networks, two-hop reachability quantifies subsequent spreading capacity, while local reciprocity adjusts the score reduction. We compare PDRS

† Corresponding author. E-mail: ruanyirun@nudt.edu.cn

* Project supported by the National Natural Science Foundation of China (Grant No. 72401286).

with 13 methods under the susceptible-infected-recovered (SIR) model on nine real-world undirected networks and nine real-world directed networks. Under these SIR settings, PDRS achieves the highest overall mean final infection proportion among the methods compared and reduces local coverage overlap among selected seeds. Ablation results show that the effective direction index changes the initial ranking, whereas the dynamic score update contributes more to the observed improvements in spreading performance and seed dispersion. PDRS therefore balances the spreading ability of individual seeds with the suppression of redundant coverage within the seed set.

Keywords: complex networks, multiple influential nodes, propagation-direction redundancy, effective direction index

录用稿件，
非最终出版稿