基于高阶反向影响采样的无向超图影响力最大化算法

芮晓彬[1,2,3]，吉嘉欣[1]，方强鹏[1]，时纪龙[1]，王志晓[1,2,3]
（1.中国矿业大学计算机科学与技术学院/人工智能学院，徐州 江苏 221116；2.矿山数字化教育部工程研究中心，徐州 江苏 221116；
3.地下空间智能感知与应急物联江苏省产业技术工程化中心，徐州 江苏 221116）

摘要：现有无向超图影响力最大化研究中的传播模型普遍缺乏对超图高阶结构的深入分析与有效建模，难以刻画信息在群体层面的复杂传播机制。同时，现有算法多依赖节点的局部拓扑特征，难以反映信息的群体传播行为。为此，深入探究无向超图中的信息传播机制，对两类典型的信息群体传播过程进行刻画，提出无向超图独立级联传播模型。在此基础上，进一步提出高阶反向影响采样方法，并设计同时适用于两种不同激活策略的无向超图影响力最大化算法，通过理论分析证明该算法能达到与最优解$\left(1-\frac{1}{\mathrm{e}}-\varepsilon\right)$的近似比。在8个真实超图数据集上的实验结果表明，所提算法的效果显著优于基线方法，影响力扩展度最高提升了42.85%，平均运行时间仅为CELF算法的0.7%，与基于启发式的算法基本持平。

关键词：影响力最大化；无向超图；独立级联模型；反向影响采样；社交网络
中图分类号：TP301.6
文献标志码：A
doi：10.11959/j.issn.1000–436x.TXXB260210

Influence maximization algorithm for undirected hypergraphs based on higher-order reverse influence sampling

Rui Xiaobin[1,2,3], Ji Jiaxin[1], Fang Qiangpeng[1], Shi Jilong[1], Wang Zhixiao[1,2,3]
1. School of Computer Science and Technology / School of Artificial Intelligence, China University of Mining and Technology, Xuzhou 221116, China
2. Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou 221116, China
3. Jiangsu Provincial Industrial Technology Engineering Center for Intelligent Sensing and Emergency IoT in Underground Space, Xuzhou 221116, China

Abstract: Existing studies on undirected hypergraph influence maximization generally lack in-depth analysis and effective modeling of hypergraph higher-order structures, making it difficult to characterize the complex group-level information diffusion mechanisms. Meanwhile, most existing algorithms rely heavily on local topological features and fail to capture collective behaviors. To address these issues, information diffusion in undirected hypergraphs was investigated systematically, two typical group-based information diffusion processes were formalized, and an undirected hypergraph independent cascade model was proposed. Based on this model and the classic reverse influence sampling framework, a hypergraph influence maximization algorithm based on higher-order reverse influence sampling was developed, and is was theoretically proved that it achieved a $(1-\frac{1}{\mathrm{e}}-\varepsilon)$ approximation to the optimal solution. Experiments on eight real-world hypergraph datasets show that the proposed method significantly outperforms all baseline approaches, improving the influence spread by up to 42.85%. Meanwhile, its average runtime is only 0.7% of the CELF algorithm and is comparable to heuristic-based methods.

Key words: influence maximization, undirected hypergraph, independent cascade model, reverse influence sampling, social network

---

收稿日期：2026–04–16；**修回日期：**2026–06–03
通信作者：王志晓，zhxwang@cumt.edu.cn
基金项目：国家自然科学基金资助项目（No.62402496）；江苏省基础研究计划资助项目（No.BK20242084）
Foundation Items: The National Natural Science Foundation of China (No. 62402496), The Basic Research Program of Jiangsu Province (No.BK20242084)

0　引言

在当今信息化社会中，社交网络平台的普及为广告营销、疾病防控、舆情监测等领域带来了便利和支持，但同时也显著提升了谣言、虚假广告、网络暴力等有害信息的扩散风险。通过深入研究社交网络中的信息传播机制，不仅有助于促进有益信息的广泛传播，还能有效遏制有害信息的负面影响，并在下游应用中展现出广阔应用潜力[1-2]，因此，吸引众多研究者开展影响力最大化问题的研究。

影响力最大化问题于21世纪初由Richardson等[3]在病毒式营销背景下提出，其目标是在网络中寻找最优的种子节点集合，以最大化信息传播范围。Kempe等[4]提出两种基本信息传播模型，即独立级联（independent cascade，IC）模型和线性阈值（linear threshold，LT）模型，并设计了经典的贪心算法。在后续研究中，如Leskovec等[5]提出的CELF算法、Goyal等[6]对CELF算法的优化以及Borgs等[7]提出的RIS算法，虽然在效率方面有所提升，但仍基于普通图模型。随着社交网络规模不断扩大和用户交互方式日益复杂，普通图已无法准确刻画现实世界中群体化、多层次的特征，因此，超图（也被称作高阶网络，不同于普通图中无向图可以看作有向图的子集，有向超图和无向超图具有本质区别，本文研究的超图特指无向超图）影响力最大化（hypergraph influence maximization，HIM）应运而生。

信息传播模型是研究超图影响力最大化的基础，而普通图上的传统传播模型仅支持用户间直接传播行为的模拟，无法反映信息通过群组讨论、社区互动等集体行为的传播过程。为解决这一问题，Suo等[8]将传染病模型引入超图，提出CP（contact process）策略和RP（reactive process）策略，但该模型中节点可重新变为易感态，信息在短时间内被接收后又会被遗忘，不符合现实中信息的存留特点。随后，Wang等[9]剔除节点从感染态恢复为易感态的设定，提出SICP（susceptible infected with contact process）模型和SIRP（susceptible infected with reactive process）模型，但其沿用的CP策略仅允许节点选择一个所属超边进行激活；RP策略则忽略超边的特殊结构，默认激活节点所在超边均处于激活状态，不符合现实场景下用户选择性地将信息转发至一个或多个群组的行为特点。此外，Ganga l等[10]在超图中运用独立级联模型，规定超边仅能被激活一次，无法模拟群体内信息的反复传播过程，缺乏对现实场景中群体激活情况的处理。综合来看，现有超图信息传播模型存在超边激活机制不完善、状态转换机制设计不合理的问题，难以准确模拟现实场景中用户选择性地将信息转发至多个群组、群体内部反复讨论等复杂过程。

此外，影响力最大化问题已被证明为NP难问题，大规模网络下的影响力估算更是#P难问题[4]，目前尚未形成可稳定求解的大模型通用框架。现有超图影响力最大化算法大多基于节点的拓扑属性设计启发式算法，难以反映信息的群体传播行为，也无法准确评估节点的传播能力。同时，基于深度强化学习的方法需要大量计算资源，且缺乏可靠的理论保障。尽管反向影响采样（reverse influence sampling，RIS）技术在普通图中已被证明能有效平衡影响力最大化问题的计算效率与理论保证，并在普通图场景下得到广泛应用与拓展[11-12]，但传统RIS依赖于特定的传播模型且仅适用于简单拓扑结构，难以直接推广至超图中的复杂高阶网络。

为此，本文对无向超图中的群体信息传播行为进行清晰的分析与刻画，提出相应的传播模型，并结合RIS技术设计一种适用于无向超图的高阶RIS方法，从而构建高效可靠的超图影响力最大化算法，主要贡献如下。

（1）根据现实场景中用户与群体之间的交互特点，明确信息的群体传播规则，提出包含两种不同超边激活策略的无向超图独立级联模型，并剖析该模型与普通图独立级联模型之间的深层映射关系，为后续算法的理论分析奠定基础。

（2）基于上述传播模型，提出一种高阶RIS方法，构建以超边为媒介的超图反向可达集（hypergraph-reverse reachable sets，H-RR sets），并设计同时适配两种超边激活策略的超图影响力最大化算法HIM-HRIS（hypergraph influence maximization based on higher-order reverse influence sampling）。

（3）理论分析表明，本文所提HIM-HRIS算法具有$1-\frac{1}{\mathrm{e}}-\varepsilon$的最优解近似比，通过计算所需的超图反向可达集数量，可在效率与准确性之间实现良好平衡。实验结果表明，该算法在种子节点质量与算法效率方面均优于现有超图影响力最大化算法。

1 相关工作

研究者围绕超图影响力最大化问题展开算法层面的深入探索，现有方法主要包括基于节点拓扑信息的启发式方法、基于激活概率动态计算的方法以及基于深度强化学习的方法。

1.1 基于节点拓扑信息的启发式方法

Xie等[13]基于SICP模型提出一种超自适应度剪枝算法HADP，利用节点度衡量节点的影响力，迭代选择当前节点度最高的节点，并采用自适应剪枝策略降低邻居节点的超度权重，以减少影响重叠。相较于HADP算法仅考虑节点度，Gong等[14]将节点的所属超边数与邻居节点数之比定义为邻域系数，将SICP模型下超图上的影响力最大化问题转化为SICP变体模型下有向加权图上的影响力最大化问题，并提出Adeff算法解决变体问题。陈彬[15]在无向超图中应用线性阈值模型，结合节点环绕集提出潜力节点优先的预算贪婪算法，以解决给定预算下无向超图的影响力最大化问题。EFS算法[16]将超图类比为静电场，节点视为点电荷，通过电荷初始化、库仑力计算、自优化更新3个步骤构建算法，实现多维度节点影响力评估与种子选择。

此类方法通过提取超图的静态拓扑特征（如节点度、邻域系数）设计启发式规则，以高效筛选种子节点，但未深入研究信息在超图网络中传播的动态过程，且缺乏理论保障。

1.2 基于激活概率动态计算的方法

Gangal等[10]基于超图独立级联模型研究HEMI问题，旨在选出使受影响超边数量最大化的种子集合。随后，Athul等[17]沿用超图独立级联模型，将研究目标转变为最大化受影响节点数，并提出Hyper IMRANK算法，通过迭代调整初始排序生成最终排名，但该算法作为排序依据的激活概率仅基于直接共享超边的节点对的原始概率进行计算。Wang等[9]基于SICP模型和SIRP模型提出MHPD（multiple-hop probability dissemination）算法，通过构建感染概率矩阵迭代更新节点在设定跳数范围内的期望感染度，以此替代蒙特卡罗模拟来计算期望传播规模。HACE算法[18]结合传播过程对节点的有效扩散能力与种子集局部感染风险进行建模，构建折扣接触能力指标，并迭代选择该指标值最大的节点加入种子集。与HACE算法相比，HCLI算法[18]更注重候选节点的整体感染风险，引入全局感染概率，精确建模候选节点被所有邻居节点感染的累积概率，并通过迭代持续更新，选择累积概率最大的节点作为新种子。王志萍等[19]提出基于遗传算法的低冗余超图影响力最大化算法LR-HGA，将影响冗余和实际传播值的节点充分考虑后融入遗传算法的选择操作和交叉操作中，以寻找全局最优解。

相较于基于节点拓扑信息的启发式方法，此类方法利用节点的初始激活概率进行动态计算，更贴近信息扩散的真实过程，但仍局限于一阶邻居节点或有限跳数内的节点信息，且随着跳数增加，时间复杂度急剧上升。

1.3 基于深度强化学习的方法

Xu等[20]提出HEDRL-IM算法，将超图影响力最大化这一离散组合优化问题转化为DQN（deep Q-network）的连续权重优化问题，利用节点对超边的潜在激活能力解决稀疏奖励问题，并结合进化算法和深度强化学习方法探索解空间并提升解的质量。Wu等[21]提出的IMH-DRL算法需通过大量合成超图完成离线训练，使智能体学习通用的种子节点选择策略，训练完成后再将模型应用于真实超图。此后，HIMH-DRL算法[22]在异质超图上引入基于类型依赖的影响力增强机制，沿用“离线训练—在线应用”模式，根据学习到的异质超图种子选择规律选择种子节点。

上述方法虽尝试借助深度强化学习解决超图影响力最大化问题，但其可靠性缺乏理论保障。同时，该类方法需预先生成大量合成超图进行离线训练，且需要额外获取超图的嵌入信息，难以实现问题的高效求解。

综上所述，目前关于超图影响力最大化的研究缺乏对超图传播模型的深入分析，也未充分考量实际场景中不同的传播模式。此外，现有算法难以在不同传播情况下兼顾对超图影响力最大化问题的求解效率、精度和理论保障。

2 无向超图独立级联模型

2.1 无向超图

为方便后续说明，首先给出无向超图的定义，即给定无向超图$\mathrm{HG}=(V,E,W)$，其中，$V=\{v_1,v_2,v_3,\cdots,v_n\}$为无向超图所包含$n$个节点的集合；$E=\{e_1,e_2,e_3,\cdots,e_m\}$为无向超图中$m$条超边构成的

集合，且对任意超边 $e \in E$，满足 $e \subseteq V$ 且 $e \neq \varnothing$；W 为权重函数，对任意节点 $v \in V$ 和超边 $e \in E$，有 $W(v,e) = \{p_{v \to e}, p_{e \to v}\}$，其中 $p_{v \to e}$ 为节点 v 对所属超边 e 的成功激活概率，$p_{e \to v}$ 为超边 e 对其所包含节点 v 的成功激活概率。

图 1 给出一个无向超图示例，其中，$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$，$E = \{e_1, e_2, e_3, e_4\}$，$e_1 = \{v_1, v_2, v_3, v_4, v_5\}$，$e_2 = \{v_2, v_3, v_4, v_5\}$，$e_3 = \{v_4, v_5, v_6\}$，$e_4 = \{v_7\}$。在图 1 中，用户的节点 v_1, v_2, v_3, v_4, v_5 处在同一个社群 e_1，信息可在该群体内传播。同时，节点 v_2 也属于超边 e_2，因此，用户 v_2 可以将在社群 e_1 中获得的信息以某种方式传播至社群 e_2。相比之下，节点 v_7 仅属于单个社群 e_4，且 e_4 只包含该单一用户，若用户 v_7 为信息的初始拥有者，则信息在该无向超图中将无法继续传播。

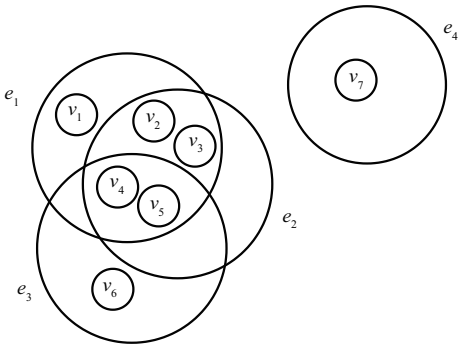


图 1　无向超图示例

由无向超图的定义与上述示例可以得出，无向超图中的一条超边可连接多个节点，表示一个讨论群组或共享信息的用户集合，每个节点均可作为信息的发送者或接收者。因此，无向超图适用于每个社交用户分属多个群体的情况，能够更准确地刻画信息在社交网络间通过群体传播的过程。

2.2　面向无向超图的独立级联模型

普通图无法刻画信息在群体中的传播过程，而现有超图信息传播模型适用的场景单一，无法准确模拟现实场景中用户选择性地将信息转发至多个群组、群体内部反复讨论等复杂过程。本文深入分析信息在群组中的实际传播情况，总结得到群体信息传播的 3 个核心特点。

（1）用户以一定概率将信息传播至其所在的社交群体。

（2）社交群体中的信息以一定概率被该群体中的其他用户接收。

（3）用户获取某条信息后，会长期维持该信息的接收状态，不会由于其他用户或群体的后续传播而改变自身已接收信息的状态。

此外，在实际传播过程中，社交群体可能一次或多次接收相同的信息。前者对应现实中转发通知信息的场景，例如，一条通知由上级发送至群组，再由群组中下级组织的各负责人转发至对应的组织，群组中的其他人不会重复发送该通知。后者刻画热点信息或重大新闻在现实中的传播场景，例如，针对某一热点新闻，同一群组内的不同用户在不同时刻获取该信息后，均可能向该群组转发。

基于以上信息传播特点，本文结合超图结构与普通图上的 IC 模型[4]，提出面向无向超图的独立级联（hypergraph independent cascade，HIC）模型，并设计相应的传播策略 OT（one touch）和 MT（multi touch），分别刻画上述两种不同的超图信息传播场景，具体传播规则如下。

给定一个无向超图 $\mathrm{HG} = (V,E,W)$，本文用 $N(e)$ 表示超边 $e \in E$ 所包含的节点集合，$N(v)$ 表示节点 $v \in V$ 所在的超边集合。节点只处于未激活与激活两种状态，且一旦被激活，状态就不再更改。每个节点在激活后仅有一次机会分别激活其所属的处于非激活状态的超边。每条超边被激活后，仅有一次机会在下一时刻分别激活其所包含的节点。在 OT 策略下，每条超边被激活后维持激活状态，即在传播过程中超边只能被激活一次。在 MT 策略下，超边向节点传播信息后被重置为未激活状态，即每条超边可多次被激活。

以下举例说明 HIC-OT 与 HIC-MT 两种模型的具体传播过程，设定节点激活所属超边的概率为 0.8，超边激活内部节点的概率为 0.5。如图 2 所示，加粗表示该超边被多次激活，给定 $\mathrm{HG} = (V,E,W)$，其中，$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$，$E = \{e_1, e_2, e_3\}$，$e_1 = \{v_1, v_2, v_3\}$，$e_2 = \{v_2, v_3, v_6, v_7\}$，$e_3 = \{v_3, v_4, v_5\}$，$S = \{v_1\}$，传播过程示例如下。

（1）$t = 1$ 时刻，种子节点 v_1 以给定概率尝试激活 e_1，假设激活成功，则在 HIC-OT 与 HIC-MT 模型中，e_1 均处于激活状态（图 2 中用斜线表示）。

（2）$t = 2$ 时刻，无论 HIC-OT 还是 HIC-MT 模型，e_1 均以给定概率尝试激活其所包含的未激活节

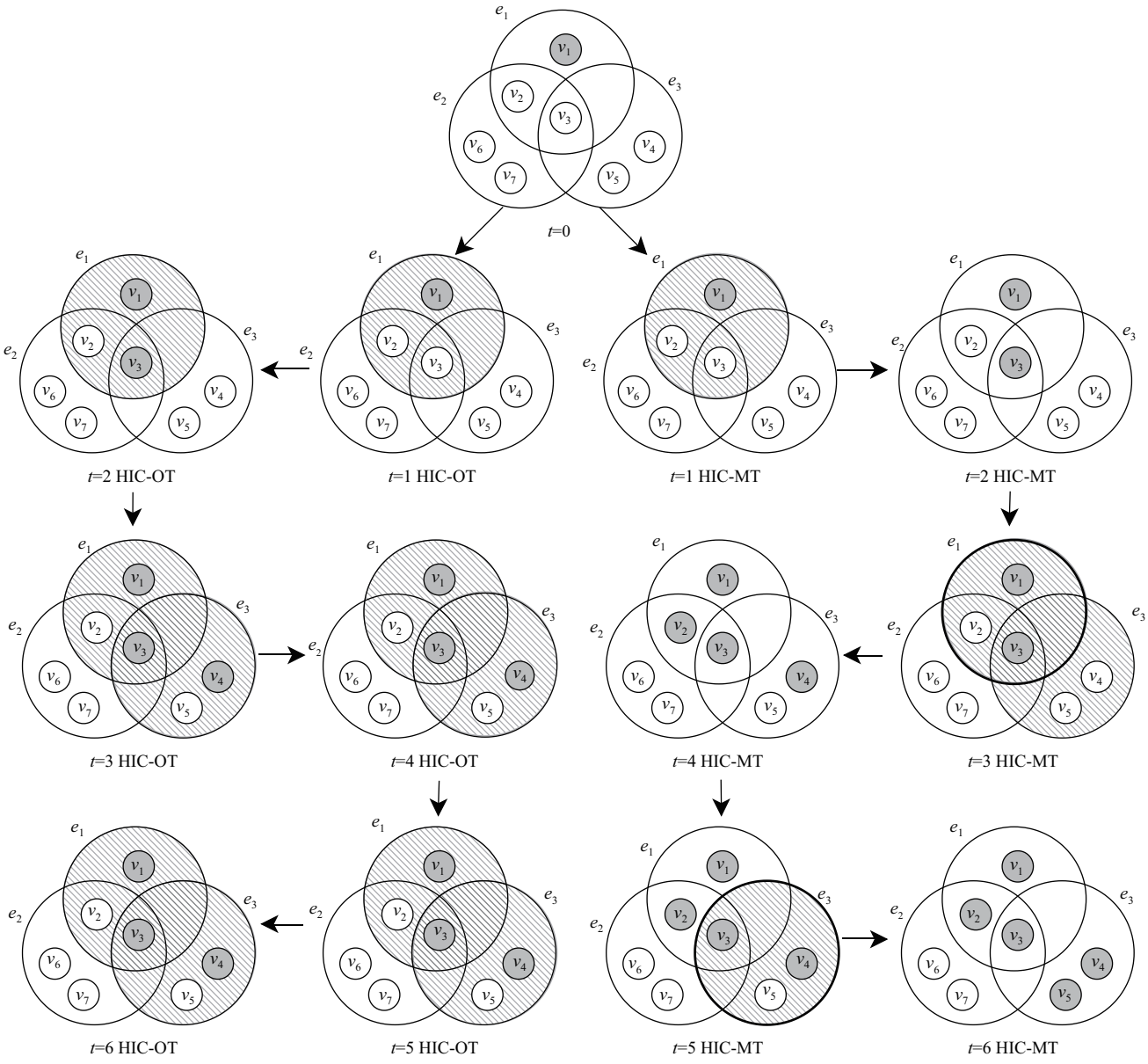


图2　无向超图独立级联模型的传播过程

点v_2和v_3，假设两种模型均仅成功激活v_3。此后，HIC-OT模型中的e_1维持激活状态，而HIC-MT模型中的e_1被重置为未激活状态。

（3）$t=3$时刻，在HIC-OT模型中，虽然v_3同时属于e_1、e_2和e_3，但由于e_1处于激活状态，v_3仅能尝试激活e_2和e_3，假设成功激活e_3。在HIC-MT模型中，v_3分别以给定概率尝试激活e_1、e_2和e_3，假设成功激活e_1和e_3。

（4）$t=4$时刻，在HIC-OT模型中，e_3尝试激活其包含的未激活节点v_4和v_5，假设仅v_4被成功激活。在HIC-MT模型中，假设e_3的激活情况与HIC-OT模型相同，同时e_1尝试激活其包含的未激活节点v_2，且激活成功。之后，HIC-OT模型中的e_3维持激活状态，而HIC-MT模型中的e_1和e_3均被重置为未激活状态。

（5）$t=5$时刻，HIC-OT模型中的传播过程结束。而在HIC-MT模型中，v_2未能激活e_1和e_2，但v_4成功激活e_3，而e_3在$t=6$时刻进一步激活v_5，至此，HIC-MT模型中的传播过程结束。

2.3　HIC模型与普通图IC模型的映射关系

通过分析HIC-OT和HIC-MT模型与普通图IC模型之间的映射关系，深入分析HIC模型的传播机制，为设计可靠的近似算法提供理论依据。

（1）在HIC-OT模型下，可将$\mathrm{HG}(V,E,W)$映射

为一个二部图$G(V',E',W)$，其中$V'=V\cup E$，即V'包含n个原始节点和m个新增的超边节点（将超图HG中的每条超边均视为二部图G中的一个新节点），因此，$|V'|=n+m$。此外，二部图G的边集E'遵循以下规则构建。对于每个节点$v\in V$及其所属超边$e\in N(v)$，对应一条包含两个权重$p_{v\to e}$和$p_{e\to v}$的无向边(v,e)，因此，总边数为各条超边所含节点数之和，即$|E'|=\sum_{e\in E}|N(e)|$。

上述映射保持HIC-OT传播过程的等价性，即原超图HG中节点激活超边的过程对应二部图G中节点激活超边节点，原超图HG中超边激活节点的过程对应二部图G中超边节点激活普通节点。在该映射中，OT策略的特性（每条超边节点只能被激活一次）与普通图IC模型中“每个节点仅能被激活一次”的规则相对应。例如，图1所示超图经映射后得到的二部图如图3所示，该二部图包含节点集$V'=\{v_1,v_2,v_3,v_4,v_5,v_6,v_7,e_1,e_2,e_3,e_4\}$以及其对应的边集$E'=\{(v_1,e_1),(v_2,e_1),(v_3,e_1),(v_4,e_1),(v_5,e_1),(v_2,e_2),(v_3,e_2),(v_4,e_2),(v_4,e_3),(v_5,e_2),(v_5,e_3),(v_6,e_3),(v_7,e_4)\}$。

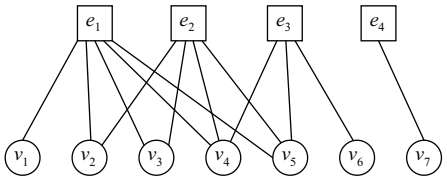


图3　无向超图示例在HIC-OT模型下的二部图映射

（2）在HIC-MT模型下，可将$\mathrm{HG}(V,E,W)$映射为普通图$G(V,E',W')$，其中，节点集$V(|V|=n)$保持不变，边集E'通过以下方式构建。对于任意两个节点u和v，若它们至少共享一条超边，即存在$e\in E$使得$u,v\in N(e)$，则在E'中建立无向边(u,v)。与HIC-OT模型的映射类似，每条边对应两个方向上的权重$W'(u,v)$和$W'(v,u)$。其中，$W'(u,v)$表示节点u通过与v共享的所有超边激活v的概率，即$W'(u,v)=1-\prod_{e\in(N(u)\cap N(v))}(1-p_{u\to e}\times p_{e\to v})$，$|E'|$的大小等于所有超边中节点组合数之和减去属于多个超边的节点组合数，即$|E'|=\sum_{e\in E}\binom{|N(e)|}{2}-\sum_{(u,v)\in V\times V}\frac{\max\{0,\varphi_{u,v}-1\}}{2}$，其中，$\varphi_{u,v}=\left|\{e\in E:u,v\in N(e)\}\right|$。

上述映射将超图中的高阶传播关系转化为普通图中的二元关系，保留MT策略允许节点通过不同超边多次尝试激活邻居节点的特性。如图4所示，节点v_1和v_2仅共同属于一条超边e_1，则有$p_{v_1\to v_2}=p_{v_1\to e_1}\times p_{e_1\to v_2}$；而节点$v_4$和$v_2$共同属于超边$e_1$和$e_2$，因此，$W'(v_4,v_2)=p_{v_4\to v_2}=1-(1-p_{v_4\to e_1}\times p_{e_1\to v_2})\times(1-p_{v_4\to e_2}\times p_{e_2\to v_2})$。

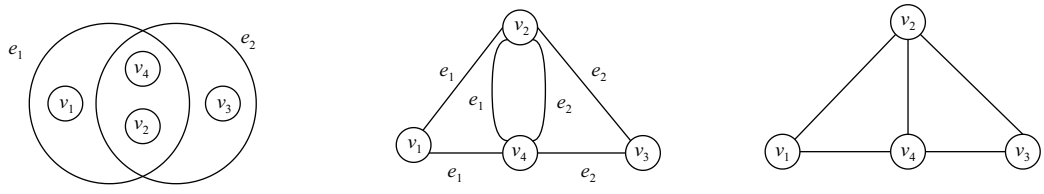
2.4　HIC模型下的无向超图影响力最大化

影响力最大化问题通常要求在给定节点数量k的条件下选择种子节点集合S，将其设置为初始激活节点，从而最大化影响力扩展度$\sigma(S)$。在普通图中，$\sigma(S)$定义为最终处于激活状态的节点数量的期望。类似地，真实世界的群组（即超图中的超边）本质上是信息扩散的媒介，因此，它们不参与最终影响力扩展度的计算，即超图中的$\sigma(S)$同样为最终处于激活状态的节点数量的期望。

基于影响力扩展度$\sigma(S)$和超图传播模型HIC，本节给出无向超图影响力最大化问题的定义，给定一个无向超图$\mathrm{HG}=(V,E,W)$和种子节点数量k，无向超图影响力最大化问题要求寻找包含k个种子节点的集合S，使得$\sigma(S)$最大，如式(1)所示：

$$S\in \mathrm{argmax}_{S_0\in V,|S_0|\le k}\sigma(S) \tag{1}$$

其中，$\sigma(S)$表示种子节点集合S在无向超图HG中基于HIC-OT或HIC-MT模型传播信息后，最终影



(a) 无向超图　　(b) “节点–超边–节点”超图　　(c) “节点–节点”图

图4　在HIC-MT模型下将无向超图转换为普通图的过程

响节点数量的期望值。此外，由式(1)可知，该问题在选择种子集合时仅允许选择节点，即由个体而非群组作为信息传播的发起者。

定理1 在HIC-OT和HIC-MT模型下，无向超图的影响力最大化问题是NP难的。

证明 首先，将最大集合覆盖（max set cover，MSC）问题规约成HIC-OT模型下的超图影响力最大化问题。已知MSC问题是经典NP难问题，给定基本集 $U=\{u_1,u_2,\cdots,u_m\}$，子集族 $F=\{F_1,F_2,\cdots,F_n\}$ 和整数 k，MSC问题要求从 F 中选择 k 个子集，使其并集覆盖 U 中元素的数量最大化。上述问题可映射为超图 $\mathrm{HG}_{\mathrm{SC}}=(V_{\mathrm{SC}},E_{\mathrm{SC}},W_{\mathrm{SC}})$，具体步骤如下。构造节点集 V_{SC}，其大小为 $n+m$，包含两类节点：将基本集 U 中的元素 u_j 转化为超图中的一个节点 v_{u_j}，将子集族 F 中的每个子集 F_i 都转化为超图中的一个节点 v_{F_i}。对于任意 $F_i\in F$ 能够覆盖元素 u_j，构造一条包含 F_i 和 u_j 所对应节点的超边 e_i，并设置超边影响节点的概率和节点影响超边的概率均为1。

在HIC-OT模型下，超边仅能被成功激活一次。种子集合 S 激活超边 e_{ij} 成功当且仅当 $v_{F_i}\in S$ 且 $u_j\in F_i$ 时，覆盖等价性成立，即 $\sigma(S)=|\bigcup_{v_{F_i}\in S}F_i|$。因此，判定该SMC问题的解等价于判定影响力最大化问题的解。由于SMC问题是一个NP难问题，因此HIC-OT模型下的超图影响力最大化问题也是一个NP难问题。在HIC-MT模型下，超边可被多次激活，但由于激活概率为1，且节点 v_{u_j} 被激活后状态不可逆，多次激活不会改变覆盖结果，因此与OT策略下的覆盖等价性完全一致。同理可证，HIC-MT模型下的超图影响力最大化问题同样为NP难问题。

定理2 在HIC-OT和HIC-MT模型下，无向超图影响力最大化问题中的影响力扩展度 $\sigma(S)$ 均为非负的单调次模函数。

证明 给定集合 $S\subseteq T\subseteq V$，记 $\Delta(v|S)=\sigma(S\cup\{v\})-\sigma(S)$。同理，$\Delta(v|T)=\sigma(T\cup\{v\})-\sigma(T)$。令 P_v 为所有从 v 出发的路径集合，$\Delta(v|S)=\sum_{P\in P_v}\Pr(P\text{ is active})*\mathbf{1}_{\{P\cap S=\varnothing\}}$，其中 $\mathbf{1}_{\{P\cap S=\varnothing\}}$ 为指示函数。对于一条路径 $P\in P_v$，将其分解为 $P=(v_0,e_0,v_1,e_1,\cdots,v_t,e_t,u)$。在OT策略下，路径中的超边 e_t 只能被其前驱节点 v_t 成功激活一次；在MT策略下，e_t 可被多次激活，因此可能重复出现。两种策略仅会导致路径所包含的元素不同，而路径 P 活跃的概率只取决于超图的结构和概率设置，与种子集 S 或 T 无关。由于 $S\subseteq T$，$\{P|P\cap S=\varnothing\}\supseteq\{P|P\cap T=\varnothing\}$，则 $1_{\{P\cap S=\varnothing\}}\geqslant 1_{\{P\cap T=\varnothing\}}$，并且 $\Pr(P\text{是活跃的})$ 独立于 S 和 T，因此 $\Delta(v|S)\geqslant\Delta(v|T)$，定理2得证。

3 算法设计

本节介绍本文提出的基于高阶RIS的无向超图影响力最大化算法（hypergraph influence maximization algorithm based on hyper RIS，HIM-HRIS），详细说明算法的设计思路和执行流程，并给出HIM-HRIS算法近似比保障的理论分析。

3.1 HIM-HRIS算法概述

整体而言，HIM-HRIS算法包括高阶RIS和种子节点选择两个核心模块。在初始化空种子集合后，将超图与传播规则传入高阶RIS模块，输出超图反向可达集合。该结果随后传入种子节点选择模块，在预算约束下完成节点择优筛选，最终输出最优种子节点集合 S，具体流程如算法1所示。

算法1 HIM-HRIS算法

输入 无向超图 $\mathrm{HG}=(V,E,W)$，策略OT或MT，预算 k

输出 种子节点集合 S

1) 初始化 $S=\varnothing$

2) HR ← 高阶RIS

3) S ← 种子节点选择

4) return S

3.2 基于HIC的高阶RIS

传统的RIS方法通常用于普通图中的IC模型和LT模型，通过预先生成大量反向可达集（reverse reachable sets，RR sets）对节点的影响力进行准确评估，从而为算法效果提供保障。然而，由于超图与普通图的结构差异以及超图中信息传播规则的变化，普通图中的RIS方法无法直接应用于无向超图中。2.3节已详细论述了本文所提HIC模型与普通图IC模型之间的映射关系，因此，可将传统的普通图上的RIS方法借助映射关系应用于无向超图。

简而言之，给定一个无向超图 $\mathrm{HG}=(V,E,W)$，

超图反向可达集的定义如下。节点v的一个随机超图反向可达集，记作$R(v)$，是在HG上的一次随机传播实例中可以激活v的所有元素的集合。

具体来说，基于HIC传播模型，给定一个无向超图$\mathrm{HG}=(V,E,W)$，从任意节点$v\in V$出发的一次高阶RIS过程，本质上是模拟HIC的反向传播过程。首先，依据给定的传播概率模拟本次能够成功激活节点v的超边；然后，模拟本次能够激活该超边的节点，持续迭代，直至不再产生新的激活节点或超边为止；最后，本次RIS过程中所涉及的节点和超边构成的集合记为$R(v)$，其中节点构成的集合记为$R_V(v)$，超边构成的集合记为$R_E(v)$，即$R(v)=R_V(v)\cup R_E(v)$，$R_V(v)\cap R_E(v)=\varnothing$。综上，基于HIC模型的一次详细RIS过程可描述如下。

（1）在V中随机选择一个节点v（通常被称为根节点）进行一次RIS，此时$R(v)=\{v\}$，$R_V(v)=\{v\}$，$R_E(v)=\varnothing$。同时，初始化最新激活节点集合$V_{\mathrm{new}}=\{v\}$，最新激活超边集合$E_{\mathrm{new}}=\varnothing$。

（2）在OT策略下，对于任意节点$\forall u\in V_{\mathrm{new}}$，以概率$p_{e\to v}$对其所属超边$e\in N(u)\backslash R_E(v)$进行采样，若采样成功，则将超边$e$加入$R_E(v)$和$E_{\mathrm{new}}$。在MT策略下，对于任意节点$\forall u\in V_{\mathrm{new}}$，以概率$p_{e\to u}$对其所属超边$e\in N(u)$进行采样，若采样成功，则将$e$加入$R(v)$、$R_E(v)$和$E_{\mathrm{new}}$。此后，将$V_{\mathrm{new}}$置为空集。

（3）对于任意超边$\forall e\in E_{\mathrm{new}}$，以概率$p_{u\to e}$对其所包含的节点$u\in N(e)\backslash R_V(v)$进行采样，若采样成功，则将节点$u$加入$R(v)$、$R_V(v)$和$V_{\mathrm{new}}$。此后，将$E_{\mathrm{new}}$置为空集。

（4）重复上述过程，直至步骤（2）或步骤（3）不再有新的元素加入$R(v)$。

上述高阶RIS过程最终得到一个以v为根节点的随机超图反向可达集H-RR sets，其物理含义为：在此次随机传播实例中，该H-RR sets中的所有节点都能激活根节点v。由于普通图IC模型与超图HIC模型之间存在映射关系，传统的RIS技术通过简单调整便可应用于超图，从而形成高阶RIS方法。

图5给出H-RR sets的示例。无向超图$\mathrm{HG}=(V,E,W)$的结构如图5(a)所示，其中，$V=\{v_1,v_2,v_3,v_4,v_5\}$，$E=\{e_1,e_2,e_3\}$，$e_1=\{v_1,v_2,v_3\}$，$e_2=\{v_3,v_4,v_5\}$，$e_3=\{v_2,v_3\}$。图5(b)~图5(d)以子图形式给出以$v_1$为根节点的3个H-RR sets示例，其中，图5(b)和图5(c)在HIC-OT模型下进行反向采样，图5(d)在HIC-MT模型下进行反向采样（加粗表示该超边被多次激活）。图5(b)中v_1反向传播至e_1，但没有进一步反向传播至e_1所包含的其他节点，该H-RR sets为$\{v_1\}$。图5(c)中v_1通过e_1反向传播至v_3，在HIC-OT模型下e_1已被加入$R_E(v_1)$，因此v_3仅能尝试反向激活e_2和e_3，成功反向激活e_3后未能进一步反向激活新的节点，该H-RR sets为$\{v_1,v_3\}$。图5(d)中v_1通过e_1反向传播至v_3，在HIC-MT模型下可以尝试反向激活e_1、e_2和e_3，假设成功反向激活e_1和e_3，那么e_1可进一步尝试反向激活v_2，激活成功后，v_2无法通过超边再反向传播至其他节点，最终该H-RR sets为$\{v_1,v_3,v_2\}$。

3.3　种子节点选择

基于以上对无向超图反向可达集的分析，可将无向超图影响力最大化问题转化为种子节点对最终生成的反向可达集的最大覆盖问题。

给定$\mathrm{HG}=(V,E,W)$，预算k和高阶反向采样生成的θ个随机H-RR sets构成的集合HR，种子节点选择过程包括以下4个步骤。

（1）遍历所有未被选中的节点$v\in V$，计算节点v所能覆盖的有效H-RR sets的数量$\mathrm{count}(v)$。

（2）从H-RR sets中选择覆盖集合数量最多即

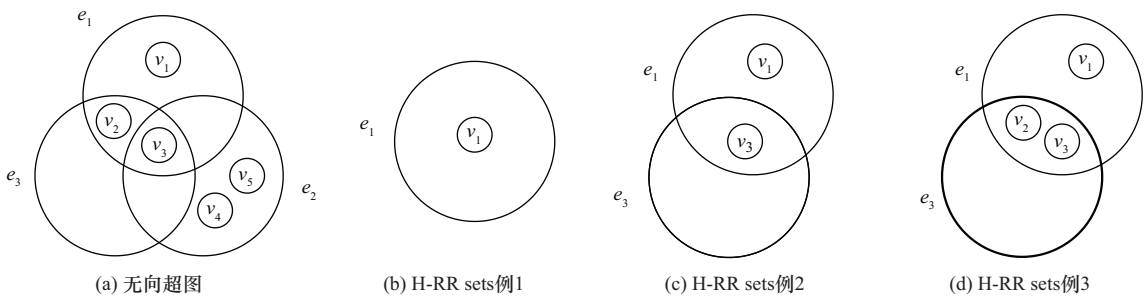


图5　超图反向可达集示例

count(v)值最大的H-RR sets的节点v_r，并将其加入种子集合S。

（3）从当前H-RR sets中移除所有包含已选择节点v_r的超图反向可达集，以消除影响力重叠。

（4）重复上述步骤，直至选出k个种子节点。

3.4 HIM-HRIS算法近似比理论分析

定理3 给定HG = $\{V,E,W\}$，在HIC-OT模型和HIC-MT模型下，对于任意种子集合$S \in V$及任意节点$v \in V$，种子集合S激活节点v的概率$\mathrm{ap}(S,v)$等于S与以v为根的一个随机超图反向可达集$R(v)$相交的概率，即$\mathrm{ap}(S,v) = \Pr\{S \cap R(v) \neq \varnothing\}$。因此，种子集合$S$的影响力扩展度满足：

$$\sigma(S) = n \times \Pr\{S \cap R \neq \varnothing\} = n \times \Pr\{S \cap R_V \neq \varnothing\} \tag{2}$$

其中，R是一个根从V中随机选取的超图反向可达集，R_V是仅包含R中节点元素的集合。

证明 在HIC-OT模型下，结合3.3节将超图HG转换为二部图$G(V',E',W)$的方法，给出HIC-OT模型基于活跃边图[4]的等价定义。给定二部图每条边上的两个权重（即$p_{v\to e}$和$p_{e\to v}$），HIC-OT模型可视为由以下元素构成的随机传播模型的特例。

（1）二部图$G = (V,E,W)$，其中，节点集$V = V^V \cup V^E$，V^V为原超图中的节点集（节点记为v^V），V^E为原超图中超边转节点集（节点记为v^E）。

（2）每个节点的状态空间$\sum = \{0,1\}$。

（3）由随机活跃边图L组成的有限传播概率空间为Ω，其中随机生成一个特定L的概率为$\Pr(L) = \prod_{(u,v)\in E(L)} p(u,v) \times \prod_{(u,v)\in E(G)\setminus E(L)} (1-p(u,v))$，$E(G)$和$E(L)$分别为图$G$和图$L$中的边。

（4）图G中传播事件的离散时间序列为$\{(t,v)|t = 1,2,3,\cdots,v \in V_V\}$。

（5）传播函数$F_v\left(X_{v,t^-},\left\{X_{u,t^-}|u \in N^-(v)\right\},L\right) = \left\|\left\{\vee_{u\in N_L^-(v)} X_{u,t^-} = 1 \vee X_{v,t^-} = 1\right\}\right\|$，其中，$N^-(v)$为$v$的入邻居节点，$X_{v,t}$为$v$在$t$时刻的激活状态，$|\{\varepsilon\}|$当事件$\varepsilon$为真时取值为1，否则为0。即当节点$v$自身或存在一个在$L$中的入邻居节点在前一时刻激活，那么$v$在当前时刻也是激活的。种子集合$S_0 = \left\{v \in V_V|X_{v,0} = 1\right\}$。$S_t$代表单次传播过程中截止到$t$时刻的活跃节点集合，$\Gamma_t(L,S)$为从节点集合$S$沿图$L$中有向边在$t$步内能到达的节点集合，$0 \leqslant t \leqslant m + n - 1$，接下来证明$S_t = \Gamma_t(L,S_0)$。

在由S_0和活跃边图L决定的一次传播实例中，设$v^V \in \Gamma_t\left(L,S_0\right)$，则存在$v_0^V \in S_0$和至少一条$L$中的路径$\left(v_0^V,v_1^E,v_2^V,\cdots,v_\tau^V\right),0 \leqslant \tau \leqslant t$，使得$v_\tau^V = v^V$，则有$X_{v_1^E,1} = F_{v_1^E}(X_{v_1^E,1},\{X_{u,1^-}|u \in N^-(v_1^E)\},L) = \|\{V_{u\in N_L^-(v_1^E)}(X_{u,0} = 1 \wedge (u,v_1^E) \in E(L)) \vee X_{v_1^E,0} = 1\}$，由于$v_0^V \in N^-(v_1^E)$，且$X_{v_0^V,0} = 1$，$(v_0^V,v_1^E) \in E(L)$，因此$X_{v_1^E} = 1$。以此类推，$v = v_\tau^V$在时刻$\tau$是活跃的，因此$v^V \in S_t$。对于任意$t$时刻最终活跃节点$v \in S_t$，必定存在一个长度不大于$t$的激活序列$(v_0^V,v_1^E,v_2^V,v_3^E\cdots,v_{2\tau}^V)$，$0 \leqslant \tau \leqslant \frac{t}{2}$，使得$v_0^V \in S_0$，$v_\tau^V = v$，$v_{2i-2}^V$激活$v_{2i-1}^E$，$v_{2i-2}^E$激活$v_{2i-1}^V$，对所有$i = 1,2,\cdots,\tau$成立，因此$(v_0^V,v_1^E,v_2^V,v_3^E,\cdots,v_{2\tau}^V)$是$L$中的一条路径，从而$v = v_\tau^V$可以从一个种子节点$v_0 \in S_0$沿$L$到达，即$v \in \Gamma_t(L,S_0)$，因此$S_t = \Gamma_t(L,S_0)$。

对于一个随机活跃边图L，种子集合S激活v^V等价于S在L中可以到达v^V，等价于存在$u \in S$使得u在L中可以到达v，即$v \in \Gamma(L,\{u\})$。根据H-RR sets的定义，$v \in \Gamma(L,\{u\})$等价于$u \in \mathrm{HR}(v)$。因此S激活v等价于存在节点$u \in S$，$u \in \mathrm{HR}(v)$，即$S \cap \mathrm{HR}(v) \neq \varnothing$。因此，当$R$的根$v$均匀选取时，$\Pr\{S \cap R \neq \varnothing\} = \sum_{v\in V} \frac{1}{n} \Pr\{S \cap R(v) \neq \varnothing\} = \frac{1}{n} \sum_{v\in V} \mathrm{ap}(S,v) = \frac{1}{n}\sigma(S)$。

相较于HIC-OT模型，HIC-MT模型下的证明过程类似，仅随机活跃边图L的概率分布计算式有所不同。综上所述，在HIC-OT模型和HIC-MT模型下，超边仅作为传播媒介，不影响种子节点影响力的计算，定理3得证。

此外，从RIS集合覆盖的角度可进一步理解超边在HIC模型下作为传播媒介的作用。对于任意$\forall e \in R_E(v)$，假设e在t时刻加入$R_E(v)$，则在$t-1$时刻必定存在$v \in R_V(v)$且$v \in e$。在最终的超图反向可达集$R(v)$中，仅保留$R_V(v)$即可完整刻画反向可达集对根节点v的影响。因此，得出式(3)为：

$$\Pr\{S \cap R \neq \varnothing\} = \Pr\left\{S \cap \left(R_E \cup R_V\right) \neq \varnothing\right\} = \Pr\{S \cap R_V \neq \varnothing\} \tag{3}$$

基于上述理解，如果一个包含k个节点的种子集合S能够覆盖最多的H-RR sets，那么该集合具备将信息传播至最广范围的能力。同时，H-RR sets数量越多，最终选出的种子节点集合的准确性越高。然而，随着H-RR sets数量的增加，生成H-RR sets的时间成本也随之上升。因此，需要在时间消耗与算法准确性之间进行平衡，以确定合适的H-RR sets数量。

定理4　定义S^*是$\sigma(S)$下的最优解，即满足$S^* \in \operatorname{argmax}_{S \in V,|S| \leqslant k}\sigma(S)$。令$\mathrm{OPT} = \sigma(S^*)$，对于任意$\delta_1,\delta_2 > 0$，$\varepsilon > 0$，$\varepsilon_1 \in \left(0, \dfrac{\varepsilon}{1-\dfrac{1}{\mathrm{e}}}\right)$，给定$\mathrm{HG} = (V,E,W)$，节点数量为$n$，种子节点集合大小为$k$，$\Omega$为所有随机元素$R_0$构成的概率空间，令$\theta_1$和$\theta_2$分别为：

$$\theta_1 = \frac{2n \cdot \ln\left(\dfrac{1}{\delta_1}\right)}{\mathrm{OPT} \cdot \varepsilon_1^2} \tag{4}$$

$$\theta_2 = \frac{\left(2-\dfrac{2}{\mathrm{e}}\right) \cdot n \cdot \ln\left(\dfrac{n}{k \times \delta_2}\right)}{\mathrm{OPT} \times \left(\varepsilon - \left(1-\dfrac{1}{\mathrm{e}}\right) \cdot \varepsilon_1\right)^2} \tag{5}$$

根据上文可得$\forall\theta > \theta_1$，$\Pr_{R_0 \sim \Omega}\{\widehat{\sigma_\theta}(S^*,R_0) \geqslant (1-\varepsilon_1) \cdot \mathrm{OPT}\} \geqslant 1-\delta_1$。$\forall\theta > \theta_2$，对于每个在$\varepsilon$约束下无法保证与最优解近似比的$S$，即$\sigma(S) < \left(1 - \dfrac{1}{\mathrm{e}} - \varepsilon\right) \times \mathrm{OPT}$，有$\Pr_{R_0 \sim \Omega}\{\widehat{\sigma_\theta}(S,R_0) \geqslant (1-\dfrac{1}{\mathrm{e}})(1-\varepsilon_1) \cdot \mathrm{OPT})\} \leqslant \dfrac{\delta_2}{\binom{n}{k}}$。

证明　本文给出Chernoff界的定义，假设$X_1X_2,\cdots,X_t$是t个$[0,1]$上相互独立的随机变量，且$\exists\mu \in [0,1]$使得$\mathrm{E}[X_i] = \mu$对任意$i \in [t]$均成立。令$Y = \sum_{i=1}^{t} X_i$，对于$\forall\gamma > 0$，则有：

$$\Pr\{y - t\mu \geqslant \gamma \cdot t\mu\} \leqslant \exp\left(-\frac{\gamma^2}{2+\dfrac{2}{3}\gamma} \times t\mu\right) \tag{6}$$

对于$\forall 0 < \gamma < 1$，则有：

$$\Pr\{y - t\mu \leqslant -\gamma \cdot t\mu\} \leqslant \exp\left(-\frac{\gamma^2}{2} \times t\mu\right) \tag{7}$$

$R_0[\theta]$为生成的θ个随机独立的超图反向可达集，对于所有节点子集S，$X_i^{R_0}(S)$表示$R_0[\theta]$的第i个集合是否与S相交，由式(2)可知，$X_i^{R_0}(S) = \dfrac{\sigma(S)}{n} \in [0,1]$，应用Chernoff界，则有式(8)：

$$\begin{aligned}
&\Pr_{R_0 \sim \Omega}\{\widehat{\sigma_\theta}(S^*,R_0) < (1-\varepsilon_1) \times \mathrm{OPT})\} = \\
&\Pr_{R_0 \sim \tau}\left\{n \times \frac{\sum_{i=1}^{\theta} X_i^{R_0}(s^*)}{\theta} < (1-\varepsilon_1) \times \sigma(S^*)\right\} = \\
&\Pr_{R_0 \sim \Omega}\left\{\sum_{i=1}^{\theta} X_i^{R_0}(s^*) - \theta \times \frac{\sigma(S^*)}{m+n} < -\varepsilon_1\left(\theta \times \frac{\sigma(s^*)}{n}\right)\right\} \leqslant \\
&\exp\left(-\frac{\varepsilon_1^2}{2}\theta \times \frac{\sigma(s^*)}{n}\right) \leqslant \\
&\left(-\frac{\varepsilon_1^2}{2} \times \frac{2n \times \ln\left(\dfrac{1}{\delta_1}\right)}{\mathrm{OPT} \times \varepsilon_1^2} \times \frac{\sigma(s^*)}{n}\right) = \delta_1
\end{aligned} \tag{8}$$

令$\varepsilon_2 = \varepsilon - \left(1-\dfrac{1}{\mathrm{e}}\right) \times \varepsilon_1$，令$S$是在$\varepsilon$约束下无法保证与最优解近似比的解，则有：

$$\begin{aligned}
&\Pr_{R_0 \sim \Omega}\left\{\widehat{\sigma_\theta}(S,R_0) \geqslant \left(1-\frac{1}{\mathrm{e}}\right)(1-\varepsilon_1) \times \mathrm{OPT}\right\} = \\
&\Pr_{R_0 \sim \Omega}\left\{\sum_{i=1}^{\theta} X_i^{R_0}(S) - \theta \times \frac{\sigma(S)}{m+n} \geqslant \frac{\theta}{n} \times \left(\left(1-\frac{1}{\mathrm{e}}\right) \times (1-\varepsilon_1) \times \mathrm{OPT}\right) - \sigma(S)\right\} \leqslant \\
&\Pr_{R_0 \sim \Omega}\left\{\sum_{i=1}^{\theta} X_i^{R_0}(S) - \theta \times \frac{\sigma(S)}{n} \geqslant \frac{\theta}{n} \times \varepsilon_2 \times \mathrm{OPT}\right\} = \\
&\Pr_{R_0 \sim \Omega}\left\{\sum_{i=1}^{\theta} X_i^{R_0}(S) - \theta \times \frac{\sigma(S)}{n} \geqslant \left(\varepsilon_2 \times \frac{\mathrm{OPT}}{\sigma(S)}\right) \times \theta \times \frac{\sigma(S)}{n}\right\} \leqslant \\
&\exp\left(-\frac{\left(\varepsilon_2 \times \dfrac{\mathrm{OPT}}{\sigma(S)}\right)^2}{2+\dfrac{2}{3}\left(\dfrac{\mathrm{OPT}}{\sigma(S)}\right)} \times \theta \times \frac{\sigma(S)}{n}\right) \leqslant
\end{aligned}$$

$$\exp\left(-\frac{\varepsilon_2^2\times \mathrm{OPT}^2}{2\sigma(S)+\frac{2}{3}(\varepsilon_2\times \mathrm{OPT})}\right)\times\theta\times\frac{1}{n}\leqslant$$

$$\exp\left(-\frac{\varepsilon_2^2\times \mathrm{OPT}^2}{2\left(1-\frac{1}{\mathrm{e}}-\varepsilon\right)\times \mathrm{OPT}+\frac{2}{3}(\varepsilon_2\times \mathrm{OPT})}\right)\times\theta\times\frac{1}{n}\leqslant$$

$$\exp\left(-\frac{\left(\varepsilon-\left(1-\frac{1}{\mathrm{e}}\right)\times\varepsilon_1\right)^2\times \mathrm{OPT}}{2-\frac{2}{\mathrm{e}}}\right)\times$$

$$\frac{\left(2-\frac{2}{\mathrm{e}}\right)\times n\times\ln\left(\frac{n}{k\times\delta_2}\right)}{\mathrm{OPT}\times\left(\varepsilon-\left(1-\frac{1}{\mathrm{e}}\right)\times\varepsilon_1\right)^2}\times\frac{1}{n}=\delta_2\times\frac{k}{n}\tag{9}$$

根据联合界可得，对于任意满足$\theta\geqslant\theta_1$和$\theta\geqslant\theta_2$的θ，算法1选出的种子节点集合S以$1-\delta_1-\delta_2$的概率，满足$\sigma(S)\geqslant\left(1-\frac{1}{\mathrm{e}}-\varepsilon\right)\mathrm{OPT}$。在此基础上，本文采用与IMM算法[23]相同的概率参数设置方式。对OPT折半猜测，设置$\delta_1=\delta_2=\frac{1}{4n^l}$，$\varepsilon_1=\varepsilon\times\frac{\alpha}{\left(1-\frac{1}{\mathrm{e}}\right)\times\alpha+\beta}$，其中，$\alpha$与$\beta$分别满足：

$$\alpha=\sqrt{l\ln n+\ln 4}$$

$$\beta=\sqrt{\left(1-\frac{1}{\mathrm{e}}\right)\times\left(\ln\binom{n}{k}+l\ln n+\ln 4\right)}\tag{10}$$

在该参数设置下，能够保证算法1选出的种子节点集合S以概率$1-\frac{1}{2n^l}$满足与最优解至少$1-\frac{1}{\mathrm{e}}-\varepsilon$的近似程度，此时$\theta_1$与$\theta_2$相等，表示为：

$$\theta_1=\theta_2=\frac{\left(2n\times\left(\left(1-\frac{1}{\mathrm{e}}\right)\times\alpha+\beta\right)\right)^2}{\mathrm{OPT}\times\varepsilon^2}\tag{11}$$

4　实验结果与分析

4.1　数据集

本文使用8个来自不同领域的真实超图数据集进行实验，这些数据集规模不同，且节点度、超边度等拓扑属性均有所区别，数据集的拓扑属性如表1所示。其中，n为节点数，m为超边数，deg为节点平均一阶度数，d^H为节点平均超边度，d^E为超边包含平均节点数，c为超图聚集系数，d为平均最短路径，ε为超图直径，ρ为从超图所派生普通网络的边密度。

4.2　对比算法

（1）MHPD-Heuristic[9]：一种基于启发式节点评估的影响力最大化算法，通过多跳传播深度（MHPD）方法计算每个节点在局部邻域内的预期传播范围。

（2）HACE[18]：综合考虑候选节点的有效扩散能力及其被现有种子集接触的概率，通过折扣接触能力这一指标迭代选择k个种子节点。

（3）HCLI[18]：对候选节点的有效扩散能力和被所有邻居节点接触的概率进行建模，将两者联合计算作为候选节点的边际收益指标。

（4）HEDRL-IM[20]：一种进化深度强化学习算法，通过DQN将超图IM问题转化为网络权重优化问题，结合EA（evolution algorithm）算法的全局探索能力与DRL（deep reinforcement learning）的局部利用优势进行种子节点选择。关键参数均根据

表1　　数据集的拓扑属性

数据集	n	m	deg	d^H	d^E	c	d	ε	ρ
Algebra	423	1 268	78.90	19.53	6.52	0.80	1.95	5	0.19
Email	143	1 542	25.17	32.50	3.01	0.59	2.07	4	0.18
Geometry	580	1 193	164.79	21.53	10.47	0.82	1.75	4	0.28
H-Com	1 290	341	195.56	9.20	34.79	0.53	1.90	4	0.15
iAF1260b	1 668	2 351	13.26	5.46	3.87	0.56	2.67	7	0.01
NDC	1 161	1 088	10.72	5.55	5.92	0.61	3.50	9	0.01
Restaurant	565	601	79.75	8.14	7.66	0.54	1.98	5	0.14
S-Bills	294	29 157	155.65	789.62	7.96	0.84	1.46	2	0.53

原文献[20]设置，学习率为0.001，折扣因子为0.8，批次大小为64，种群大小为100，交叉概率和变异概率分别为0.8和0.2。

（5）H-IMRANK[17]：将IMRANK算法应用于超图，通过迭代估算节点的边际影响并对节点排序，直至节点排名收敛，最终选取稳定排名序列中分值最高的k个节点。

（6）基于超边度的启发式算法（H-Degree）：该算法计算各节点的超边度，并选取前k个超边度较大的节点。

（7）基于度的启发式算法（Degree）：该算法先计算所有节点的度值，再选取前k个度最大的节点。

（8）H-CELF：将CELF算法[5]应用于超图，通过动态迭代计算所有节点的边际影响力，直至选出影响力较大的k个种子节点。

为验证HIM-HRIS算法的有效性，在HIC-OT和HIC-MT模型下，计算8个数据集下各算法选取的k个种子节点的影响力范围。设置k取值为1~30，蒙特卡罗模拟次数为5 000。实验采取两种概率设置方式：①权重设置，设置节点影响其所属超边的概率$p_{v\to e}$为$\frac{1}{N(v)}$，超边影响其所包含节点的概率$p_{e\to v}$为$\frac{1}{N(e)}$；②常数设置，设置$p_{v\to e}$和$p_{e\to v}$均为固定的常数。在S-Bills数据集上，由于HEDRL-IM算法的空间复杂度较高，在96 GB内存的实验环境下运行时发生内存溢出，因此未获得该数据集上的对应实验结果。

实验结果如图6~图9所示，在iAF1260b、NDC和Restaurant数据集上，由于网络结构相对稀疏且传播范围有限，HIM-HRIS与其他对比算法之间的性能差距较小。HEDRL-IM算法在S-Bills数据集上出现内存溢出，而HIM-HRIS仍能稳定运行。MHPD-Heuristic算法仅考虑限定跳数内节点的预期传播范围，而HIC-MT模型允许超边多次激活，当种子数量较小时，其对节点影响力的评估误差被进一步放大，导致种子质量较差。相比之下，基于简单拓扑结构的算法由于超边可被重复激活提升了性能，但HIM-HRIS通过精准的集合覆盖计算，在k较小时仍能选取高质量的种子节点集合。尽管HIM-HRIS在HIC-OT模型下的性能略逊于HCLI，但在HIC-MT模型下能够获得影响力效果最优的种子集合。在其他数据集中，HIM-HRIS同样表现出明显优势。

总体而言，在不同数据集、不同传播策略以及不同传播概率设置下，HIM-HRIS均能选择高质量种子节点集合，且在不同数据集中表现稳定。

在运行效率方面，本文比较了在8个数据集

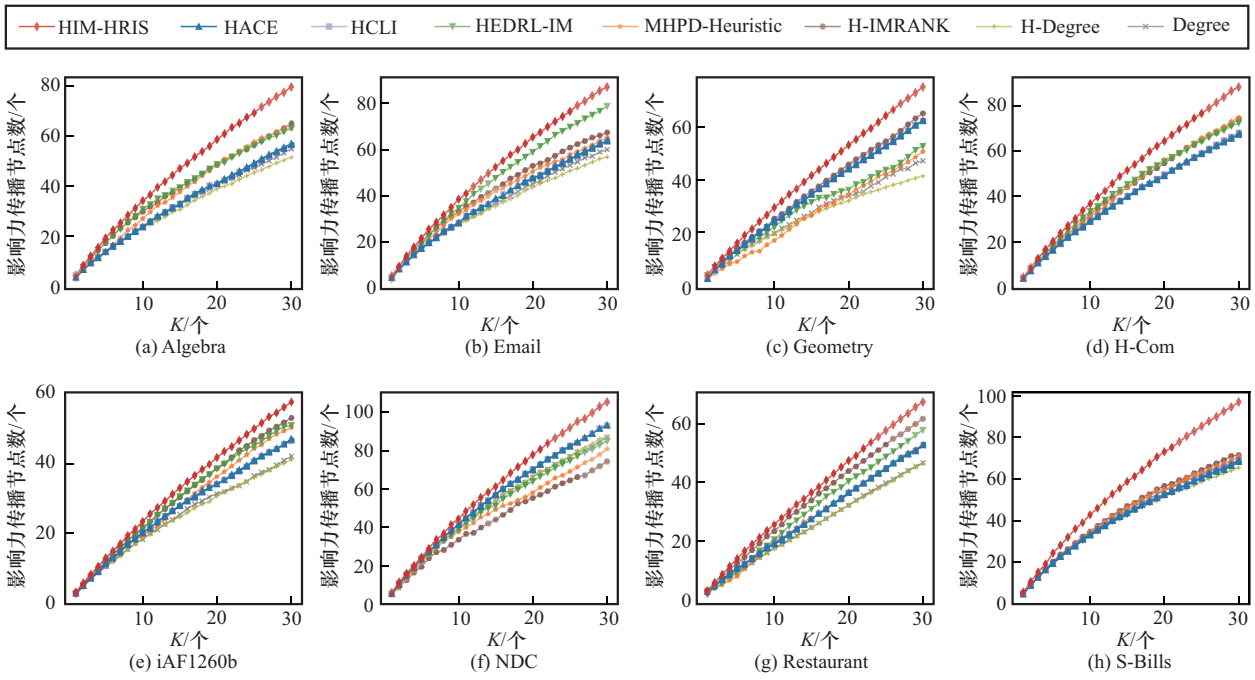


图6　HIC-OT模型下各算法的影响力范围（权重设置）

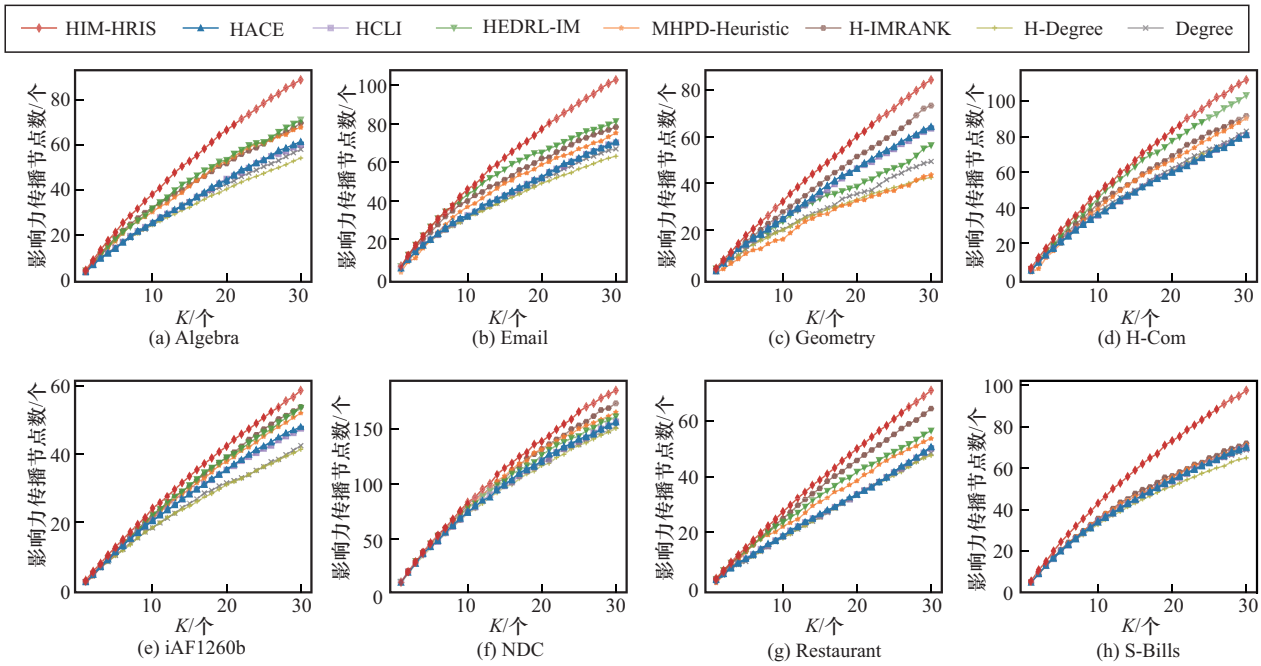


图 7　HIC-MT 模型下各算法的影响力范围（权重设置）

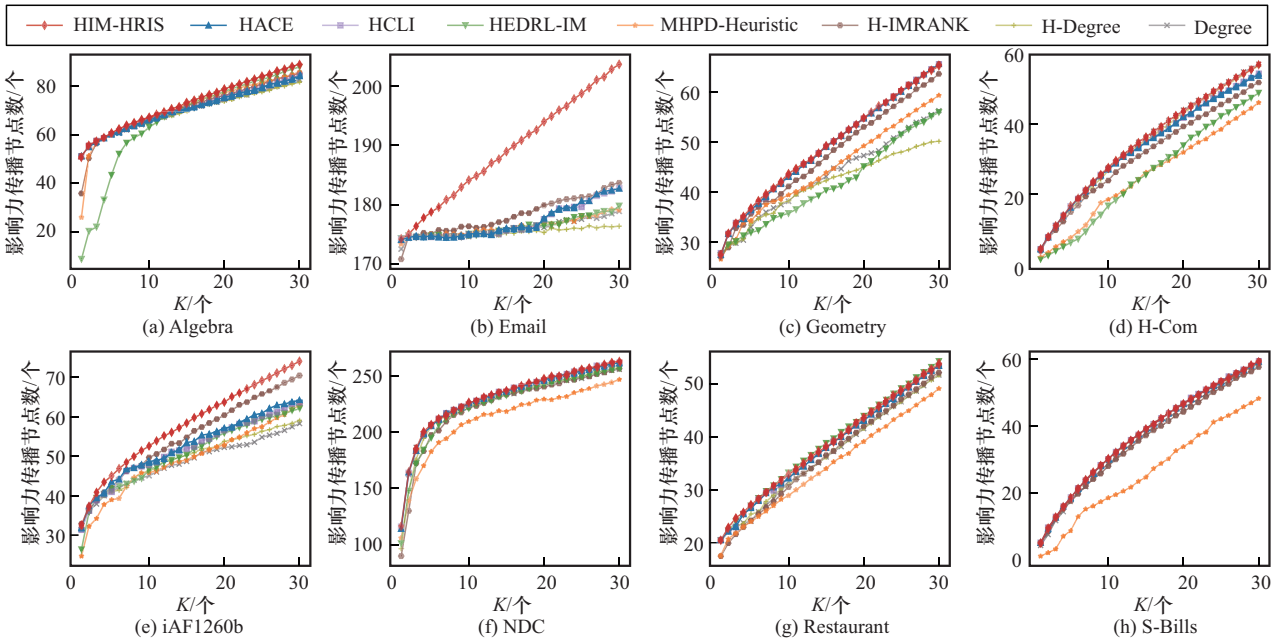


图 8　HIC-OT 模型下各算法的影响力范围（常数设置）

上（k=5，$p_{v\to e}$ 和 $p_{e\to v}$ 均设置为 0.01）HIM-HRIS 算法与对比算法分别在 HIC-OT 和 HIC-MT 模型下的运行时间。实验结果如表 2 所示，其中，MHPD-Heuristic 简写为 MHPD。HIM-HRIS 的运行效率明显高于 H-CELF 算法，在 HIC-OT 模型下 8 个数据集的平均加速比达 210.03 倍，在 HIC-MT 模型下 8 个数据集的平均加速比达 427.15 倍。HIM-HRIS 算法在所有数据集上的运行速度均显著快于 HEDRL-IM。与其他启发式算法相比，HIM-HRIS 的运行时间虽略高于 H-Degree、Degree 等基于简单拓扑的启发式方法，但在传播影响范围方面具有显著优势，实现效果与效率的良好平衡。同时，相较于 HACE、HCLI、MHPD-Heuristic 等复杂启发式方法，HIM-HRIS 在多数数据集上仍保持相当或更高的运行效率，综合表现更均衡稳定。

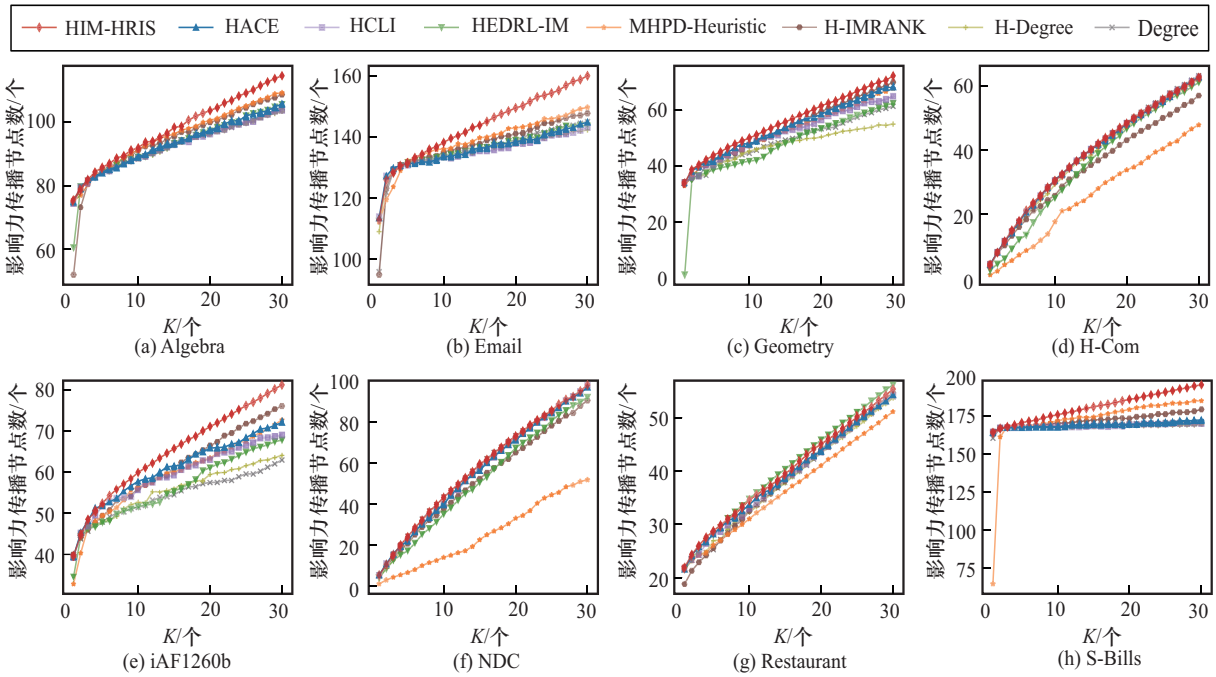


图9　HIC-MT 模型下各算法的影响力范围(常数设置)

表 2　　**HIC-OT/HIC-MT 模型下 8 个数据集上各算法的运行时间**

数据集	运行时间/s								
	HIM-HRIS	H-CELF	HEDRL-IM	HACE	HCLI	MHPD	H-IMRANK	H-Degree	Degree
Algebra	5.2/4.2	1 675.9/1 591.1	1 247.3/1 107.8	2.2/0.7	2.5/0.7	2.1/2.4	1.7/2.0	0.02/0.03	0.6/0.7
Email	1.3/1.1	161.9/276.2	773.1/519.2	0.3/0.1	0.4/0.1	0.2/0.2	0.3/0.3	0.01/0.01	0.3/0.3
Geometry	6.6/5.5	3 251.8/3 908.7	3 528.3/1 898.4	4.2/1.6	6.5/4.4	3.9/3.6	5.5/5.2	0.03/0.04	0.9/0.8
H-Com	18.3/10.2	4 919.3/18 430.5	2 695.8/2 402.2	7.5/4.3	18.8/3.9	20.3/23.4	5.1/4.4	0.06/0.07	1.0/0.9
iAF1260b	20.1/16.6	2 871.3/3 885.6	1 288.3/1 233.6	9.4/12.2	9.7/10.1	39.1/28.4	5.6/4.7	0.10/0.07	0.9/0.7
NDC	11.4/9.5	583.0/605.3	1 712.6/1 668.7	3.0/2.2	3.23/2.1	18.2/13.6	3.1/2.6	0.05/0.04	0.6/0.5
Restaurant	5.1/4.2	1 074.1/1 437.7	1 476.9/1 027.5	4.2/0.8	3.11/1.0	4.2/3.0	1.1/0.9	0.03/0.02	0.4/0.3
S-Bills	179.5/220.5	15 114.6/10 375.6	内存溢出	39.2/10.4	64.36/9.9	1.2/0.9	1 620.6/1 147.6	0.04/0.02	22.4/12.2

5　结束语

本文针对社交网络中信息的群体传播行为，系统研究无向超图中的影响力最大化问题。通过深入分析无向超图中的信息传播过程，本文提出具有两种不同超边激活策略的无向超图独立级联模型，并揭示其与普通独立级联模型之间的本质映射关系。该模型突破普通图仅能模拟节点间两两交互的局限，精准适配现实社交网络中“用户选择性转发至多群组”“热点信息在群组中反复传播”“通知类信息在群组中单次传播”等多样化场景。在此基础上，本文将 RIS 技术拓展至超图，形成高阶 RIS 方法。通过生成给定数量的超图反向可达集，将无向超图影响力最大化这一 NP 难问题转化为可高效求解的最大覆盖问题，进而设计具有理论性能保证的 HIM-HRIS 算法。理论分析结果表明，该算法能够达到与最优解（$1-\frac{1}{\mathrm{e}}-\varepsilon$）的近似比。在 8 个真实网络数据集上的实验结果表明，HIM-HRIS 在传播效果方面优于现有方法，且在计算效率方面明显优于同样具有理论保障的传统贪心算法，平均运行时间仅为传统贪心算法的 0.7%，实现了求解效率与传播效果的双重优化，验证了本文模型与算法的有

效性和实用性。

基于本文的研究成果，未来可从以下两个方向进一步深入探索：①拓展传播模型的适用场景，考虑节点异质性、信息时效性等现实因素，构建更具泛化性的超图传播模型，以适配更复杂的社交网络环境；②结合新兴技术探索HIC模型与大模型、强化学习等方法的融合路径，进一步提升算法的实用性与扩展性。

参考文献：

[1] 曾志林, 张超群, 吴国富, 等. 一种预测未知节点的融合影响力最大化的知识可迁移GNN模型[J]. 中文信息学报, 2025, 39(2): 89-99, 110.
Zeng Z L, Zhang C Q, Wu G F, et al. A knowledge transferable GNN model integrating influence maximization for predicting unknown nodes[J]. Journal of Chinese Information Processing, 2025, 39(2): 89-99, 110.

[2] 陈晋音, 张敦杰, 林翔, 等. 基于影响力最大化策略的抑制虚假消息传播的方法[J]. 计算机科学, 2020, 47(S1): 17-23, 33.
Chen J Y, Zhang D J, Lin X, et al. False message propagation suppression based on influence maximization[J]. Computer Science, 2020, 47 (S1): 17-23, 33.

[3] Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing[C]//Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2002: 61-70.

[4] Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network[C]//Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2003: 137-146.

[5] Leskovec J, Krause A, Guestrin C, et al. Cost-effective outbreak detection in networks[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2007: 420-429.

[6] Goyal A, Lu W, Lakshmanan L V S. CELF++: optimizing the greedy algorithm for influence maximization in social networks[C]//Proceedings of the 20th International Conference Companion on World Wide Web. New York: ACM Press, 2011: 47-48.

[7] Borgs C, Brautbar M, Chayes J, et al. Maximizing social influence in nearly optimal time[C]//Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms. New York: ACM Press, 2014: 946-957.

[8] Suo Q, Guo J L, Shen A Z. Information spreading dynamics in hypernetworks[J]. Physica A: Statistical Mechanics and Its Applications, 2018, 495: 475-487.

[9] Wang H S, Pan Q T, Tang J. MHPD: an efficient evaluation method for influence maximization on hypergraphs[J]. Communications in Nonlinear Science and Numerical Simulation, 2024, 139: 108268.

[10] Gangal V, Ravindran B, Narayanam R. HEMI: Hyperedge majority influence maximization[C]//Proceedings of the 2nd International Workshop on Social Influence Analysis. New York: CEUR Workshop Proceedings, 2016: 38-47.

[11] 王璿, 张瑜, 周军锋, 等. 基于社交网络的影响力最大化算法[J]. 通信学报, 2022, 43(8): 151-163.
Wang X, Zhang Y, Zhou J F, et al. Influence maximization algorithm based on social network[J]. Journal on Communications, 2022, 43(8): 151-163.

[12] 李子豪, 张锋巍. 基于可信执行环境的联邦学习平台[J]. 信息网络安全, 2026, 26(5): 788-808.
Li Z H, Zhang F W. TEE-based federated learning platform[J]. Netinfo Security, 2026, 26(5): 788-808.

[13] Xie M, Zhan X X, Liu C, et al. An efficient adaptive degree-based heuristic algorithm for influence maximization in hypergraphs[J]. Information Processing & Management, 2023, 60(2): 103161.

[14] Gong X L, Wang H C, Wang X Y, et al. Influence maximization on hypergraphs via multi-hop influence estimation[J]. Information Processing & Management, 2024, 61(3): 103683.

[15] 陈彬. 基于超图的社交网络中的影响力传播问题的算法研究[D]. 北京: 北京邮电大学, 2022.
Chen B. Research on influence propagation algorithm in social networks based on hypergraph[D]. Beijing: Beijing University of Posts and Telecommunications, 2022.

[16] Li S Y, Li X. Influence maximization in hypergraphs: a self-optimizing algorithm based on electrostatic field[J]. Chaos, Solitons & Fractals, 2023, 174: 113888.

[17] Athul M A, Arun R. Hyper-IMRANK: ranking-based influence maximization for hypergraphs[C]//Proceedings of the 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD). New York: ACM Press, 2022: 100-104.

[18] Wu L Y, Li C, Qu B, et al. Adaptive overlap penalization and probabilistic modeling in hypergraph influence maximization[J]. Information Processing & Management, 2026, 63(4): 104594.

[19] 王志萍, 赵嘉乐, 刘凯, 等. 基于遗传算法的低冗余超图影响力最大化[J]. 复杂系统与复杂性科学, 2025, 22(2): 97-104.
Wang Z P, Zhao J L, Liu K, et al. Genetic algorithm-based low redundant hypergraph influence maximization[J]. Complex Systems and Complexity Science, 2025, 22(2): 97-104.

[20] Xu L, Ma L J, Lin Q Z, et al. Influence maximization in hypergraphs based on evolutionary deep reinforcement learning[J]. Information Sciences, 2025, 698: 121764.

[21] Wu J, Li D. Modeling and maximizing information diffusion over hypergraphs based on deep reinforcement learning[J]. Physica A: Statistical Mechanics and Its Applications, 2023, 629: 129193.

[22] Sun Y H, Wu J, Song N, et al. Deep reinforcement learning-based influence maximization for heterogeneous hypergraphs[J]. Physica A: Sta-

tistical Mechanics and Its Applications, 2025, 660: 130361.

[23] Tang Y Z, Shi Y C, Xiao X K. Influence maximization in near-linear time: a martingale approach[C]//Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2015: 1539-1554.

[作者简介]



芮晓彬（1992– ），男，博士，中国矿业大学副教授，主要研究方向为社交网络分析、图数据挖掘、影响力最大化。



吉嘉欣（2003– ），女，中国矿业大学硕士生，主要研究方向为社交网络影响力最大化。



方强鹏（2000– ），男，中国矿业大学博士生，主要研究方向为社交网络影响力阻断最大化、公平性算法。



时纪龙（2000– ），男，中国矿业大学硕士生，主要研究方向为社交网络影响力阻断最大化。



王志晓（1979– ），男，博士，中国矿业大学教授，主要研究方向为社交网络分析、图数据挖掘、图表示学习。