

基于谱图理论的大规模复杂网络 重要节点组挖掘算法*

邢梓涵¹⁾ 刘丝语¹⁾ 刘慧^{1)2)3)†} 陈凌霄¹⁾

1) (华中科技大学人工智能与自动化学院, 武汉 430074)

2) (华中科技大学图像信息处理与智能控制教育部重点实验室, 武汉 430074)

3) (华中科技大学类脑智能系统湖北省重点实验室, 武汉 430074)

(2025年3月31日收到; 2025年5月28日收到修改稿)

本文研究了无向复杂网络中基于谱图理论的节点组重要性挖掘问题. 依据复杂网络牵制控制理论中节点重要性评价指标, 删后 Laplacian 矩阵最小特征值较大者为重要受控节点. 本文提出一种基于多重图特征线性融合与改进贪心搜索的重要节点组挖掘方法 (multi-metric fusion and enhanced greedy search algorithm, MFG 算法). 该方法首先通过融合度中心性、介数中心性、K-Shell 值和电阻距离等多重指标, 结合全局图特征 (如图密度、平均路径长度等) 构建线性加权融合模型, 预筛选候选节点组以克服单一指标的局限性; 其次, 设计二阶邻域局部扰动与全局随机游走搜索策略, 优化传统贪心算法的短视性, 在预筛选节点组中迭代选择使得删后 Laplacian 矩阵最小特征值最大的节点, 从而平衡局部最优与全局搜索能力; 并利用改进的反幂法进行最小特征值的计算, 降低了传统计算特征谱的复杂度, 从而使得算法总体计算性能提升. 最后, 在经典网络模型和多个真实网络中进行仿真分析, 利用不同算法挖掘重要节点组, 计算删后拉普拉斯矩阵的最小特征值, 利用 SIR 模型进行传播模拟, 并从网络拓扑上分析不同算法筛选出的重要节点组特征. 结果表明 MFG 算法相比其他几种算法挖掘重要节点组的效果更好, 对于社交网络信息传播控制具有指导意义.

关键词: 复杂网络, 节点组重要性, 预筛选算法, 删后拉普拉斯矩阵谱图理论

PACS: 64.60.aq, 07.05.Dz, 87.55.kd, 89.75.-k

DOI: 10.7498/aps.74.20250416

CSTR: 32037.14.aps.74.20250416

1 引言

近年来, 社交网络已成为信息传播的重要平台. 随着网络规模的不断扩大, 如何有效管理和控制信息传播, 尤其是识别和利用网络中的关键节点组, 变得愈加重要. 多年以来, 许多学者在复杂网络牵制控制领域已经取得非常重要的成果^[1-3]. 复杂网络中的重点节点挖掘不仅有助于理解信息流动的机制, 还能为社交网络中的舆论引导、市场推

广和危机管理提供重要参考. 目前对于社交网络重要节点组的挖掘方法已经有了较多进展. 首先考虑网络中单个节点影响力的大小, 已知许多基于节点特性和网络拓扑结构的重要性排序算法, 例如 LNC 算法^[4], 综合考虑了节点自身的影响力和最近及次近邻节点的贡献, 适合于结构紧凑型网络, 但对于稀疏网络其性能下降明显. 孔江涛等^[5]结合了网络拓扑结构信息和节点自身特性, 基于偏离均值的方差两级节点重要性评估标准, 相比 LNC 方法在处理异质性强的网络时具有一定优势, 但对参

* 国家自然科学基金 (批准号: 62176009, U24A20272) 和华中科技大学学科交叉研究项目 (批准号: 5003170102) 资助的课题.

† 通信作者. E-mail: hliu@hust.edu.cn

数选择较为敏感. 汪亭亭等^[6]提出了结合信息熵与迭代因子的 IE+算法, 通过 K-Shell 分层与改进的信息熵计算, 对节点在传播过程中的重要性进行排序, 但是较为依赖局部结构信息, 对全局拓扑的敏感性不足. 杨松青等^[7]提出了一种融合局部结构洞特征与全局 K 壳中心性的节点重要性评估方法, 通过引入 Tsallis 非广延熵度量节点及邻域的结构复杂性, 能较为准确区分网络中不同重要性节点, 但仍需合理选择熵参数以避免评估结果对网络结构变化过于敏感. 姜廷帅等^[8]提出基于信息熵赋权的多通道卷积神经网络 (EMCNN), 通过熵权法客观融合全局和局部拓扑指标, 结合 CNN 与注意力机制实现特征解耦与融合, 但对低度密集结构的 Top 节点识别精度不稳定且对大规模网络的计算效率有待优化. Rezaei 等^[9]提出了一种基于机器学习的数据驱动方法 (EML), 通过训练模型来捕捉节点的活力, 即节点在传播过程中的影响力, 相比传统基于规则的方法具有更好的自适应性, 但对数据量和标注质量依赖较高. Zhang 等^[10]提出了一种基于多属性决策 (CRITIC) 方法的复杂网络节点重要性度量指标 (MCNDI), 用于更全面地评估节点在网络中的重要性, 在多因素综合权衡上具有一定优势, 但计算复杂度相对较高. 王博雅等^[11]受物理系统中不同节点间信息的多维度、多层次相互影响过程的启发, 提出了一种基于图卷积神经网络的多维参数的节点重要性评估方法, 能够有效捕捉节点间复杂依赖关系, 但需要大量训练数据, 且模型解释性较差. 文献^[12]通过构建一个数学算子 H , 将度、 H 指数和核心性联系起来, 并提出了一种基于 H 指数的节点重要性评估方法, 在计算效率上有明显优势. Kou 等^[13]提出了一种基于图多头注意力回归模型 (graph multi-head attention regression model) 的方法, 用于识别社交网络中的影响力节点. 该方法将识别影响力节点的问题转化为回归任务, 并引入多头注意力机制和密集连接 (dense connection) 来提高模型性能, 在准确性上优于传统方法, 但计算成本高昂. 此外还有许多单个节点重要性的衡量算法. 例如 EP^[14] 算法通过竞争性渗透过程来量化节点的重要性, 能够有效模拟竞争扩散过程, 但依赖启发式策略, 易陷入局部最优.

在挖掘单个重要节点的基础上, 人们对关键节点组的挖掘问题有了进一步的研究. 相对于单个节点重要性排序, 关键节点组的挖掘更为复杂, 需要

考虑更多因素, 例如文章节点之间影响力的重叠, 网络中广泛存在的“富人俱乐部”因素^[15]等. 影响力最大化问题的求解被证明是 NP 难的, 因此目前所有算法都尽可能地综合考虑效果和复杂度. Fan 等^[16]引入了一个深度强化学习框架 FINDER, 开辟了利用深度学习技术理解复杂网络组织原理的新方向, 该方法能够挖掘到隐藏的关键节点组. Zhao 等^[17]基于图着色算法将网络中的节点划分为独立集, 再从独立集中选出个重要性靠前的节点, 保证了所选节点之间不相邻, 该方法在减少节点间冗余方面效果良好, 但对节点影响力的评估仅依赖于单一中心性指标, 且独立集划分可能导致部分高影响力节点被忽略, 尤其在社区结构明显的网络中表现受限. Bao 等^[18]提出的启发式聚类算法 (HC) 基于节点相似性指标将网络划分为多个簇, 并选择每个簇的中心节点作为传播源. 该方法通过平衡节点分散性和自身重要性, 在多种真实网络中展现出优于传统方法的传播效果. 其优势在于能够自适应地识别网络中的局部密集结构, 但算法性能受相似性指标选择的影响较大, 且对参数 λ 的设定较为敏感. Ji 等^[19]采用链路恢复和随机断边的思路, 寻找网络中紧密连接结构, 再从中选出个重要节点, 具有较强的局部结构探索能力, 但在大规模稀疏网络中性能下降. 以上工作大多仅仅基于网络局部信息, 虽然有更低的算法时间复杂度以及可扩展性, 然而算法效果不一定能得到有效保证.

另外, 复杂网络中传播现象的研究是一个热点. 历史上有许多相关的研究, 并建立了几类数学模型. Anderson 和 May^[20]设计了一个简单的确定性模型, 这就是现代数学流行病研究领域著名的 SIR 传染模型, 其成功预测了许多有记录的流行病的暴发行为. Pastor-Satorras 等^[21]对传播动力学基础理论方法和建模进行总结, 为预测病毒疾病的传播和控制提供了理论支持, 该总结覆盖面广但细节处理较为宏观. Feng 等^[22]为在线社交网络和线下传染病传播网络提出了一个具有不同时间尺度的耦合模型, 研究信息扩散与传染病传播相互作用动态中的社区结构所起的作用, 能够较好反映跨网络传播特性, 但模型参数较多, 实际应用中难以准确设定. Avraam 等^[23]提出了一个基于个体的随机 SEIR 模型, 用于模拟传染病在不同接触网络上的传播过程, 相比经典模型更能刻画实际传播过程中的随机性, 但计算复杂度随网络规模增加显

著上升.

Liu 等^[1]提出了牵制控制理论中节点重要性评价指标,即删后拉普拉斯矩阵最小特征值.我们据此结合网络图特征提出了基于多重指标(度指标、介数中心性、K-Shell 指标、电阻距离)预筛选的重要节点组贪心挖掘算法,通过二阶邻域局部扰动和全局随机游走搜索来优化贪心搜索方法,防止简单的贪心搜索会导致产生局部最优解,并用反幂法进行最小特征值计算过程的加速.我们在复杂网络节点传播的 SIR 模型上进行仿真分析验证了该算法能比较快速地找到最优的受控节点集合.经过多个复杂网络的仿真实验,得出了该算法不局限于特定属性的复杂网络,对任何类型的网络都适用的结论.

2 模型介绍、问题描述

2.1 复杂网络动力学模型

本部分首先建立复杂网络动力学分析模型,并引入基于牵制控制的同步判据.对复杂网络进行牵制控制的目的是通过控制网络中的部分节点,使得网络中所有节点的状态都收敛于目标状态 $s(t)$ ^[24].由控制理论文献^[1]可知,对于无向网络拓扑,当选定受控节点集时,系统可达同步的判定条件可表示为代数判据:

$$\lambda_1(\mathbf{L}_{N-k}) > \frac{\alpha}{c}, \quad (1)$$

其中, k 表示受控节点个数; \mathbf{L}_{N-k} 是网络拉普拉斯矩阵 \mathbf{L}_N 删去受控节点对应行和列所得到的矩阵,即 Grounded-Laplacian 矩阵^[25], $N-k$ 表示删后子矩阵的维数; $\lambda_1(\mathbf{L}_{N-k})$ 表示该删后子矩阵的最小

特征值.此外,式中 α 是由节点自身动力学函数和内连耦合矩阵决定的,独立于网络拓扑结构和耦合强度 c .

由此引出本文研究的核心命题,即在给定控制节点数量约束 l 的条件下,如何构建最优控制节点集合使得 $\lambda_1(\mathbf{L}_{N-k})$ 最大化.然而,该组合优化问题面临严峻的计算复杂性挑战,尤其是在涉及多个被控制节点的情况下.

为了解决这一问题,本文提出了一种贪心算法,该算法通过剪枝策略,能够剔除低价值的被控制节点组合,显著降低计算复杂度,从而保证在较少的时间内精确求解最优控制节点集合.

2.2 电阻距离与网络谱图最小特征值的相关性分析

考虑给定网络图 $G = (V, E)$ 及其 Laplacian 矩阵 \mathbf{L}_N ,若 \mathbf{H} 是 \mathbf{L}_N 的一个广义逆,则节点之间的电阻距离定义^[26]如下:

$$r(i, j) = \mathbf{e}_{ij}^T \mathbf{H} \mathbf{e}_{ij} = h_{ii} + h_{jj} - h_{ij} - h_{ji}, \quad (2)$$

其中, \mathbf{e}_{ij} 是 $N \times 1$ 阶的列向量,其中第 i 行元素为 1,第 j 行元素为 -1,其他行元素为 0. $i = j$ 时, $r(i, j)$ 为 0.不同于节点之间的物理距离,电阻距离可以理解为节点之间的等效电阻.

网络中某节点的电阻距离 σ_t ,等于该节点到其他节点电阻距离之和:

$$\sigma_t = \sum_{j=1, j \neq t}^N r(t, j). \quad (3)$$

由图 1 可见,在节点距离矩阵中,节点 1 和节点 3 的距离为 2,而在电阻距离矩阵中,两者电阻距离为 0.6667,这是因为电阻距离能够考虑到存在

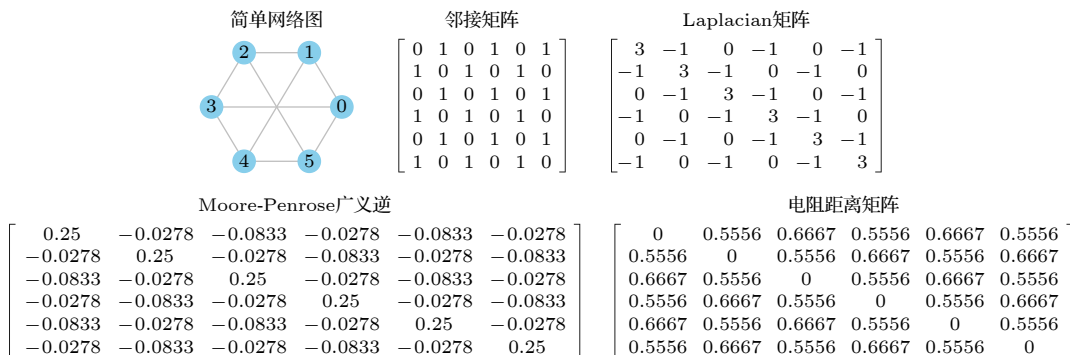


图 1 一个简单矩阵的电阻距离计算过程

Fig. 1. Process of calculating the resistance distance of a simple matrix.

“并联”(节点间存在多条连边)情况的影响,因而电阻距离更能真实反映节点之间的相对距离.因此,在后续的多特征融合指标中,我们把节点的电阻距离作为一个重要因素考虑.

结合以上定义及引理可以推导出电阻距离与 $\lambda_1(\mathbf{L}_{N-k})$ 的相关数值关系不等式,该不等式能建立节点电阻距离与 $\lambda_1(\mathbf{L}_{N-k})$ 的关系,通过对该不等式的恰当转换,能有助于剔除低价值节点,减少牵制单节点选点问题的计算量,下面给出节点电阻距离与 $\lambda_1(\mathbf{L}_{N-k})$ 不等式的推导.

考虑上述网络和 Laplacian 矩阵 \mathbf{L}_N , 若节点 t 是唯一受控节点,则有以下定理成立.

定理 1:

$$\lambda_1(\mathbf{L}_{N-1}) \leq \frac{N-1}{N} \frac{1}{\left[\sigma_t - \sum_{i=2}^N \frac{1}{\lambda_i(\mathbf{L}_N)} \right]}, \quad (4)$$

其中,特征值序列 $\lambda_N(\mathbf{L}_N) \geq \dots \geq \lambda_1(\mathbf{L}_N)$, 即 \mathbf{L}_N 所有特征值从大到小排列.

证明: 设 \mathbf{L}_N 是图 G 的 Laplacian 矩阵, \mathbf{L}_{N-1} 是 \mathbf{L}_N 删去第 t 行和第 t 列后的主子矩阵,可以通过以下方法得到 \mathbf{L}_N 的一个一般广义逆 $\mathbf{H} \in \mathbf{R}^{N \times N}$. 首先令 $\mathbf{B} = \mathbf{L}_{N-1}^{-1}$, 则 $\mathbf{B} \in \mathbf{R}^{(N-1) \times (N-1)}$, 其中 \mathbf{B} 没有负数元素,然后将 \mathbf{B} 分块为

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_3 & \mathbf{B}_4 \end{pmatrix},$$

其中 $\mathbf{B}_1 \in \mathbf{R}^{(t-1) \times (t-1)}$, $\mathbf{B}_2 \in \mathbf{R}^{(t-1) \times (N-t)}$, $\mathbf{B}_3 \in \mathbf{R}^{(N-t) \times (t-1)}$, $\mathbf{B}_4 \in \mathbf{R}^{(N-t) \times (N-t)}$, 然后在矩阵 \mathbf{B} 的第 t 行和第 t 列插入 0 元素,令得到的 $N \times N$ 矩阵为 \mathbf{H} , 即

$$\mathbf{H} = \begin{bmatrix} \mathbf{B}_1 & 0 & \mathbf{B}_2 \\ 0 & \dots & 0 \\ \mathbf{B}_3 & 0 & \mathbf{B}_4 \end{bmatrix}. \quad (5)$$

由于 \mathbf{B} 是对称的,所以此时所求 \mathbf{H} 也是几何对称的,因此有:

$$r(i, j) = e_{ij}^T \mathbf{H} e_{ij} = h_{ii} + h_{ij} - 2h_{ij}. \quad (6)$$

又令 s_i 为 \mathbf{H} 中第 i 行元素的总和,即 $s_i = \sum_{j=1}^N h_{ij}$, 可得

$$2s_i = N \cdot r(i, t) + \sigma_t - \sigma_i. \quad (7)$$

又由于 \mathbf{H} 的第 t 行元素全为零,从方程 (7) 可得

$$2 \sum_{i=1}^N s_i = 2N\sigma_t - \sum_{i=1}^N \sigma_i. \quad (8)$$

又根据文献 [27,28] 知:

$$\sum_{i=1}^N \sigma_i = 2N \sum_{i=2}^N \frac{1}{\lambda_i(\mathbf{L}_N)}. \quad (9)$$

将 (9) 式代入 (8) 式,可得

$$\sum_{i=1}^N s_i = N \left[\sigma_t - \sum_{i=2}^N \frac{1}{\lambda_i(\mathbf{L}_N)} \right]. \quad (10)$$

通过以上分析,可知 \mathbf{H} 由 \mathbf{B} 添加零元素所得, $\mathbf{B} = \mathbf{L}_{N-1}^{-1}$, 显然有 $\sum_{i=1}^N s_i = \sum_{i,j}^{N-1} b_{ij}$ 即得

$$\sum_{i,j}^{N-1} b_{ij} = N \left[\sigma_t - \sum_{i=2}^N \frac{1}{\lambda_i(\mathbf{L}_N)} \right] > 0. \quad (11)$$

且由于 \mathbf{B} 没有负数元素,对于向量 $\mathbf{x}_0 = \frac{1}{\sqrt{N-1}}(1, 1, \dots, 1)^T$, 有

$$\begin{aligned} \lambda_{N-1}(\mathbf{B}^{-1}) &\geq \mathbf{x}_0^T \mathbf{A} \mathbf{x}_0 = \frac{1}{N-1} \sum_{i,j}^{N-1} b_{ij} \\ &= \frac{N}{N-1} \left(\sigma_t - \sum_{i=2}^N \frac{1}{\lambda_i(\mathbf{L}_N)} \right) > 0. \end{aligned} \quad (12)$$

又因为 $\lambda_{N-1}(\mathbf{B}^{-1}) = \frac{1}{\lambda_1(\mathbf{B})}$, 于是有

$$\lambda_1(\mathbf{L}_{N-1}) = \lambda_1(\mathbf{B}) \leq \frac{N-1}{N} \frac{1}{\left[\sigma_t - \sum_{i=2}^N \frac{1}{\lambda_i(\mathbf{L}_N)} \right]}. \quad (13)$$

研究发现,节点的电阻距离决定了 $\lambda_1(\mathbf{L}_{N-1})$ 的上界.因此,较小的电阻距离通常意味着 $\lambda_1(\mathbf{L}_{N-1})$ 可能更大,因此该节点可能更重要.但需注意,这仅是必要条件而非充分条件.

讨论 $\lambda_1(\mathbf{L}_{N-1})$ 和电阻距离的相关性,在不同规模和连接概率的随机网络 (Erdős-Rényi 网络) 中验证了电阻距离和 $\lambda_1(\mathbf{L}_{N-1})$ 具有强相关性.实验中计算了每个节点的度、电阻距离以及 $\lambda_1(\mathbf{L}_{N-1})$, 并对每个指标进行排名,节点度越大排名越高,总电阻距离越小排名越高, $\lambda_1(\mathbf{L}_{N-1})$ 越大排名越高.以 $\lambda_1(\mathbf{L}_{N-1})$ 排名为横坐标,度排名和电阻距离排名为纵坐标.比如一个节点在度排名中排列为 20,而在 $\lambda_1(\mathbf{L}_{N-1})$ 中排列为 15,则它在图中的坐标为 (15, 20).依此方法,绘制了所有节点构成的散点图.容易发现,对于不同的纵坐标对应的指标,散点连成的曲线越靠近 $F(x) = x$,说明该指标

与 $\lambda_1(\mathbf{L}_{N-1})$ 相关性越强. 采用 ($N = 500, p = 0.05$) 和 ($N = 2000, p = 0.1$) 的 ER 随机网络进行实验, 实验结果如图 2 所示. 从图 2 可以发现, 电阻距离对应的散点图与 $F(x) = x$ 几乎完全重合, 证明电阻距离和 $\lambda_1(\mathbf{L}_{N-1})$ 具有非常强的相关性, 从而电阻距离可以在某种程度上成为 $\lambda_1(\mathbf{L}_{N-1})$ 的下位替代指标, 对于重要节点组的挖掘具有启发意义.

根据以上讨论, 可得出电阻距离决定 $\lambda_1(\mathbf{L}_{N-1})$ 的上界, 并且电阻距离与 $\lambda_1(\mathbf{L}_{N-1})$ 有强相关性的结论. 因此在节点组的预筛选时, 将电阻距离作为重要指标加入计算, 能够有效筛除低价值节点; 在预选节点组中再筛选节点时, 利用电阻距离能够以较低的复杂度达到较高的筛选精度. 但是值得注意的是, 电阻距离虽然与 $\lambda_1(\mathbf{L}_{N-1})$ 关联度较高, 但是其仍然无法全面地筛选出总体上最有价值的节点, 在筛选过程中仍需结合其他指标.

2.3 多特征融合指标

度中心性 (degree centrality)、介数中心性 (betweenness centrality)、K-Shell 和电阻距离 (resistance-distance) 是复杂网络节点影响力评估

的重要指标. 但单一指标对节点重要性的描述都有所不足. 因此, 如何整合各种传统度量方法的优点以克服其各自的不足, 已成为学术界研究的焦点. 为了解决这一问题, 引入全局图特征, 将 4 种指标线性融合为一个综合指标, 从而弥补单一指标的不足.

首先, 计算图的全局图属性, 包括图密度 D 、平均度 \bar{d} 、平均网页排名 P 、平均加权度 \bar{d}_w 、平均路径长度 L_{ap} 和平均聚类系数 \bar{C} . 其中, 各图属性均使用 Z-score 标准化.

然后利用图谱信息, 分别计算度中心性、介数中心性、K-Shell 值和电阻距离的特征评估分数, 计算方法如下所示:

$$\begin{aligned} f_{\text{deg-ree}} &= a_1 \cdot \bar{d} + a_2 \cdot \bar{d}_w, \\ f_{\text{K-Shell}} &= b_1 \cdot D + b_2 \cdot P + b_3 \cdot \bar{d}, \\ f_{\text{bc}} &= c_1 \cdot L_{ap} + c_2 \cdot \bar{C}, \\ f_{\text{Res-dis}} &= d_1 \cdot D + d_2 \cdot L_{ap} + d_3 \cdot \bar{C}. \end{aligned} \quad (14)$$

进行评估分数归一化后, 分别计算各个特征的归一化权重:

$$w_i = e^{f_i} / \sum_{i=1}^4 e^{f_i}. \quad (15)$$

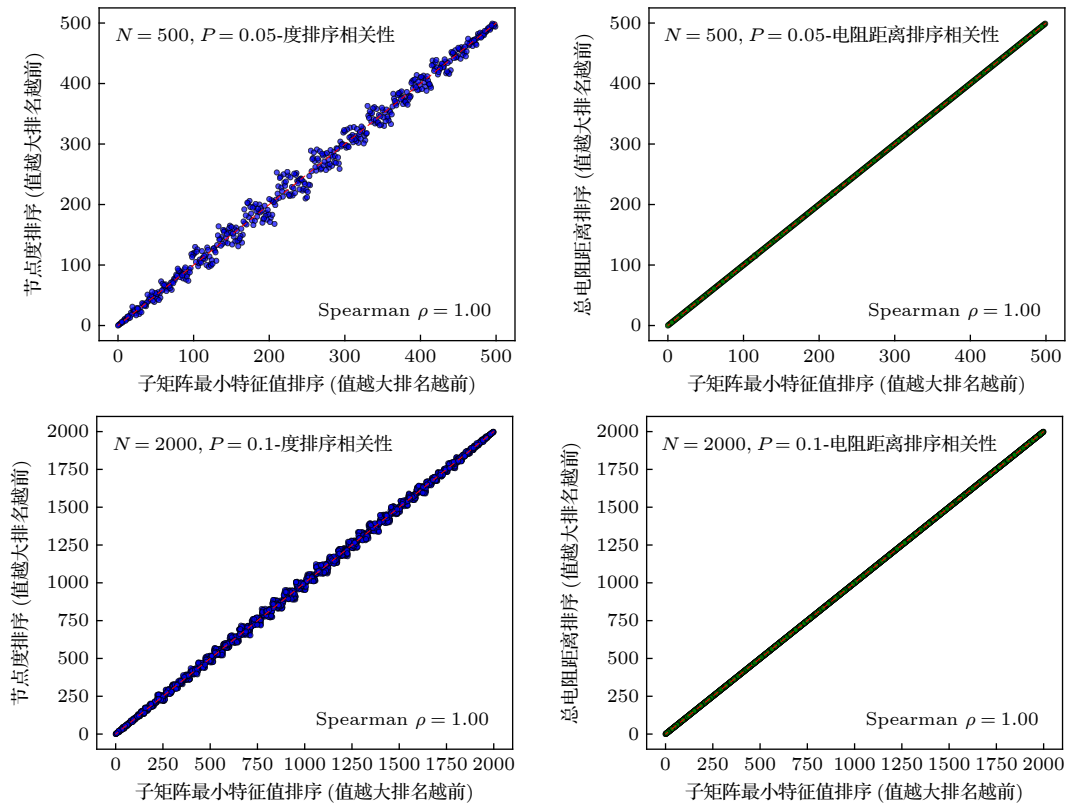


图 2 ER 随机网络节点的度及电阻距离分别与 $\lambda_1(\mathbf{L}_{N-1})$ 的排序相关性

Fig. 2. Correlation between the degree of nodes and the resistance distance in ER random networks, respectively, with the $\lambda_1(\mathbf{L}_{N-1})$.

在计算特征评估分数和归一化权重时, 需要考虑网络本身特性. 例如, 在低密度网络中, 节点之间的连接较少, 节点的介数和电阻距离影响可能较大. 因此, 在计算特征评估分数时, 需要提高平均聚类系数 \bar{C} 和路径长度 L_{ap} 的权重. 在高密度网络中, 节点的度影响较大, 此时应当提高平均度 \bar{d} 和平均加权度 \bar{d}_w 的权重. 对于其他网络, 需要根据其特定网络形态合理调整权重.

分别计算所有节点的度中心性 d_c 、介数中心性 b_c 、K-Shell 值 K_s 和电阻距离 R_d . 计算融合指标:

$$F(v) = w_{K-Shell}(K_s(v)) + w_{degree}(d_c(v)) + w_{Res-dis}(R_d(v)) + w_{bc}(b_c(v)). \quad (16)$$

更一般地, 我们提出以下多特征融合指标:

$$G(v) = \Theta [\text{feature}_i(v) \cdot \Phi(w_i)], \quad (17)$$

其中 Θ 表示对于 $\text{List}(\text{feature}_i)$ 的运算, Φ 表示 feature_i 与 w_i 的运算. 该指标可运用于所有基于图特征的算法.

在本文实验中, 通过综合指标筛选出的重要节点组对网络整体有着不俗的牵制控制效果. 由此可见, 该综合指标具有较强的说服力.

3 基于多重谱图特征线性融合指标的重要节点组挖掘算法

3.1 基于反幂法的矩阵最小特征值求解

反幂法是求解矩阵最小特征值的经典算法^[29]. 带原点偏移的反幂法可用于估计矩阵 $A_{N \times N}$ 中某个特定特征值及其对应的特征向量. 然而, 在本文的主要关注点是矩阵的最小特征值, 因此不需要输出对应的特征向量.

Liu 和 Lu^[30] 详细介绍了利用 LU 分解的反幂法的具体过程. 在此简要概述了该方法的基本思想以及 LU 分解在其中所起的作用, 而不深入探讨具体的操作步骤.

反幂法的基本思想是通过迭代寻找与给定矩阵的最小特征值对应的近似特征向量. 在这一过程中, LU 分解将矩阵分解为一个下三角矩阵 L 和一个上三角矩阵 U , 以便于求解迭代过程中产生的线性系统. 具体而言, LU 分解通过提供高效的前向和后向代换, 帮助加速向目标特征向量的收敛. LU 分解在该方法中起着至关重要的作用, 其提供了一种计算上高效的方式, 用于解决反幂迭代中每

一步所需的线性系统. 这反过来又使得该方法对于直接处理计算上不可行的较大矩阵变得切实可行.

3.2 MFG 算法流程

为了找到社交网络中的重要节点组, 我们希望建立算法挖掘社交网络模型上的关键节点组. 上文和文献^[1]证明了在牵制控制中, 节点组的删后拉普拉斯矩阵最小特征值的大小与节点组的重要性有关. 找到使得删后拉普拉斯矩阵最小特征值最大的节点组, 对其针对性进行免疫措施或优先控制, 将取得最优效果. 但最优的搜索仍然是一个求解时间复杂度会随着问题规模的增大而指数级增长的问题 (NP 难问题), 高计算复杂度主要来自两个方面: 1) 高维矩阵特征值的计算; 2) 最优节点组的大量组合.

为了一定程度上解决这两个问题, 我们提出了“基于多重图特征线性融合指标的重要节点组挖掘算法”, 利用度中心性、介数中心性、K-Shell 值以及电阻距离基本指标, 用网络图特征将指标线性融合为综合指标, 通过 MapReduce 分治法^[31] 和简单贪心搜索筛选网络重要节点组, 可以一定程度上解决节点组组合数过大的问题, 并通过二阶邻域局部扰动和全局随机搜索优化的贪心搜索避免传统贪心算法容易陷入局部最优的弊端; 进一步地, 采用反幂法计算矩阵最小特征值, 从而加快了算法收敛速度, 使大规模网络问题可解.

下面首先介绍本文提出的挖掘关键节点组算法. 定义 T 为目标筛选出的节点组集合, T 为中间操作的节点组集合, k 为受控节点个数. 算法思路可以分为两步: 第 1 步计算每个节点的删后拉普拉斯矩阵的最小非零特征值, 选择此值最大的节点作为初始迭代节点; 第 2 步是以已选节点为基础, 通过计算复杂度低的预筛选算法预选出影响力较大的一些节点作为接下来挖掘节点的候选, 然后分别计算已选节点与某个候选节点组成节点组的 Laplacian 删后矩阵最小特征值, 选择此值最大的节点作为挖掘节点, 以此类推, 直至选至 k 个节点. MFG 算法

步骤 1: 计算图的全局图属性, 图密度 D 、平均度 \bar{d} 、平均网页排名 P 、平均加权度 \bar{d}_w 、平均路径长度 L_{ap} 和平均聚类系数 \bar{C} .

步骤 2: 利用步骤 1 得到的图结构信息, 根据 (14) 式分别计算度中心性、介数中心性、K-Shell

值和电阻距离的特征评估分数, 对评估分数进行归一化, 并根据 (15) 式计算各个特征的归一化权重.

步骤 3: 分别计算所有节点的度中心性、介数中心性、K-Shell 值和电阻距离. 再结合各特征的归一化权重, 根据 (16) 式计算融合指标.

步骤 4: 将所有节点切分为 m 个子数据集, 筛选出每个子数据集中融合指标最大的前 p 个节点, 再利用归并排序汇总各个数据集的筛选结果, 筛选出最大的前 p 个节点, 构成预选节点组 S .

步骤 5: 依次选择某个预选节点组 S 中尚未选择过的节点, 将此节点与节点组 T 求并得到节点组 T' , 用反幂法计算删去节点组 T' 后的 Laplacian 矩阵的最小特征值 $\lambda_1(\mathbf{L}_{N-T'})$, 选取使得最小特征值最大的预选节点 v^* .

步骤 6: 选取 v^* 的二阶邻接节点集 $N(v^*)$, 筛去不在预选节点组中的节点和在节点组 T 里的节点, 如果结果为空, 则跳过进行下一步. 对 $N(v^*)$ 中的节点进行电阻距离排序 (步骤 3 已经计算出了每个节点的电阻距离), 选择电阻距离最小的节点 u^* , 计算不删去 v^* 而删去 u^* 的最小特征值 $\lambda_1(\mathbf{L}_{N-T-u^*})$ 和不删去 u^* 而删去 v^* 的最小特征值 $\lambda_1(\mathbf{L}_{N-T-v^*})$. 若 $\lambda_1(\mathbf{L}_{N-T-u^*}) > \lambda_1(\mathbf{L}_{N-T-v^*})$, 则选取 u^* 加入 T' , 反之则选取 v^* 加入 T' .

步骤 7: 每进行 s 步搜索 (其中 $(s < k, k \equiv 0(\text{mod } s))$), 尝试随机选取一个未被选取过的节点 $e \in S \cap e \notin T$, 移除 T 中任意一个节点 g , 计算以节点 e 代替节点 g 后的删后 Laplacian 矩阵的最小特征值 $\lambda_1(\mathbf{L}_{N-1})$ 的变化. 若变大, 则维持更改; 若变小, 则放弃更改.

步骤 8: 循环操作步骤 5—步骤 7, 直到筛选出的受控节点组规模达到 k , 将节点组 T 输出, 得到目标筛选节点组 Γ .

MFG 算法流程图如图 3 所示. MFG 算法伪代码详见补充材料 (online).

3.3 MFG 算法减少 $\lambda_1(\mathbf{L}_{N-1})$ 计算量的理论分析

与 MATLAB 库中的 eig 函数相比, 使用反幂法可以有效地减少计算量. 接下来, 我们将对两种计算特征值的方法进行比较. eig 函数采用的是 QR 算法. 当目标是求解最小特征值时, 必须先计算所有特征值, 对它们进行排序, 然后确定最小的非零特征值. 首先我们将分析 QR 方法的计算复杂度.

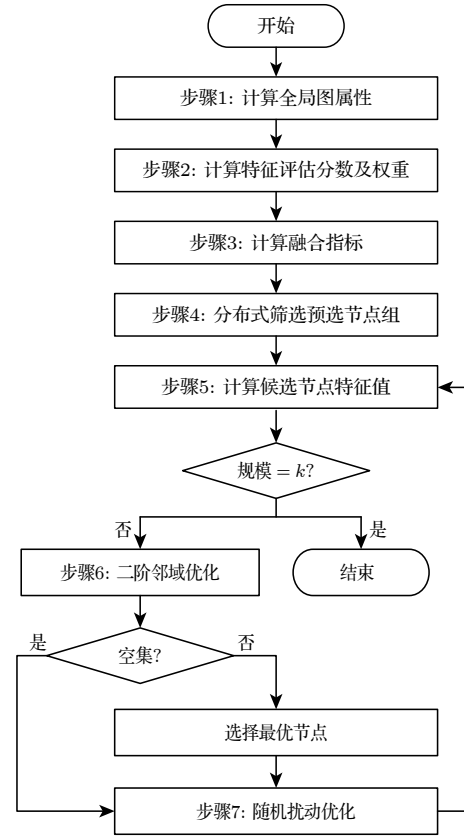


图 3 MFG 算法的流程图
Fig. 3. Flowchart of MFG Algorithm.

根据文献 [30], 如果 QR 算法的迭代次数为 T , 则执行的乘法总次数为

$$\frac{2}{3}N^3 + (4T - 1)N^2 + \frac{7}{3}N - 4T - 5, \quad (18)$$

其中 N 为矩阵维度. 而反幂法求解最小特征值, 若迭代次数记为 t , 此过程乘法次数约为 $t \cdot N(N - 1)$. 此方法求解一次最小特征值所需乘法计算次数约为

$$\frac{1}{3}N^3 - \frac{1}{2}N^2 + \frac{1}{6}N + t \cdot N(N - 1). \quad (19)$$

当 $T = t$, 计算一次 n 阶矩阵的最小特征值, 反幂法的乘法计算数量减少:

$$d(t, n) = (3N^2 + N - 4) + \left(\frac{1}{3}N^3 - \frac{1}{2}N^2 + \frac{13}{6}N - 5 \right). \quad (20)$$

根据 3.2 节的算法流程, 若每次选择 p 个节点为预选点, 目标为挖掘 k 个节点组成的节点组, 在挖掘第 i 个节点 ($i = 2, 3, \dots, k$) 时, 需要计算最小特征值的矩阵阶数为 $N - i$, 减少的乘法次数为 $p \cdot d(t, N - i)$, 挖掘后 $k - 1$ ($k \geq 2$) 个节点共减少乘法计算数目为

$$\begin{aligned}
 \sum_{i=2}^k p \cdot d(t, N-i) &= p \sum_{i=2}^k d(t, N-i) = p \sum_{i=2}^k \left\{ [3(N-i)^2 + (N-i) - 4]t + \left[\frac{1}{3}(N-i)^3 - \frac{1}{2}(N-i)^2 \right. \right. \\
 &+ \left. \left. \frac{13}{6}(N-i) - 5 \right] \right\} = p \left\{ \frac{1}{3} \left[\frac{(N-2)(N-1)}{2} \right]^2 + \left(3t - \frac{1}{2} \right) \frac{(N-2)(N-1)(2N-3)}{6} \right. \\
 &+ \left. \left(t + \frac{13}{6} \right) \frac{(N-2)(N-1)}{2} - 4t - 5 \right\} - p \left\{ \frac{1}{3} \left[\frac{(N-k)(N-k+1)}{2} \right]^2 \right. \\
 &+ \left. \left(3t - \frac{1}{2} \right) \frac{(N-k)(N-k+1)(2N-2k-1)}{6} + \left(t + \frac{13}{6} \right) \frac{(N-k)(N-k+1)}{2} - 4t - 5 \right\}. \quad (21)
 \end{aligned}$$

若考虑较大总计算量的情况, 设 $k = N - 1$, 即选取了 $N - 1$ 个节点组成受控节点组, 此时减少的乘法计算数目为

$$\begin{aligned}
 &p \left\{ \frac{1}{3} \left[\frac{(N-2)(N-1)}{2} \right]^2 \right. \\
 &+ \left. \left(3t - \frac{1}{2} \right) \frac{(N-2)(N-1)(2N-3)}{6} \right. \\
 &+ \left. \left(t + \frac{13}{6} \right) \frac{(N-2)(N-1)}{2} \right\}, \quad (22)
 \end{aligned}$$

计算量的减少达到了 N^4 量级.

此外, 当迭代次数设为 $t = 100$ 时, 我们为矩阵阶数 N 和预选节点数 p 赋值, 得到减少的乘法次数, 如表 1 所示. 与 QR 算法相比, 反幂法在乘法次数上有显著的减少.

表 1 与 QR 算法相比, 反幂法在乘法次数上的减少情况
Table 1. In comparison to the QR algorithm, the inverse power method exhibits a reduction in the number of multiplications.

N	p	乘法次数减少
100	4	4.280556×10^8
	8	8.56112×10^8
	12	1.28417×10^9
1000	4	7.31605×10^{11}
	8	1.46321×10^{12}
	12	2.19481×10^{12}
10000	4	3.73196×10^{15}
	8	7.46392×10^{16}
	12	1.11959×10^{17}

3.4 MFG 算法减少计算量的实验分析

为了验证 MFG 算法减少 $\lambda_1(\mathbf{L}_{N-1})$ 计算量的有效性, 我们在 E-mail 网络^[32] 上进行代码耗时实

验. 测试软件为 Matlab R2023b, 代码部分除了计算 $\lambda_1(\mathbf{L}_{N-1})$ 部分以外均相同, 计算 $\lambda_1(\mathbf{L}_{N-1})$ 分别使用反幂法和 Matlab 的 eig 函数, 取 $1 \leq k \leq 12 (k \in \mathbb{Z})$, $s = 3$, $p = 2k$, 记录每个 k 取值的程序耗时 (实验取 5 次测试平均值), 实验结果如图 4 所示.

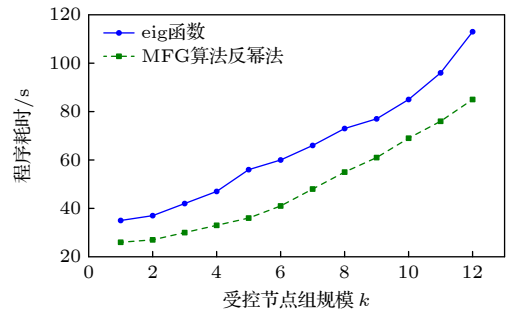


图 4 在 E-mail 网络中, 取 $1 \leq k \leq 12 (k \in \mathbb{Z})$, $s = 3$, $p = 2k$ 时, 使用 eig 函数和反幂法的程序耗时对比
Fig. 4. In the E-mail network, when taking $1 \leq k \leq 12 (k \in \mathbb{Z})$, $s = 3$ and $p = 2k$, the comparison of the computational time between the eig function and the inverse power method.

实验结果表明, MFG 算法能够大幅提升 $\lambda_1(\mathbf{L}_{N-1})$ 的计算效率, 降低计算复杂度和程序耗时. MFG 算法对反幂法的使用不仅有利于挖掘重要节点组, 也为其他涉及谱分析的问题提供了有益思路.

3.5 MFG 算法参数

根据 3.3 节的讨论, 预选节点数 p 、搜索步长 s 和受控节点组规模 k 是 MFG 算法中最重要的几个影响算法精度和复杂度的参数. 一般而言, 受控节点组规模 k 依据执行任务不同而确定, 因而无需进行技术性讨论, 但是其对计算复杂度有显著影响.

3.5.1 MFG 算法预选节点数 p

预选节点数 p 越大, 可能越容易找到最优节点组, 但是也意味着计算复杂度的大幅增加. 我们在 Email 网络上选择不同受控节点组规模 $k = 6, 12, 24$, 分别取 $6 \leq p \leq 24$, $12 \leq p \leq 48$, $24 \leq p \leq 96$ ($p \in \mathbb{Z}$), $s = 3$, 记录最终得到的最小特征值 $\lambda_1(\mathbf{L}_{N-1})$, 结果如图 5 所示.

实验结果表明, 只要 $p > k$, 最后得到的最小特征值 $\lambda_1(\mathbf{L}_{N-1})$ 几乎完全相同, 进而说明筛选出的节点组牵制控制效果几乎完全一致, 这是 E-mail 网络的网络特性决定的, 即除了个别牵制控制能力较大的几个节点外, 其余节点牵制控制能力几乎相同, 这在网络特征上可能体现为: 具有少数强中心性的节点和其余大部分中心性基本一致的节点. 因此在类似 E-mail 网络结构的网络中, 更大的 p 并不更有利于挖掘重要节点组, 反而会增加计算复杂度.

类似地, 在节点数 $N = 1000$, 邻居数 $\text{Nei} = 4$,

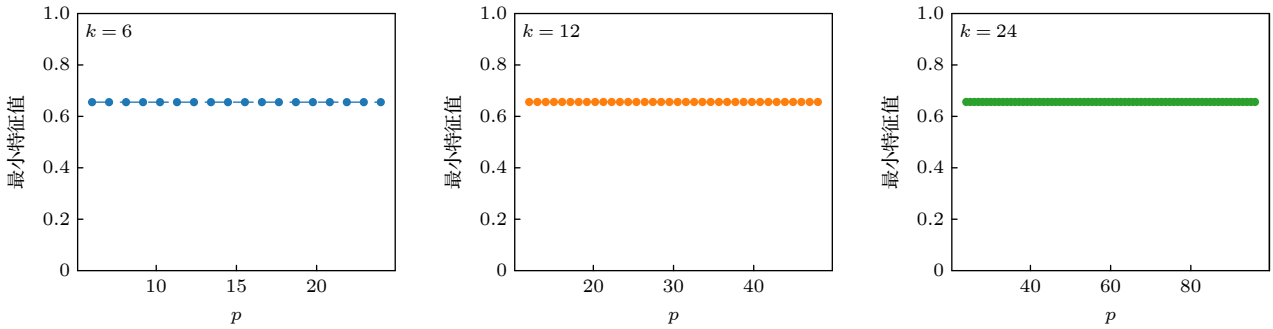


图 5 在 E-mail 网络中, 分别取 $k = 6, 12, 24$, $6 \leq p \leq 24$, $12 \leq p \leq 48$, $24 \leq p \leq 96$ ($p \in \mathbb{Z}$), $s = 3$ 时, 最终得到的最小特征值 $\lambda_1(\mathbf{L}_{N-1})$

Fig. 5. In the E-mail network, when taking respectively $k = 6, 12, 24$ and $6 \leq p \leq 24$, $12 \leq p \leq 48$, $24 \leq p \leq 96$ ($p \in \mathbb{Z}$), $s = 3$, the final minimum eigenvalue $\lambda_1(\mathbf{L}_{N-1})$.

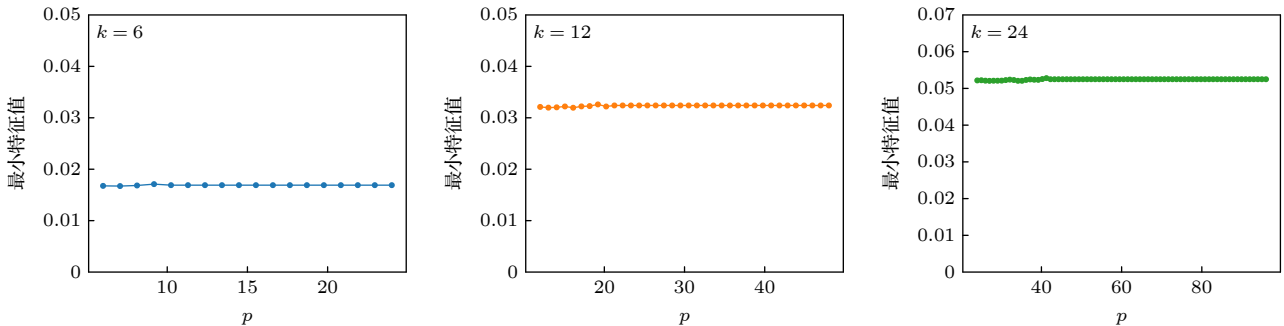


图 6 在 NW 网络 ($N = 1000$, $\text{Nei} = 4$, $p_c = 0.1$) 中, 分别取 $k = 6, 12, 24$, $6 \leq p \leq 24$, $12 \leq p \leq 48$, $24 \leq p \leq 96$ ($p \in \mathbb{Z}$), $s = 3$ 时, 最终得到的最小特征值 $\lambda_1(\mathbf{L}_{N-1})$

Fig. 6. In the NW network ($N = 1000$, $\text{Nei} = 4$, $p_c = 0.1$), when taking respectively $k = 6, 12, 24$ and $6 \leq p \leq 24$, $12 \leq p \leq 48$, $24 \leq p \leq 96$ ($p \in \mathbb{Z}$), $s = 3$, the final minimum eigenvalue $\lambda_1(\mathbf{L}_{N-1})$.

连接概率 $p_c = 0.1$ 的 NW 小世界网络上进行同上述参数相同的实验, 结果如图 6 所示.

实验结果表明, 当 $k < p \leq 2k$ 时, 最终得到的最小特征值 $\lambda_1(\mathbf{L}_{N-1})$ 有些许提升, 但总体不超过 4%; 当 $p > 2k$ 时, 无论如何增大 p , $\lambda_1(\mathbf{L}_{N-1})$ 总是不变. 因此, 略微增大 p ($k < p \leq 2k$) 能够略微增强筛选出节点组的牵制控制能力, 提升挖掘重要节点组的效果. 但是再增大 p , 并无益于挖掘重要节点组, 反而会增加计算复杂度.

综上所述, 当受控节点个数为 k 时, 预选节点数 p 介于 $k < p \leq 2k$ 较为适宜. 既能增强挖掘重要节点组的效果, 也能减少不必要的计算.

进一步看, 运用 MFG 算法时, 增大 p 对于 $\lambda_1(\mathbf{L}_{N-1})$ 提升很小, 说明了 MFG 算法能够进行非常精确有效的预筛选和贪心搜索, 体现了 MFG 算法的优势.

3.5.2 MFG 算法全局搜索步长 s

全局搜索的主要作用是增强 MFG 算法的鲁

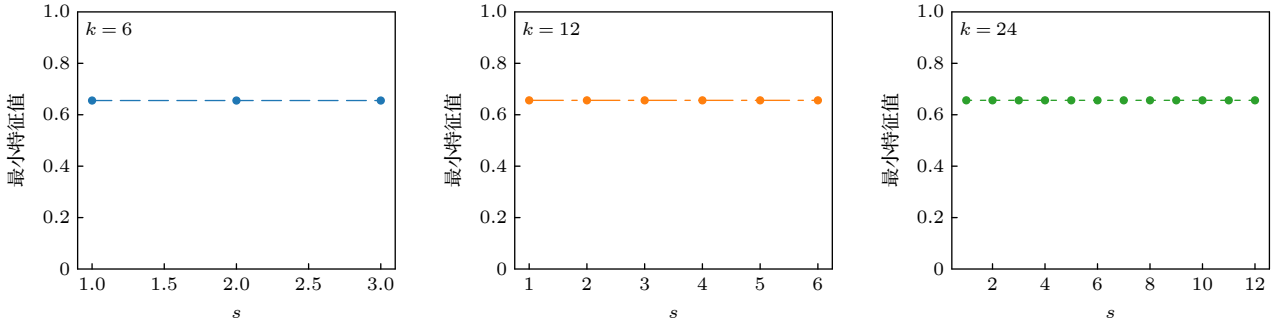


图 7 在 E-mail 网络中, 分别取 $k = 6, 12, 24$, $p = 9, 15, 30$, 当 $1 \leq s \leq 3(s \in \mathbb{Z})$, $1 \leq s \leq 6(s \in \mathbb{Z})$, $1 \leq s \leq 12(s \in \mathbb{Z})$ 时, 最终得到的最小特征值 $\lambda_1(\mathbf{L}_{N-1})$

Fig. 7. In the E-mail network, when taking respectively $k = 6, 12, 24$, $p = 9, 15, 30$ and $1 \leq s \leq 3(s \in \mathbb{Z})$, $1 \leq s \leq 6(s \in \mathbb{Z})$, $1 \leq s \leq 12(s \in \mathbb{Z})$, the final minimum eigenvalue $\lambda_1(\mathbf{L}_{N-1})$.

棒性, 避免 MFG 算法陷入局部最优的问题. 我们在 E-mail 网络上选择不同受控节点组规模 $k = 6, 12, 24$ 和预筛选节点规模 $p = 9, 15, 30$, 分别令 $1 \leq s \leq 3(s \in \mathbb{Z})$, $1 \leq s \leq 6(s \in \mathbb{Z})$, $1 \leq s \leq 12(s \in \mathbb{Z})$, 记录最终得到的最小特征值 $\lambda_1(\mathbf{L}_{N-1})$, 结果如图 7 所示.

可以发现, $\lambda_1(\mathbf{L}_{N-1})$ 几乎不随 s 变化而变化. 这是 E-mail 网络本身网络特性所决定的, 即每次挖掘得到的最优节点就是全局最优节点. 但这并不意味着全局搜索没有作用, 其依然可以以较低的时间复杂度增强 MFG 算法对于不同网络的泛化能力. 因此, s 取适中偏大的值较好, 能让 MFG 算法在复杂环境中保持稳定性能.

4 实验仿真和算法有效性分析

本节将 MFG 算法分别应用于不同网络中, 进行选点效果的对比, 通过和其他挖掘算法在不同受控节点组规模下进行对比说明 MFG 算法在挖掘删除节点组后 Laplacian 矩阵最小特征值 $\lambda_1(\mathbf{L}_{N-1})$ 较大的节点组时更优, 最后以这些节点作为初始感染节点进行 SIR 传播模拟进一步验证了所选节点组的重要性. 若感染节点数峰值更大、上升更快, 表明对应的初始感染节点在复杂网络中牵制控制能力强. 在 SIR 仿真实验中, 得到的结果为时间步 $\text{step} = 1000$ 中 7 种算法下感染数. 为了降低偶然性, 最终结果中每一时刻感染人数取 10 次实验的平均值. 为了检测 MFG 算法的参数敏感度和鲁棒性, 对于每个 SIR 仿真实验, 采取 $\beta = 0.01, \gamma = 0.01$; $\beta = 0.015, \gamma = 0.01$; $\beta = 0.02, \gamma = 0.01$; $\beta = 0.03, \gamma = 0.01$; $\beta = 0.03, \gamma = 0.015$; $\beta = 0.03, \gamma = 0.02$;

$\beta = 0.03, \gamma = 0.025$; $\beta = 0.03, \gamma = 0.03$ 共 8 组参数. 实验对比的近年算法如下.

1) Wu 等 [33] 提出了节点传播熵度量, 该度量结合了局部和全局拓扑信息, 综合考虑了节点聚类系数、邻居数量、K-Shell 分解以及最短路径, 以捕捉每个节点在网络中传播信息的潜力, 下文简称 NPE.

2) Meng 等 [34] 提出了一种非对称引力模型 (asymmetric gravity model, AGM), 通过考虑非对称吸引效应来识别复杂网络中的关键节点. 该算法的核心思想是通过累积指定影响距离内邻居节点的吸引力来计算一个节点的影响力.

3) Zhu 等 [35] 提出了一种新的基于 H 指数的引力中心性方法 HVGC, 基于 H 指数, 综合考虑邻近节点的影响、路径信息和节点位置. 这种方法能更好地识别在不同网络部分间起桥梁作用的节点.

4.1 NW 小世界网络算法仿真实验分析

4.1.1 NW 小世界网络最小特征值和选点分析

例 1 在本实验中, 选择节点数目 $N = 1000$ 、连边概率 $p = 0.1$ 、邻居数目 $\text{Nei} = 4$ 作为生成 NW 小世界网络的参数组合, 该选择在理论与实践上均具有良好的合理性与代表性. 首先, 设定 $\text{Nei} = 4$ 意味着每个节点在初始规则网络中与四个邻居相连, 形成了相对紧密的局部结构, 有助于保持较高的聚类系数. 这种局部稠密连接使得网络在初始状态下已经具备一定程度的社区结构, 为小世界性质的形成提供基础. 与此同时, 选择较小但非极小的

连边概率 $p_c = 0.1$, 使得在原有规则结构上适度引入了长程随机边, 从而显著降低网络的平均路径长度, 有效提升全局信息传播效率. 根据小世界网络理论, 适中的 p_c 值 (通常在 0.01—0.1 之间) 是实现小世界特性的关键区域, 能够在保持较高聚类系数的同时大幅度压缩路径长度, 达到“局部有序、全局高效”的结构特性. 相比更大的 p_c , 该设置避免了网络过度随机化而导致聚类性下降; 相比更小的 p_c , 又能更明显地增强节点间的远程可达性. 因此, $N_{ei} = 4$ 和 $p_c = 0.1$ 的组合能够在规则性与随机性之间实现良好平衡, 既保留了小世界网络的关键特征, 又为后续的分析任务提供了稳定而具有代表性的网络基础. 生成的 NW 小世界网络见图 8, 其中度越大的节点在尺寸上越大. NW 小世界网络属性, 如平均度、平均加权密度、密度、平均聚类系数、特征向量中心性和平均路径长度分别为 4.000, 8.000, 0.004, 0.383, 0.281 和 9.182.

将受控节点组规模 k 依次增加, 7 种算法删除节点组后的 Laplacian 矩阵最小特征值如表 2 所示, 各算法挖掘所得的节点重要性排序如表 3 所示, 图 9 显示了 7 种算法删除节点后的删后 Laplacian 矩阵最小特征值 $\lambda_1(L_{N-k})$ 的可视化比较.

表 2 和图 9 的结果表明, 在受控节点组规模不同的情况下, MFG 算法在小世界网络中得到的 Laplacian 删后矩阵最小特征值 $\lambda_1(L_{N-k})$ 均大于其他算法, 证明了 MFG 算法的有效性, 而且随着节点组规模的不断增大, 删后 Laplacian 矩阵的最小特征值都在不断增大.

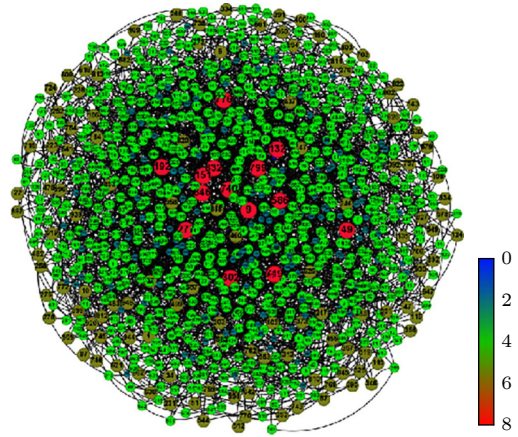


图 8 生成的 NW 小世界网络 ($N = 1000$, $N_{ei} = 4$, $p_c = 0.1$), 标度尺反映了节点度大小的情况

Fig. 8. Generated NW small world network ($N = 1000$, $N_{ei} = 4$, $p_c = 0.1$), the scale reflects the magnitude of node degrees.

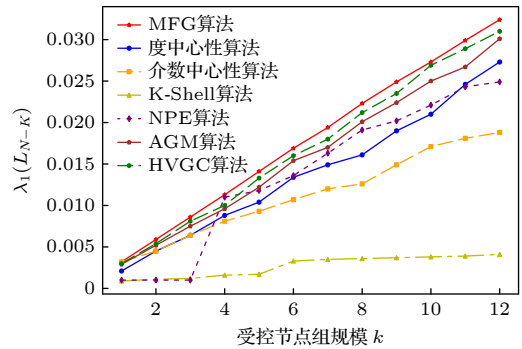


图 9 NW 网络中不同算法去除节点后不同受控节点组规模下最小特征值的比较

Fig. 9. Comparison of minimum eigenvalue with different target node counts following node removal by different algorithms in NW network.

表 2 不同算法在小世界网络中 $\lambda_1(L_{N-k})$ 的对比

Table 2. Comparison of $\lambda_1(L_{N-k})$ in the NW network for different algorithms.

受控节点组规模 k	度中心性算法	介数中心性算法	K-Shell算法	NPE算法	AGM算法	HVGC算法	MFG算法
1	0.0021	0.0032	0.0009	0.0010	0.0029	0.0030	0.0032
2	0.0045	0.0045	0.0011	0.0010	0.0052	0.0054	0.0059
3	0.0064	0.0064	0.0012	0.0010	0.0075	0.0081	0.0086
4	0.0088	0.0081	0.0016	0.0110	0.0096	0.0100	0.0113
5	0.0104	0.0093	0.0017	0.0118	0.0122	0.0133	0.0141
6	0.0134	0.0107	0.0033	0.0136	0.0154	0.0160	0.0169
7	0.0149	0.0120	0.0035	0.0163	0.0170	0.0180	0.0194
8	0.0161	0.0126	0.0036	0.0191	0.0201	0.0212	0.0223
9	0.0190	0.0149	0.0037	0.0202	0.0224	0.0235	0.0249
10	0.0210	0.0171	0.0038	0.0221	0.0250	0.0269	0.0273
11	0.0246	0.0181	0.0039	0.0243	0.0267	0.0289	0.0299
12	0.0273	0.0188	0.0041	0.0249	0.0301	0.0310	0.0324

表 3 小世界网络不同算法挖掘所得的节点重要性排序

Table 3. Node importance ranking by the different algorithms in NW network.

度中心性算法	介数中心性算法	K-Shell算法	NPE算法	AGM算法	HVGC算法	MFG算法
4	616	838	616	616	616	616
121	924	839	329	207	523	523
198	329	837	595	523	207	207
236	207	840	207	371	595	236
329	595	238	924	236	924	371
363	236	239	145	887	307	595
371	382	868	307	417	417	417
417	145	869	523	329	329	329
523	307	237	339	339	701	887
560	146	240	676	409	887	701
612	814	836	409	382	382	382
616	915	841	614	937	937	937

此外,由表 3 结果可看出,当受控节点组规模过小时,由于少数几个节点对于网络的重要性十分显著,7种算法都会筛选到这些节点,导致所选出的节点重合率较高,当受控节点组过大时,很容易囊括所有较为重要的节点,使得 MFG 算法的优势不明显.从表 2 也可看出,当节点组规模较为适中

时 MFG 算法所筛选出的节点组对应的删后 Laplacian 矩阵最小特征值 $\lambda_1(L_{N-1})$ 相对于其余算法的增加量更显著.

4.1.2 NW 小世界网络 SIR 模型仿真分析

NW 小世界网络模型在 SIR 传播仿真中起着至关重要的作用,因为其能有效平衡局部连接性和全局连通性.小世界特性使得网络中大部分节点仅通过少数中介节点就能相互连接,这种结构促使信息在网络中迅速传播.这些特性使得小世界网络成为研究社交网络信息传播机制的理想模型,提供了对复杂网络中传播动态的深刻理解.

考虑网络信息同上的 NW 小世界网络模型,将上述得到的重要节点组分别作为初始感染者,进行感染仿真.感染数随时间的变化如图 10 所示.可以明显看出, MFG 算法相比于其他算法在早期阶段能够更快速地实现感染的扩散,在数据上体现为曲线更大的极值和更大的斜率,显示出其在动态网络传播中的优势.这一现象表明, MFG 算法在识别和选择复杂网络中的关键节点方面具有显著的能力,能够有效地促进传播速度的提升.在实际应用中,能够为快速响应和控制重大社交传播事件提供重要的理论支持.

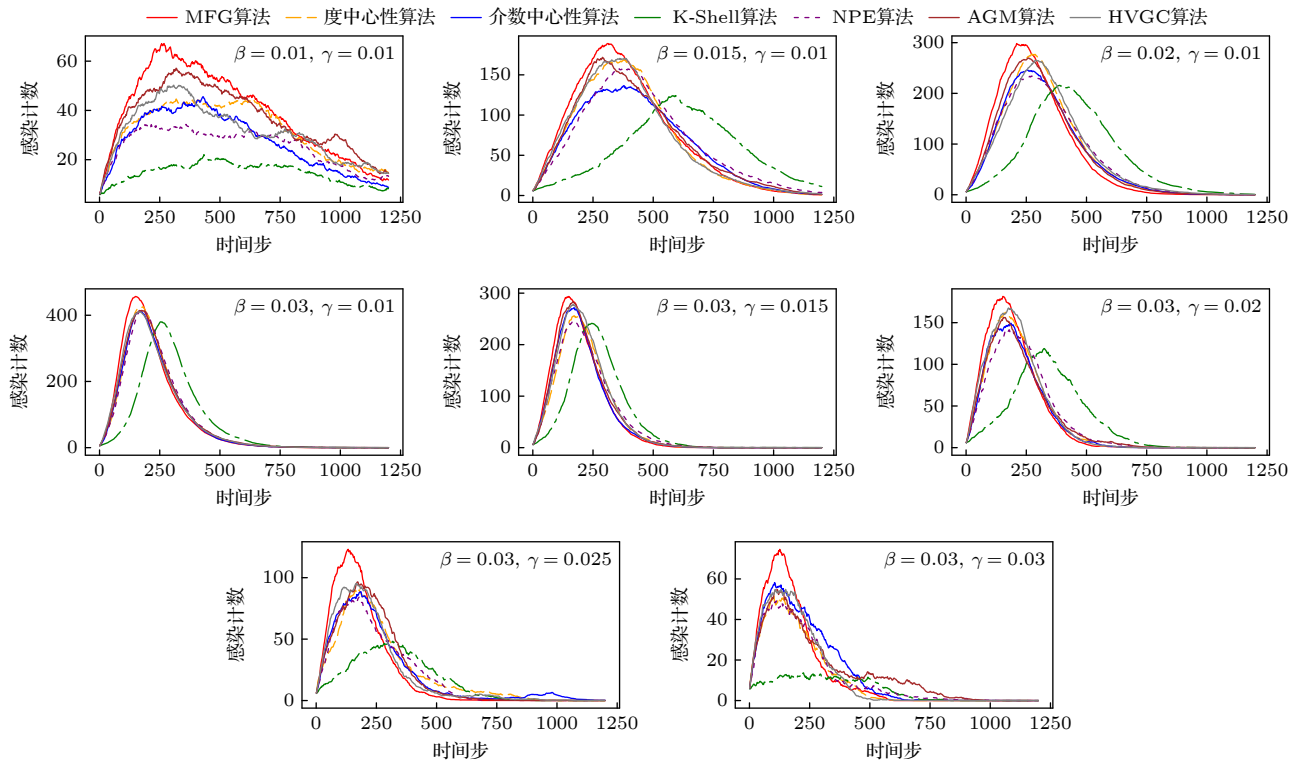


图 10 小世界网络模型中感染数随时间的变化

Fig. 10. Changes in the number of infections over time in the NW network model.

4.2 真实网络算法仿真实验分析

虽然 NW 小世界网络在模拟短路径特性方面表现良好, 但真实社交网络在多个方面更具优势. 一方面, 真实社交网络的节点连接通常具有显著的异质性, 而 NW 模型假设连接概率均匀, 无法有效反映这一特征. 另一方面, 真实社交网络的连接受到地理和功能性限制, NW 网络的随机重连机制无法捕捉这些复杂关系. 此外, 真实社交网络展现出动态演化和明显的社区结构, 而 NW 小世界网络在这些方面显得力不从心. 因此, 尽管 NW 小世界网络提供了有价值的视角, 真实社交网络在描述复杂性和动态性上更为准确.

4.2.1 真实网络最小特征值和选点分析

例 2 在网站^[36]上搜集了规模大小为 $N = 7642$ 的真实社交网络 lastfm_asia 数据用于仿真对比实验. 网络节点图见图 11, 网络属性平均度、平均加权密度、密度、平均聚类系数、特征向量中心性和平均路径长度分别为 7.294, 14.589, 0.001, 0.285, 0.3148 和 5.232.

将受控节点组规模 k 依次增加, 7 种算法删除节点组后的 Laplacian 矩阵最小特征值 $\lambda_1(L_{N-k})$ 如表 4 和图 12 所示, 各算法挖掘所得的节点重要性排序如表 5 所示

例 3 我们搜集了规模大小为 $N = 1005$ 的真实社交网络 E-mail 用于仿真对比实验. 网络节点图见图 13, 网络属性平均度、平均加权度、图密度、平均聚类系数、特征向量中心性和平均路径长度分别为 33.246, 65.214, 0.033, 0.473, 0.02275, 2.587.

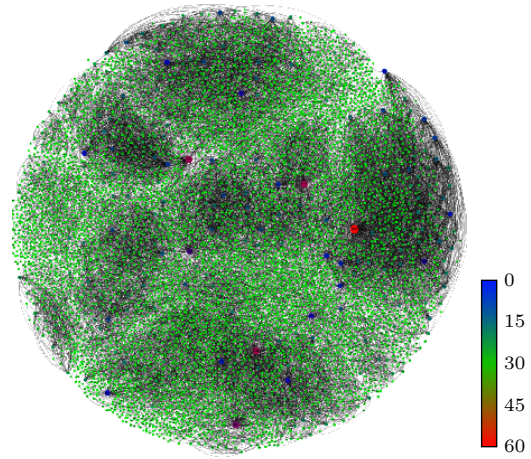


图 11 生成的真实社交网络 lastfm_asia ($N = 7642$), 标度尺反映了节点度大小的情况

Fig. 11. Generated real social network lastfm_asia ($N = 7642$), the scale reflects the magnitude of node degrees.

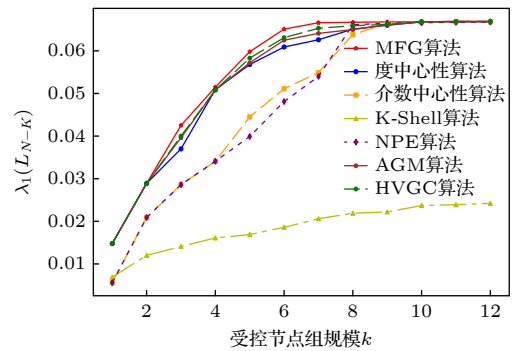


图 12 lastfm_asia 网络中不同算法去除节点后不同受控节点组规模下最小特征值 $\lambda_1(L_{N-k})$ 的比较

Fig. 12. Comparison of minimum eigenvalue $\lambda_1(L_{N-k})$ with different target node counts following node removal by different algorithms in lastfm_asia network.

表 4 不同算法在 lastfm_asia 网络中的 $\lambda_1(L_{N-k})$ 对比

Table 4. Comparison of $\lambda_1(L_{N-k})$ in lastfm_asia network for different algorithms.

受控节点组规模 k	度中心性算法	介数中心性算法	K-Shell算法	NPE算法	AGM算法	HVGC	MFG算法
1	0.0148	0.0056	0.0069	0.0056	0.0148	0.0148	0.0148
2	0.0289	0.0209	0.0120	0.0209	0.0289	0.0289	0.0289
3	0.0370	0.0286	0.0141	0.0286	0.0396	0.0400	0.0425
4	0.0509	0.0341	0.0161	0.0341	0.0508	0.0508	0.0515
5	0.0568	0.0445	0.0169	0.0399	0.0572	0.0583	0.0598
6	0.0609	0.0511	0.0186	0.0481	0.0625	0.0631	0.0651
7	0.0626	0.0549	0.0206	0.0540	0.0641	0.0653	0.0666
8	0.0651	0.0638	0.0219	0.0661	0.0650	0.0659	0.0667
9	0.0660	0.0665	0.0222	0.0665	0.0661	0.0662	0.0668
10	0.0668	0.0667	0.0237	0.0666	0.0668	0.0668	0.0668
11	0.0668	0.0668	0.0239	0.0667	0.0669	0.0669	0.0669
12	0.0668	0.0668	0.0242	0.0668	0.0669	0.0669	0.0669

表 5 lastfm_asia 网络不同算法挖掘所得的节点重要性排序

Table 5. The node importance ranking by the different algorithms in lastfm_asia network.

度中心性算法	介数中心性算法	K-Shell算法	NPE算法	AGM算法	HVGC算法	MFG算法
7238	7200	379	7200	7238	7238	7238
3531	7238	764	7238	6102	7200	3531
4786	2855	952	2855	3531	6102	6102
525	4357	1335	4357	4786	3531	4786
3451	6102	2453	5455	4357	525	525
2511	5455	3241	5128	3451	2855	1796
3598	4339	3545	3451	1796	5275	3451
2855	5128	3598	6102	5128	3451	5275
5128	3451	4810	3545	4812	4812	4812
6102	4786	4901	4901	5128	4901	2855
4812	3531	5091	4339	525	4357	4357
5579	3104	6109	3531	2855	5128	5128

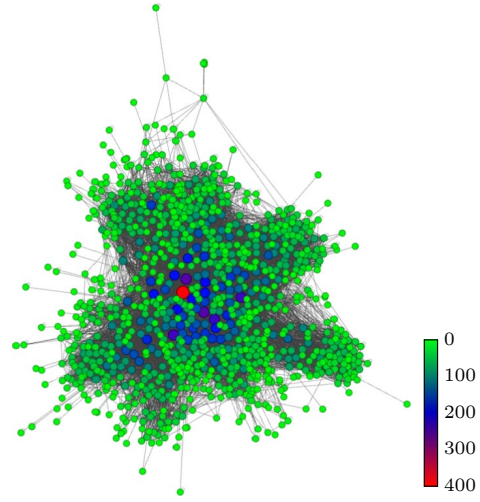


图 13 生成的真实社交网络 E-mail ($N = 1005$), 标度尺反映了节点度大小的情况

Fig. 13. Generated real social network E-mail ($N = 1005$), the scale reflects the magnitude of node degrees.

表 6 不同算法在 Email 网络中的 $\lambda_1(\mathbf{L}_{N-1})$ 对比

Table 6. Comparison of $\lambda_1(\mathbf{L}_{N-1})$ in Email network for different algorithms.

受控节点组规模 k	度中心性算法	介数中心性算法	K-Shell算法	NPE算法	AGM算法	HVGC算法	MFG算法
1	0.64884	0.64884	0.62029	0.64884	0.64884	0.64884	0.64884
2	0.65218	0.65255	0.64221	0.65255	0.65288	0.65288	0.65355
3	0.65322	0.65350	0.64845	0.65350	0.65402	0.65402	0.65420
4	0.65371	0.65402	0.65117	0.65393	0.65451	0.65451	0.65469
5	0.65420	0.65429	0.65237	0.65419	0.65478	0.65478	0.65497
6	0.65437	0.65446	0.65307	0.65445	0.65517	0.65473	0.65529
7	0.65448	0.65459	0.65389	0.65459	0.65528	0.65487	0.65539
8	0.65458	0.65470	0.65401	0.65466	0.65534	0.65498	0.65550
9	0.65466	0.65476	0.65414	0.65474	0.65542	0.65505	0.65563
10	0.65469	0.65482	0.65426	0.65477	0.65550	0.65524	0.65572
11	0.65483	0.65486	0.65429	0.65481	0.65562	0.65530	0.65588
12	0.65487	0.65487	0.65433	0.65486	0.65577	0.65546	0.65601

将受控节点组规模 k 依次增加, 7 种算法删除节点组后的 Laplacian 矩阵最小特征值 $\lambda_1(\mathbf{L}_{N-1})$ 如表 6 和图 14 所示, 各算法挖掘所得的节点重要性排序在表 7 列出.

从表 4、表 6、图 12 和图 14 结果可知, MFG 算法所筛选出的节点组对应的删后 Laplacian 矩阵最小特征值 $\lambda_1(\mathbf{L}_{N-1})$ 依旧始终领先其余 6 种算法. 说明了其在复杂网络控制中具有较好效果, 表明其在保留网络连通性和稳定性方面的显著优势.

由上述仿真可见, MFG 算法通过综合多种指标的优势, 能够在选点数较少的情况下找到最优牵制节点组, 为寻找重要节点组算法提供了一些有益思路.

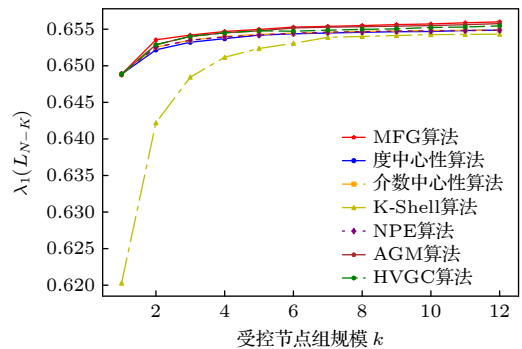


图 14 E-mail 网络中不同算法去除节点后不同受控节点组规模下最小特征值 $\lambda_1(\mathbf{L}_{N-1})$ 的比较

Fig. 14. Comparison of minimum eigenvalue $\lambda_1(\mathbf{L}_{N-1})$ with different target node counts following node removal by different algorithms in E-mail network.

表 7 E-mail 网络不同算法挖掘所得的节点重要性排序

Table 7. The node importance ranking by the different algorithms in E-mail network.

度中心性算法	介数中心性算法	K-Shell算法	NPE算法	AGM算法	HVGC算法	MFG算法
161	161	22	161	161	161	161
122	87	29	122	122	87	87
83	6	82	83	83	83	6
108	83	115	108	108	63	83
87	122	129	63	63	14	122
63	108	130	87	250	6	378
435	14	161	435	435	122	14
14	378	170	250	184	108	108
167	63	213	184	130	65	334
184	65	250	167	167	534	435
6	212	304	130	129	302	63
65	534	372	65	87	167	167

4.2.2 真实网络 SIR 模型仿真分析

在 E-mail 网络上对 SIR 模型进行了仿真分析. 当受控节点组规模 $k = 12$ 时, 选取筛选出的节点作为初始感染节点, 感染数量随时间的变化如图 15 所示.

除了上述网络外, 我们考虑了另一种形态的网

络, 在节点数规模为 $N = 1519$ 的真实网络中, MFG 算法同样得到了较好的结果. 我们在网站^[36]上, 获取了该网络的数据, 如图 16 所示, 该网络具有明显的强中心性和高聚集性, 两个主要中心节点作为枢纽, 能够快速将信息或感染扩散至周围节点. 同时, 邻近节点间的紧密连接形成了易于传播的“热点”区域, 整体传播路径较短, 提升了远距离传播效率. 网络中的连接度分布不均, 导致部分节点 (尤其是中心节点) 具有更高的影响力, 而其他节点连接较少. 此外, 两个中心节点在传播过程中存在明显竞争, 影响传播路径、速率及感染分布格局.

双中心的真实网络结构与社交网络中的关键用户情景非常相似. 在社交媒体或在线社区中, 通常会存在一些具有较高影响力的用户, 他们与大量其他用户连接. 这些中心用户的活动能够迅速影响其周围的许多关注者, 从而实现信息或观点的快速传播. 我们在 facebook_combine 网络上对 SIR 模型进行仿真分析, 当受控节点组规模 $k = 7$ 时, 选取筛选出的节点作为初始感染节点, 感染数量随时间的变化如图 17 所示. 图 15 和图 17 的实验结果表明, MFG 算法的传播能力更强, 可以通过以下观察得到.

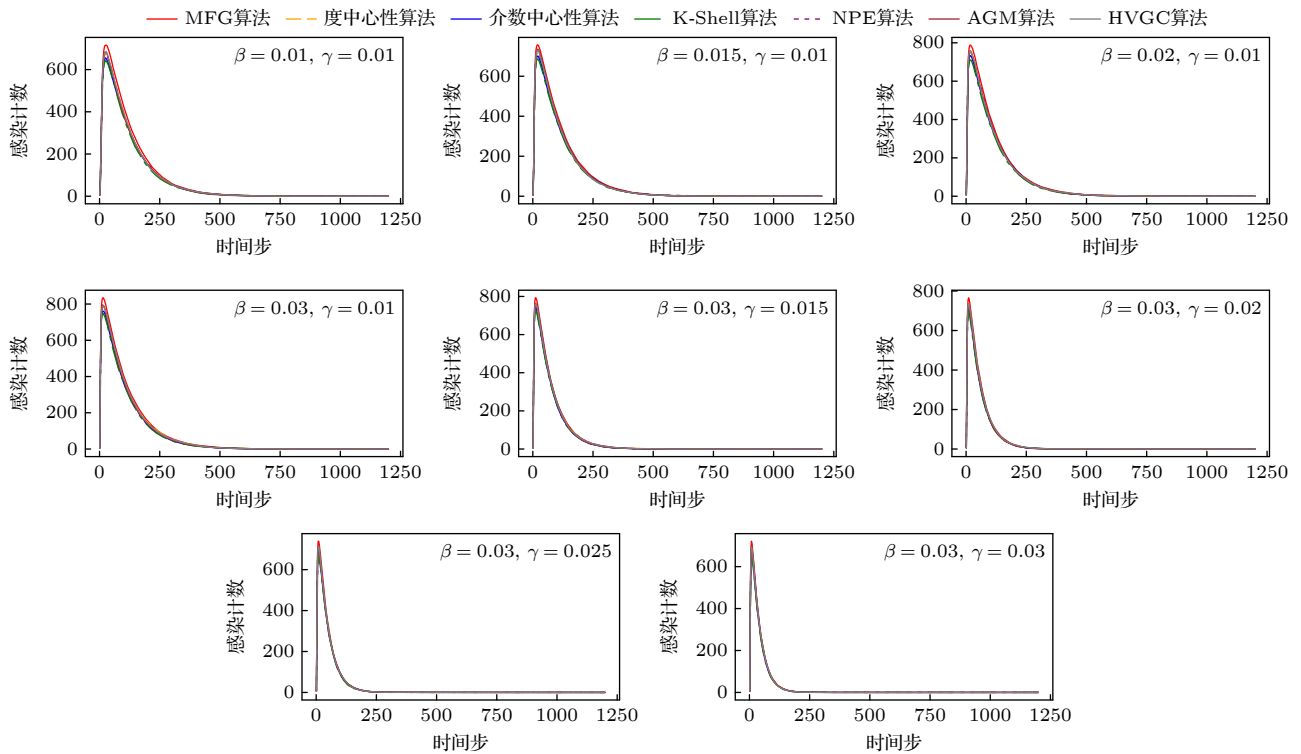


图 15 E-mail 网络感染人数随时间的变化

Fig. 15. Change in the number of infections over time steps in E-mail network.

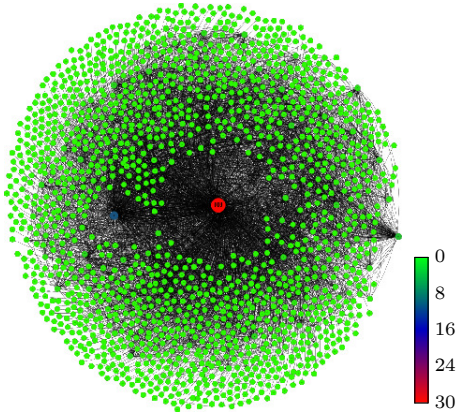


图 16 Facebook combine 网络 ($N = 1519$)

Fig. 16. Structure of the facebook combine Network ($N = 1519$).

1) 感染峰值水平: 与其他算法相比, MFG 算法的感染峰值水平明显更高. 这表明感染的传播更加有效和迅速, 表明在相同条件下的传播能力增强.

2) 累计感染计数: MFG 算法的感染曲线下面积明显较大, 反映了较高的累积感染计数. 这证明 MFG 算法在较长时间内维持感染传播的能力, 覆盖了更广泛的网络部分.

3) 传播速度和稳定性: MFG 算法不仅更快达到峰值感染水平, 而且保持较长时间的高感染率.

这种跨越不同网络结构的稳定性表明了它在不同网络传播情况下的稳健性和适应性.

因此, 根据以上论述证明了 MFG 算法的有效性.

5 基于最小特征值视角对复杂网络受控节点组的选取

5.1 复杂网络受控节点组数目

复杂网络中的受控节点选取直接影响网络控制效果. 过多受控节点会导致资源浪费和控制冗余, 增加系统不稳定性; 过少则会造成控制不足, 形成信息孤岛. 因此, 需要在控制能力和资源效率间寻求最优平衡, 这对提升网络性能和鲁棒性至关重要.

以节点数 $N = 1519$ 的 facebook_combine 网络为例, 在上文中已经说明该网络具有强中心性, 仅有两个较大的中心节点, 表 8 给出了当受控节点组规模 k 不同时 ($1 \leq k \leq 12, k \in \mathbb{Z}$) 7 种算法得到的 Laplacian 删后矩阵最小特征值 $\lambda_1(\mathbf{L}_{N-1})$. 可以看出, 当受控节点组规模不断增加时, 度中心性算法得到的 $\lambda_1(\mathbf{L}_{N-1})$ 基本保持不变, 并且较小. 因为删去两个中心节点后, 其余节点度都较小, 难以找到除了中心节点以外对网络有较强控制作用

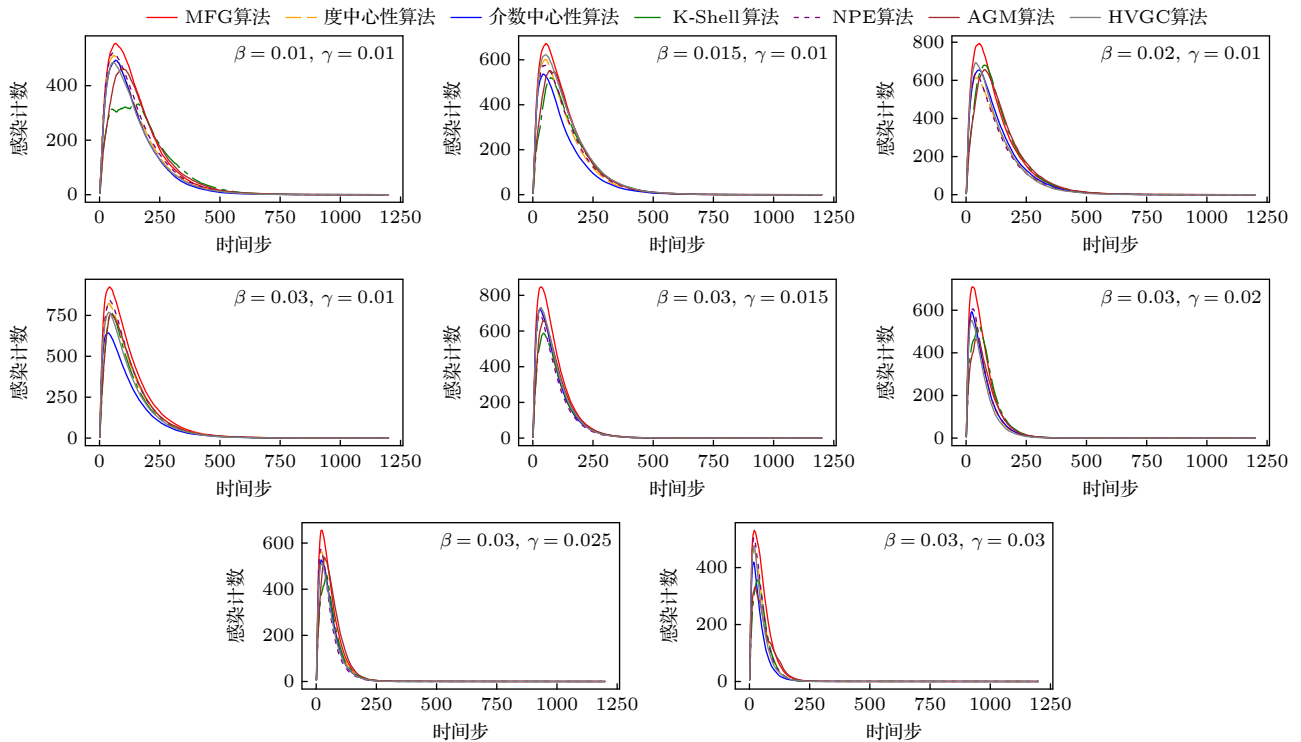


图 17 facebook_combine 网络模型 ($N = 1519$) 中感染数随时间的变化

Fig. 17. Variation of the number of infections over time in the facebook_combine network model ($N = 1519$).

表 8 不同受控节点组规模下 $\lambda_1(\mathbf{L}_{N-1})$ 对比
Table 8. Comparison of $\lambda_1(\mathbf{L}_{N-1})$ under different controlled node sizes.

受控节点组规模 k	度中心性算法	介数中心性算法	K-Shell算法	NPE算法	AGM算法	HVGC算法	MFG算法
1	0.0137	0.0137	0.0061	0.0056	0.0137	0.0137	0.0137
2	0.1507	0.1507	0.0061	0.0209	0.1507	0.1507	0.1507
3	0.2262	0.2262	0.0061	0.0286	0.2262	0.2262	0.2262
4	0.2263	0.2262	0.0061	0.0341	0.3674	0.5315	0.5789
5	0.2263	0.5789	0.0061	0.0399	0.5899	0.6987	0.7033
6	0.2263	0.7028	0.0061	0.0481	0.7675	0.7543	0.8032
7	0.2263	0.7035	0.0061	0.0540	0.8236	0.9275	1.0000
8	0.2263	0.7288	0.0061	0.0661	1.0000	1.0000	1.0000
9	0.2263	1.0000	0.0061	0.0665	1.0000	1.0000	1.0000
10	0.2263	1.0000	0.0061	0.0666	1.0000	1.0000	1.0000
11	0.2263	1.0000	0.0061	0.0667	1.0000	1.0000	1.0000
12	0.2263	1.0000	0.0062	1.0000	1.0000	1.0000	1.0000

的节点. K-Shell 算法得到的 $\lambda_1(\mathbf{L}_{N-1})$ 也较小且保持不变, 这是由于该网络的中心性结构决定了 K-Shell 值较大的节点也位于网络的核心地位, 与度中心性算法重合度较高. 介数中心性算法得到的 $\lambda_1(\mathbf{L}_{N-1})$ 随着受控节点组规模 k 增加而逐渐提升, 在受控节点组规模 k 达到一定规模时 $\lambda_1(\mathbf{L}_{N-1})$ 保持不变. 这是因为介数中心性算法通常会找出网络中具有路径中心性的节点, 而不仅仅依赖连通性, 因此其选取的受控节点组对其他节点控制力较强, $\lambda_1(\mathbf{L}_{N-1})$ 较大. 而 k 达到一定规模后 $\lambda_1(\mathbf{L}_{N-1})$ 不变, 则表明对于特定网络, 选取的受控节点组规模量过多时并无益于网络的控制, 反而会浪费大量计算资源. AGM 算法和 HVGC 算法虽较度算法和 K-Shell 算法有所提升, 但仍需 7—8 个节点才能接近完全控制. 而 NPE 算法表现出明显的滞后性, 直到控制 12 个节点时 $\lambda_1(\mathbf{L}_{N-1})$ 才突增至 1.0000, 这说明其节点评估策略存在较大优化空间, 难以有效识别网络中的关键控制节点. 在此背景下, MFG 算法展现出较大的性能优势: 它不仅以 7 个节点的最小规模使 $\lambda_1(\mathbf{L}_{N-1})$ 达到 1.0000, 比次优算法减少 12.5% 的控制成本, 更在中间过程 (5 节点时 $\lambda_1(\mathbf{L}_{N-1}) = 0.7033$) 就达到介数中心性算法 9 节点的效果, 这表明其能精准识别网络控制的重要节点组, 为资源受限环境下的网络控制提供了较优解决方案.

5.2 复杂网络最佳控制点拓扑

最佳受控制点拓扑的选择与复杂网络的结构

特征密切相关, 尤其是中心性、连通性和社区结构等方面. 例如, 在强中心性网络中, 通常会选择具有高连接度的节点作为受控制点, 因为这些节点能够快速传播信息并影响周围节点, 上文 $N = 1519$ 的 facebook_combine 网络已经说明了这个结论. 而在弱中心性网络中, 由于节点的连接度相对均匀, 选择基于介数中心性或特征向量中心性的方法来确定受控制点更为有效. 在具有明确社区结构的网络中, 最佳受控制点应该是那些跨社区的节点, 这样的选择可以增强不同社区之间的连接性, 从而提高整体网络的控制能力. 在动态适应性网络中, 受控制点的选择则需要实时评估节点的适应性和影响力, 以便根据网络的变化及时调整控制策略. 对于高度不均匀的网络, 通常优先选择连接度高的“超级节点”作为控制点. 这是我们在大量实验过程中体会到的偏好. 基于此, MFG 算法能兼顾网络特征给出最优的节点组选择.

6 结 论

本文首先介绍了传播动力学模型, 第 3 节提出了一种挖掘网络中重要性节点组的算法, 对算法复杂度的改进进行理论分析, 算法在计算复杂度及网络规模上有较大的提升. 第 4 节中对算法性能、传播效果进行了仿真实验分析, 实验结果表明, 我们提出的重要节点组挖掘算法针对重要目标节点有较好的挖掘效果, 在此基础上给出了节点优先控制建议. 在以后的研究中, 将分析节点组重要性在有

向加权网络^[37]中的情况,并希望进一步探讨牵制控制理论在动态网络谱图扰动分析及在信息传播调控策略上的指导.

参考文献

- [1] Liu H, Xu X H, Lu J A, Chen G R, Zeng Z G 2021 *IEEE Trans. Syst. Man Cybern. Syst.* **51** 786
- [2] Liu H, Wang B J, Lu J A, Li Z Y 2021 *Acta Phys. Sin.* **70** 056401 (in Chinese) [刘慧, 王炳珺, 陆君安, 李增扬 2021 物理学报 **70** 056401]
- [3] Zhou F, Su C, Xu S Q, Lü L Y 2022 *Chin. Phys. B* **31** 068901
- [4] Dai J Y, Wang B, Sheng J F, Sun Z J, Khawaja F R, Ullah A 2019 *IEEE Access* **7** 131719
- [5] Kong J T, Huang J, Gong J X, Li E Y 2018 *Acta Phys. Sin.* **67** 098901 (in Chinese) [孔江涛, 黄健, 龚建兴, 李尔玉 2018 物理学报 **67** 098901]
- [6] Wang T T, Lang 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]
- [7] 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, 蒋沅, 童天驰, 严玉为, 淦各升 2021 物理学报 **70** 216401]
- [8] Jiang T S, Ruan Y R, Li H, Bai L, Yuan Y F, Yu T Y 2025 *Acta Phys. Sin.* **74** 126401 (in Chinese) [姜廷帅, 阮逸润, 李海, 白亮, 袁逸飞, 于天元 2025 物理学报 **74** 126401]
- [9] Rezaei A A, Munoz J, Jalili M, Khayyam H 2023 *Expert Syst. Appl.* **214** 119086
- [10] Zhang Y H, Lu Y L, Yang G Z, Hang Z J 2022 *Appl. Sci.* **12** 1944
- [11] Wang B Y, Yang X C, Lu S R, Tang Y P, Hong S Q, Jiang H Y 2024 *Acta Phys. Sin.* **73** 226401 (in Chinese) [王博雅, 杨小春, 卢升荣, 唐勇平, 洪树权, 蒋惠园 2024 物理学报 **73** 226401]
- [12] Lü L Y, Zhou T, Zhang Q M, Stanley H E 2016 *Nat. Commun.* **7** 10168
- [13] Kou J H, Jia P, Liu J Y, Dai J Q, Luo H R 2023 *Neurocomputing* **530** 23
- [14] Qiu Z H, Fan T L, Li M, Lü L Y 2021 *New J. Phys.* **23** 033036
- [15] Chakrabarti S, Dom B, Raghavan P, Rajagopalan S, Gibson D, Kleinberg J 1998 *CNIS* **30** 65
- [16] Fan C J, Zeng L, Sun Y Z, Liu Y Y 2020 *Nat. Mach. Intell.* **2** 317
- [17] Zhao X Y, Huang B, Tang M, Zhang H F, Chen D B 2014 *Europhys. Lett.* **108** 68005
- [18] Bao Z K, Liu J G, Zhang H F 2017 *Phys. Lett. A* **381** 976
- [19] Ji S G, Lü L Y, Yeung C H, Hu Y Q 2017 *New J. Phys.* **19** 073020
- [20] Anderson R M, May R M 1991 *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press)
- [21] Pastor-Satorras R, Castellano C, Mieghem P V, Vespignani A 2015 *Rev. Mod. Phys.* **87** 925
- [22] Feng M L, Zhang S F, Xia C Y, Zhao D W 2024 *Chaos* **34** 073128
- [23] Avraam D, Hadjichrysanthou C 2024 *J. Theor. Biol.* **599** 112010
- [24] Lu J A, Liu H, Chen J 2016 *Synchronization in Complex Dynamical Networks* (Vol. 1) (Beijing: Higher Education Press) p49
- [25] Pirani M, Sundaram S 2016 *IEEE Trans. Autom. Control* **61** 509
- [26] Bapat R B 2010 *Graphs and Matrices* (Springer London)
- [27] Zhu H Y, Klein D J, Lukovits I 1996 *J. Chem. Inf. Model.* **36** 420
- [28] Gutman I, Mohar B 1996 *J. Chem. Inf. Comput. Sci.* **36** 982
- [29] Guan Z, Chen J L 1990 *Numerical Calculation Method* (Beijing: Tsinghua University Press) (in Chinese) [关治, 陈景良 1990 数值计算方法 (北京: 清华大学出版社)]
- [30] Liu Y Q, Lu J A 2007 *Complex Syst. Complexity Sci.* **4** 13 (in Chinese) [刘砚青, 陆君安 2007 复杂系统与复杂性科学 **4** 13]
- [31] Dean J, Ghemawat S 2008 *Commun. ACM* **51** 107
- [32] Yin H, Benson A R, Leskovec J, Gleich D F 2017 *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)* Halifax, NS, Canada, August 13–17, 2017 p555
- [33] Wu Y H, Tian K, Li M D, Hu F 2023 *Comput. Appl. Eng. Educ.* **19** 66 (in Chinese) [吴英晗, 田阔, 李明达, 胡枫 2023 计算机工程与应用 **19** 66]
- [34] Meng L, Xu G Q, Dong C 2025 *Physica A* **657** 130237
- [35] Zhu S Q, Zhan J, Li X 2023 *Sci. Rep.* **13** 16404
- [36] Snap: Stanford network analysis project <https://snap.stanford.edu/>.
- [37] Jiang W C, Wang Y H 2020 *IEEE Access* **8** 32432

An algorithm for mining key node groups in large-scale complex networks based on spectral graph theory*

XING Zihan¹⁾ LIU Siyu¹⁾ LIU Hui^{1)2)3)†} CHEN Lingxiao¹⁾

1) (*School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China*)

2) (*Key Laboratory of Image Processing and Intelligent Control, Huazhong University of Science and Technology, Ministry of Education, Wuhan 430074, China*)

3) (*Key Laboratory of Brain-like Intelligent Systems, Huazhong University of Science and Technology, Wuhan 430074, China*)

(Received 31 March 2025; revised manuscript received 28 May 2025)

Abstract

In this paper, we investigate the saliency identification of node groups in undirected complex networks by utilizing spectral graph theory of pinning control. According to the node significance criterion in network pinning control theory, where important controlled nodes are those maximizing the minimum eigenvalue of the grounded Laplacian matrix after their removal, we propose multi-metric fusion and enhanced greedy search algorithm (MFG), a novel key node group identification framework that integrates multi-metric linear fusion and an enhanced greedy search strategy. First, a linear weighted fusion model that synergistically integrates local centrality metrics with global graph properties is constructed to pre-screen potentially more important node groups, effectively reducing the inherent limitations of a single-metric evaluation paradigm. Second, a dual search strategy combining second-order neighborhood perturbation and global random walk mechanisms is developed to optimize the myopic nature of traditional greedy algorithms. Through iterative selection within pre-screened node groups, the nodes maximizing the minimum eigenvalue of the grounded Laplacian matrix are identified, achieving an optimal balance between local optimization and global search capabilities. Third, computational efficiency is enhanced by using a modified inverse power method for eigenvalue calculation, reducing the complexity of traditional spectral computations. Comprehensive simulations of generated networks and real-world networks demonstrate the framework's superiority. The evaluation of the proposed algorithm includes three aspects: 1) comparison of the minimum eigenvalues between different algorithms; 2) SIR epidemic modeling for propagation capability assessment; 3) topological analysis of identified key nodes. The simulation results reveal the following two significant points: a) Our method outperforms state-of-the-art benchmarks (NPE, AGM, HVGC) in maximizing the ground Laplacian minimum eigenvalue in synthesized (NW small-world, ER) and real-world networks, especially at critical control sizes; b) The identified critical node groups exhibit unique topological features, typically combining high-level hubs with strategically located bridges to best balance local influence and global connectivity. Importantly, the SIR propagation model confirms that these topologically optimized populations accelerate the early outbreak of epidemics and maximize global saturation coverage, directly linking structural features with superior dynamic influence. These findings provide guidance for controlling information propagation in social networks.

Keywords: complex network, node group importance, pre-screening algorithm, spectral graph theory of the grounded Laplacian matrix

PACS: 64.60.aq, 07.05.Dz, 87.55.kd, 89.75.-k

DOI: [10.7498/aps.74.20250416](https://doi.org/10.7498/aps.74.20250416)

CSTR: [32037.14.aps.74.20250416](https://cstr.cn/32037.14.aps.74.20250416)

* Project supported by the National Natural Science Foundation of China (Grant Nos. 62176099, U24A20272) and the Interdisciplinary Research Program of Huazhong University of Science and Technology, China (Grant No. 5003170102).

† Corresponding author. E-mail: hliu@hust.edu.cn



基于谱图理论的大规模复杂网络重要节点组挖掘算法

邢梓涵 刘丝语 刘慧 陈凌霄

An algorithm for mining key node groups in large-scale complex networks based on spectral graph theory

XING Zihan LIU Siyu LIU Hui CHEN Lingxiao

引用信息 Citation: *Acta Physica Sinica*, 74, 166401 (2025) DOI: 10.7498/aps.74.20250416

CSTR: 32037.14.aps.74.20250416

在线阅读 View online: <https://doi.org/10.7498/aps.74.20250416>

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

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>

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>

基于信息熵与迭代因子的复杂网络节点重要性评价方法

Importance evaluation method of complex network nodes based on information entropy and iteration factor

物理学报. 2023, 72(4): 048901 <https://doi.org/10.7498/aps.72.20221878>

基于加权圈比的复杂网络关键节点识别方法

A method of identifying key nodes in complex networks based on weighted cycle ratio

物理学报. 2025, 74(14): 148902 <https://doi.org/10.7498/aps.74.20250338>

基于图卷积神经网络的多维度节点重要性评估方法

A multidimensional node importance evaluation method based on graph convolutional networks

物理学报. 2024, 73(22): 226401 <https://doi.org/10.7498/aps.73.20240937>

基于Tsallis熵的复杂网络节点重要性评估方法

A method of evaluating importance of nodes in complex network based on Tsallis entropy

物理学报. 2021, 70(21): 216401 <https://doi.org/10.7498/aps.70.20210979>