



计算机科学

COMPUTER SCIENCE

基于NMTF的自适应复杂网络社团检测算法

李希龙, 刘琰, 贾萌萌, 张子林

引用本文

李希龙, 刘琰, 贾萌萌, 张子林. [基于NMTF的自适应复杂网络社团检测算法](#)[J]. 计算机科学, 2026, 53(4): 215-223.

LI Xilong, LIU Yan, JIA Mengmeng, ZHANG Zilin. [NMTF-based Adaptive Algorithm for Community Detection in Complex Networks](#) [J]. Computer Science, 2026, 53(4): 215-223.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[复杂网络下的概念认知学习与增量学习](#)

Concept-cognitive Learning and Incremental Learning in Complex Networks
计算机科学, 2026, 53(4): 208-214. <https://doi.org/10.11896/jsjcx.250600216>

[基于梯度引导的社团隐匿扰动子结构优化方法](#)

Gradient-guided Perturbed Substructure Optimization for Community Hiding
计算机科学, 2025, 52(9): 376-387. <https://doi.org/10.11896/jsjcx.240800107>

[基于可控偏好抽样的复杂网络度分布推断方法](#)

Degree Distribution Inference Method for Complex Networks Based on Controllable Preferential Sampling
计算机科学, 2025, 52(7): 82-91. <https://doi.org/10.11896/jsjcx.241200098>

[基于网络分解的故障树自动生成方法研究](#)

Research on Automatic Generation Method of Fault Tree Based on Network Decomposition
计算机科学, 2025, 52(6A): 240900108-6. <https://doi.org/10.11896/jsjcx.240900108>

[基于类脑脉冲神经网络的边缘联邦持续学习方法](#)

Edge-side Federated Continuous Learning Method Based on Brain-like Spiking Neural Networks
计算机科学, 2025, 52(3): 326-337. <https://doi.org/10.11896/jsjcx.240900070>

基于 NMTF 的自适应复杂网络社团检测算法

李希龙 刘 琰 贾萌萌 张子林

信息工程大学网络空间安全学院 郑州 450001

(lxllxlong@163.com)

摘要 针对现有基于非负矩阵分解(NMF)的社团检测方法存在需预设社团数量、易陷入局部最优及模型泛化能力有限等问题,提出了一种基于非负矩阵三因子分解(NMTF)的自适应社团检测算法(Adp-NMTF)。该算法引入动态评估和反馈机制,能够自动搜索并确定合理的社团划分数,无需人工预设;引入图正则化、稀疏性及社团独立性约束,平衡模型泛化能力与可解释性;同时利用半监督初始化与热启动策略,加速 NMTF 收敛过程,提高计算效率。实验结果表明,Adp-NMTF 能够自适应确定网络社团数,在合成和真实网络数据集测试中,模块度(Q)、标准化互信息(NMI)和调整兰德指数(ARI)等评价指标均优于主流基线方法,同时显著提升了矩阵分解的收敛速度和计算效率。

关键词: 复杂网络; 社团检测; 非负矩阵分解; 非负矩阵三因子分解; 正则约束

中图分类号 TP391.9

NMTF-based Adaptive Algorithm for Community Detection in Complex Networks

LI Xilong, LIU Yan, JIA Mengmeng and ZHANG Zilin

School of Cybersecurity, Information Engineering University, Zhengzhou 450001, China

Abstract To address the limitations of existing non-negative matrix factorization(NMF)-based community detection methods, such as the requirement for preset the number of communities, susceptibility to local optima, and limited model generalization, this paper proposes Adp-NMTF, an adaptive community detection algorithm based on non-negative matrix tri-factorization(NMTF). The algorithm incorporates a dynamic evaluation and feedback mechanism to automatically search and determine the optimal number of communities without manual intervention. It introduces graph regularization, sparsity constraints, and inter-community independence constraints to balance generalization capability and interpretability. Additionally, semi-supervised initialization and warm-start strategies are employed to accelerate NMTF convergence and improve computational efficiency. Experimental results demonstrate that Adp-NMTF can autonomously determine a reasonable number of communities and outperforms mainstream baseline methods in both synthetic and real-world networks across evaluation metrics including modularity(Q), normalized mutual information(NMI), and adjusted Rand index(ARI). Furthermore, the convergence rate of matrix factorization is significantly improved.

Keywords Complex networks, Community detection, Non-negative matrix factorization, Non-negative matrix tri-factorization, Regularization constraints

1 引言

现实世界的复杂系统及事物间关系可以建模为复杂网络^[1](如社交、交通和生物网络等),其中节点表示实体,边表示实体间关系或交互行为。复杂网络普遍具有社团结构特征,即网络可划分为多个社团,社团内部连接紧密,社团之间连接稀疏^[2]。社团的具体含义因网络所表示的实际系统而异,例如,其在社交网络中代表兴趣或偏好相似的团体,在蛋白质网络中则代表不同的功能单元。

社团检测有助于揭示复杂系统的功能模块划分和信息

传播规律,在社交推荐、生物功能发现和城市规划等领域具有重要应用价值^[1-3]。现有社团检测方法主要包括模块度优化、谱图理论、随机游走和机器学习及深度学习等^[4],但各有局限:模块度优化方法易受分辨率极限问题影响,难以识别小规模社团^[5];谱图理论方法计算复杂度高,难以处理大规模网络^[6];随机游走方法对初始条件敏感,结果稳定性较差^[7];深度学习需要大量标注数据且可解释性差^[8]。近年来,非负矩阵分解(Non-negative Matrix Factorization, NMF)因兼具良好的拓扑保持性、可扩展性和可解释性,成为社团检测的研究热点。

到稿日期:2025-05-15 返修日期:2025-09-03

基金项目:国家重点研发计划(2022YFB3102904)

This work was supported by the National Key Research and Development Program of China(2022YFB3102904).

通信作者:刘琰(ms.liuyan@foxmail.com)

非负矩阵三因子分解(Non-negative Matrix Tri-Factorization, NMTF)作为 NMF 的扩展模型^[9],在社团检测中备受关注。NMTF 可以将网络特征矩阵分解为节点-社团隶属度矩阵、社团关联矩阵和社团特征矩阵,能够同时考虑节点与社团、社团与社团之间的关系,更充分地揭示网络社团结构,且分解矩阵元素物理意义明确,便于分析社团间关系及其演化规律^[10-12]。

尽管基于 NMTF 的社团检测已取得一定成果,但仍存在几个关键问题:一是社团划分数 k 需预先设定,而实际网络中 k 通常未知,限制了其实际应用^[13];二是传统目标函数缺乏对网络局部结构、节点隶属稀疏性和社团间层次关系的综合约束,易过拟合或陷入局部最优;三是现有算法计算复杂度高,收敛速度慢。

针对上述问题,本文提出了一种基于 NMTF 的自适应社团检测算法(Adaptive Non-negative Matrix Tri-Factorization, Adp-NMTF),通过自适应确定社团划分数、引入正则化约束等,求解最优 NMTF 分解,克服传统 NMTF 依赖预设 k 值的缺陷,提升模型泛化能力。本文的主要贡献如下:

1)提出一种动态反馈调整机制,基于目标评估与历史

信息,自适应确定社团划分数 k ,探索最优分解,避免人工预设的局限;

2)扩展 NMTF 模型,增加了多约束联合优化,平衡了拓扑质量、泛化能力与可解释性;

3)引入半监督初始化和热启动机制,提高了 NMTF 的收敛速度,降低了陷入局部最优的风险;

4)通过合成网络和真实网络测试,与主流基线方法对比,验证了 Adp-NMTF 的有效性和检测准确性。

本文第 2 章介绍基本概念与符号定义;第 3 章综述基于 NMF/NMTF 的社团检测相关研究;第 4 章阐述 Adp-NMTF 算法设计及关键技术;第 5 章讨论扩展 NMTF 模型及其求解方法;第 6 章分析算法复杂度;第 7 章通过合成与真实数据集验证算法有效性;最后总结全文并展望未来研究方向。

2 基本概念和符号定义

复杂网络(Complex Network)通常表示为图结构 $G=(V, E)$,其中 $V=v_1, v_2, \dots, v_n$ 为节点集合, $n=|V|$ 表示节点数, $E=\{e_{ij} | v_i$ 与 v_j 存在连接} 为边集合, $m=|E|$ 表示边数。复杂网络的相关概念及定义如表 1 所列。

表 1 复杂网络相关概念和表示

Table 1 Related concepts and representations of complex networks

| 符号 | 数学定义/描述 | 维度 | 物理意义 |
|-----|--|-----------------------------|--|
| G | $G=(V, E)$ | — | 复杂网络,包含节点集 V 、边集 E |
| A | $A=[a_{ij}], a_{ij} \in \{0, 1\}$ | $\mathbb{R}^{n \times n}$ | 邻接矩阵, $a_{ij}=1$ 表示节点 i 与 j 相连 |
| D | $D=\text{diag}(d_1, d_2, \dots, d_n), d_i=\sum_j a_{ij}$ | $\mathbb{R}^{n \times n}$ | 度矩阵, 对角元素 d_i 为节点 i 的度数 |
| L | $L=D-A$ | $\mathbb{R}^{n \times n}$ | 拉普拉斯矩阵,刻画网络全局结构特性 |
| U | $U \in \mathbb{R}_+^{n \times k}$ | $\mathbb{R}_+^{n \times k}$ | 节点-社团隶属度矩阵, U_{ij} 表示节点 i 对社团 j 的隶属度 |
| S | $S \in \mathbb{R}_+^{k \times k}$ | $\mathbb{R}_+^{k \times k}$ | 社团关联矩阵, S_{pq} 表示社团 p 与 q 的连接强度 |

社团(Community)尚无统一严格定义,普遍指网络中一组节点的集合,其内部节点连接紧密(边数显著多于随机情况),外部连接稀疏(社团间边数或权重较少)^[1]。社团结构体现了网络的层次性与模块化。

复杂网络通常可划分为社团集合 $C=C_1, C_2, \dots, C_k$,其中 k 为社团数, $C_i \subseteq V$ 且 $\bigcup_{i=1}^k C_i = V$ 。若 $C_i \cap C_j = \emptyset (\forall i \neq j)$,则为非重叠社团划分;若 $C_i \cap C_j \neq \emptyset (\forall i \neq j)$,则为重叠社团划分。

对于矩阵 X, X_i 为第 i 行向量, X_j 为第 j 列向量, X_{ij} 为第 i 行第 j 列元素, I 表示单位矩阵;矩阵 X 的迹、转置分别表示为 $\text{tr}(X)$ 和 X^T , 矩阵的 Frobenius 范数为 $\|X\|_F =$

$$\sqrt{\sum_{i=1}^m \sum_{j=1}^n X_{ij}^2}, X \in \mathbb{R}^{m \times n}.$$

3 相关工作

非负矩阵分解(NMF)作为契合社团检测需求的无监督学习技术,通过低秩非负分解,将网络结构映射到潜在社团空间,其分解结果可直观解释为节点对社团的隶属度及社团特征^[14],且理论上等价于 K-means、谱聚类等方法^[15-16]。

3.1 基于 NMF 的社团检测

在社团检测中, NMF 将网络邻接矩阵 $A \in \mathbb{R}^{n \times n}$ 分解为非负矩阵 $U \in \mathbb{R}_+^{n \times k}$ (节点-社团隶属度矩阵)和 $V \in \mathbb{R}_+^{k \times n}$ (社团特征矩阵),满足 $A \approx UV$ 。Lee 等^[17]最早提出基于欧氏距离或

KL 散度优化的 NMF 算法,后续研究不断拓展;Kuang 等^[18]提出对称非负矩阵分解(SNMF),简化无向网络检测;Sheldon 等^[19]将其扩展为概率模型,以支持重叠社团检测;Li 等^[20]针对无向网络,提出对称与非负约束模型(SNCMF),通过多 LF 矩阵增强表示学习能力;Guan 等^[21]提出多跳非负矩阵分解(MHNMF),融合多跳连接信息提升性能;Ye 等^[22]基于图正则化 NMF 提出 HPNMF 模型,整合网络拓扑与节点同质性信息。

3.2 基于 NMTF 的社团检测

非负矩阵三因子分解(NMTF)作为 NMF 的扩展,将矩阵分解为 $A \approx USV^T$,其中 U, S, V 分别表示节点-社团隶属度、社团间关联强度及社团特征。Ding 等^[9]率先将其应用于社团检测,通过 S 矩阵建模社团交互;Song 等^[23]提出对称三因子分解模型(SNLF),解决了传统双因子模型在无向稀疏网络中的对称性、非负性及预测精度问题;Li 等^[24]提出半正交非负矩阵三因子分解(semi-ONMTF),利用 Cayley 变换确保正交性,以提高效率与准确性;Zhang 等^[25]基于 NMTF 提出多路社区检测方法,通过共识矩阵整合多层特征,提升跨平台检测质量。

3.3 问题与挑战

基于 NMF 和 NMTF 的社团检测虽在模型拓展、算法优化等方面取得了成果,但仍面临挑战:1)需预先设定社团数量 k ,现有枚举(如通过模块度选优^[26])和交叉验证等方法缺乏

与分解过程的协同优化,导致效率与精度受限;2)模型在初始化敏感及结果稳定性等方面仍需深入研究。

4 基于 NMTF 的自适应社团检测

基于 NMTF 的自适应社团检测算法 Adp-NMTF 通过

Adp-NMTF 总体结构如图 1 所示。

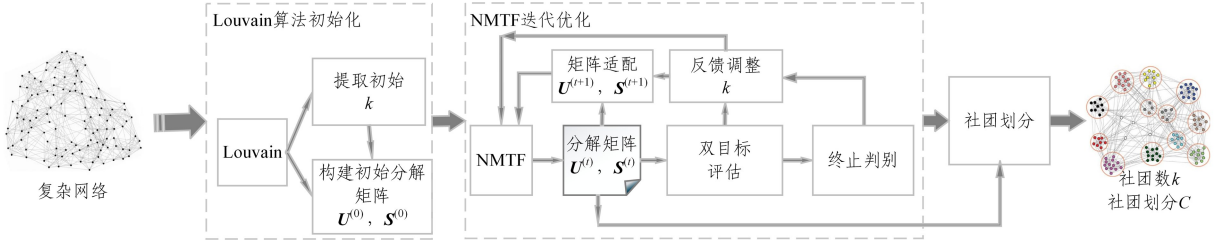


图1 Adp-NMTF 算法总体结构框架

Fig. 1 Structural framework of Adp-NMTF algorithm

4.1 基于 Louvain 的初始化

Louvain 算法^[27]是一种基于模块度优化的无监督社团检测算法,其实现简单、计算效率高($O(n \log n)$),但对初始条件敏感,易受初始随机因素影响,导致社团划分结果不稳定。本文利用该算法对复杂网络进行社团预划分,以初始化社团数 k 及 NMTF 分解矩阵 U 和 S 。

4.1.1 初始化社团数 k

为降低随机性的影响,初始化采用多次采样策略:通过 Louvain 算法运行 3 次生成候选划分集合 $\{C_i\}_{i=1}^3$,选择模块度最高的划分 $C^{(0)} = \arg \max Q_{\text{mod}}(C_i)$ 作为初始社团划分,其中 Q_{mod} 为模块度函数,并设置初始社团数为 $k^{(1)} = \left\lceil \frac{1}{3} \sum_{i=1}^3 |C_i| \right\rceil$ 。

4.1.2 初始化矩阵适配

NMTF 模型的目标函数求解过程与分解矩阵初始化密切相关。传统随机初始化策略存在收敛慢、结果不稳定的问题,Adp-NMTF 以 Louvain 算法的检测结果为指导进行初始化。

由于 Louvain 算法输出为硬划分(每个节点仅属于一个社团),而 NMTF 的隶属度矩阵 $U \in \mathbb{R}^{n \times k}$ 为软划分(非负连续值),因此需对 Louvain 划分结果进行软化和模糊化处理,避免初始矩阵过于稀疏导致优化困难。具体步骤如下。

1) 基于 Louvain 初始社团划分,采用 one-hot 编码初始化硬隶属度矩阵 U_0 :若节点 i 属于社团 j ,则 $U_{ij} = 1$,否则为 0。

2) 对 U_0 进行模糊化处理:保留节点原分配社团 80% 的隶属度,引入 20% 的均匀随机噪声,即 $\tilde{U}_{ij} = 0.8 \cdot U_{ij} + 0.2 \cdot U(0,1)$,其中 $U(0,1)$ 是区间 $(0,1)$ 上的均匀分布随机数。

3) 对每行进行归一化处理:确保每行和为 1,满足概率分布约束 $U_{ij}^{(1)} = \frac{\tilde{U}_{ij}}{\sum_{m=1}^k \tilde{U}_{im}}$,得到初始化隶属度矩阵 $U^{(1)} \in \mathbb{R}^{n \times k^{(1)}}$ 。

4.2 自适应迭代

自适应迭代是 Adp-NMTF 框架的关键,它通过在每次迭代中交替执行 NMTF 与动态调整,以寻找最优的 NMTF 划分。其关键技术包括 NMTF 性能评估、社团数调整和分解矩阵适配等。

3 步实现:1)用 Louvain 算法生成初始社团划分并构建分解矩阵;2)迭代执行 NMTF 分解,借助双目标评估和动态反馈调整机制自适应探索最优社团划分数 k ;3)基于 k 值下的分解结果输出最终检测结果。

4.2.1 NMTF 性能评估

每次迭代先固定当前社团数 $k^{(t)}$,对网络特征矩阵 \tilde{A} 执行 NMTF 分解,得到隶属度矩阵 $U^{(t)}$ 和社团关联矩阵 $S^{(t)}$ 。NMTF 性能评估用于测评本次划分性能,为下次迭代提供优化引导。

本文结合模块度 Q 和社团规模分布熵 H ,构建双目标混合评估函数 Q' 来评估划分质量:

$$Q' = Q + \lambda \cdot (H / \ln k) \quad (1)$$

其中, $H / \ln k$ 表示对熵进行归一化使其值域为 $[0,1]$; λ 为权重系数 ($0 \leq \lambda \leq 1$)。通过调整 λ ,可灵活平衡模块度与熵的相对重要性,实验中设 $\lambda = 0.5$ 。模块度 Q 关注社团内部连接紧密性,熵 H 反映全局分布,二者互补。引入熵调节项来保留合理小规模社团,避免单一模块度优化的过拟合问题,增强目标函数平滑性,以稳定优化过程。

4.2.2 自适应 k 值调整

本文提出动态反馈搜索算法实现社团数量 k 的自适应优化。在获得初始社团数量 $k^{(1)}$ 后,基于 NMTF 结果评估和历史信息调整 $k^{(t+1)}$,最终得到优化的社团数 k^* 。

1) 初始设置:利用 Louvain 算法生成初始划分,其社团数作为 Adp-NMTF 的初始值 $k^{(1)}$ 。根据初始划分统计特征设定搜索方向:若初始划分社团规模差异大(熵 $H^{(0)}$ 较低),则设方向标志 $d^{(0)} = 1$ (表示增加 k 以提升社团的均匀性);若初始模块度 $Q^{(0)}$ 低,则设 $d^{(0)} = -1$ (表示减少 k 以提升模块度)。

2) 调整方向和步长:基于相邻迭代的双目标函数变化 $\Delta Q'^{(t)} = Q'^{(t)} - Q'^{(t-1)}$,按以下规则调整。

$$k^{(t+1)} = k^{(t)} + d^{(t+1)} \cdot \Delta k^{(t+1)} \quad (2)$$

(1) 方向:若 $\Delta Q'^{(t)} > 0$,则保持当前方向 $d^{(t+1)} = d^{(t)}$;否则反转方向 $d^{(t+1)} = -d^{(t)}$ 。

(2) 步长:若 $|\Delta Q'^{(t)}| < \epsilon$ ($\epsilon = 0.01$),则步长 $step^{(t+1)} = 1$;否则步长为 2。

(3) 范围约束:社团数 k 需满足 $k \in [2, \lfloor \sqrt{n} \rfloor]$,若超出范围,则令 $k^{(t+1)} = \text{RandInt}(2, \lfloor \sqrt{n} \rfloor)$ 。

4.2.3 矩阵维度适配

NMTF 分解要求 $A \approx USU^T$,当划分数为 k 时,需满足 $U \in \mathbb{R}_+^{n \times k}$, $S \in \mathbb{R}_+^{k \times k}$,否则矩阵乘法失效。因此需根据 k 值变

化动态调整 U 和 S 的初始维度(即维度适配),确保模型一致性。当 k 增减时,分别通过分裂或合并矩阵的行/列实现维度适配,保证社团结构变化与矩阵表示同步更新。

1) 社团分裂($k \rightarrow k+1$)

(1) 隶属度矩阵 U 适配。选择隶属度熵最大的社团 $j^* = \arg \max_j (-\sum_i U_{ij} \log U_{ij})$ 作为待分裂社团(熵高说明节点归属模糊,可能存在子结构)。对其隶属度向量 U_{\cdot, j^*} 进行中位数分割,生成两列新向量并添加高斯噪声 $\epsilon \sim \mathcal{N}(0, 0.1)$ 。具体步骤如下。

① 计算中位数:

$$m_{j^*} = \text{median}(U_{1j^*}, U_{2j^*}, \dots, U_{nj^*})$$

② 生成二值掩码向量:

$$M = \begin{cases} 1, & U_{ij^*} \geq m_{j^*} \\ 0, & \text{otherwise} \end{cases}$$

③ 分裂为两列:

$$U_{\cdot, \text{new1}} = M \odot U_{\cdot, j^*} + \epsilon_1 \quad (3)$$

$$U_{\cdot, \text{new2}} = (1 - M) \odot U_{\cdot, j^*} + \epsilon_2 \quad (4)$$

其中, \odot 表示逐元素乘法, $\mathbf{1}$ 为全 1 向量。

④ 删除原列 U_{\cdot, j^*} , 插入新列形成 $n \times (k+1)$ 矩阵 U' 。

(2) 社团关联矩阵 S 适配。将 S 扩展为 $(k+1) \times (k+1)$ 矩阵, 新行列初始化如下。

$$S_{\text{new}, l} = \frac{S_{j^*, l}}{2} + \eta \quad (5)$$

$$S_{l, \text{new}} = \frac{S_{l, j^*}}{2} + \eta, l = 1, \dots, k+1 \quad (6)$$

其中, $\eta \sim \mathcal{U}(0, 0.05)$ 为微小随机扰动, 避免初始值对称导致优化停滞。

当 $\Delta k > 1$ 时, 重复上述分裂过程。

2) 社团合并($k \rightarrow k-1$)

当减少社团数时, 通过合并相似社团的隶属度与交互关系实现矩阵适配, 步骤如下。

(1) 选择合并对象。计算社团相似度矩阵 $\text{Sim} = U^T D U \in \mathbb{R}^{k \times k}$ (采用余弦相似度), 选取最相似非对角元素对应的社团对 $(j_1, j_2) = \arg \max_{p \neq q} \text{Sim}_{pq}$ 。

(2) 隶属度矩阵 U 合并。将社团 j_1 和 j_2 对应的列按均值合并为新列:

$$U_{\cdot, \text{merged}} = \frac{U_{\cdot, j_1} + U_{\cdot, j_2}}{2} \quad (7)$$

移除原列并插入新列得到 $U' \in \mathbb{R}^{n \times (k-1)}$ 。

(3) 社团关联矩阵 S 合并。

① 行/列合并: 合并后的社团交互强度取原社团的均值。

$$S_{\text{merged}, l} = \frac{S_{j_1, l} + S_{j_2, l}}{2} \quad (8)$$

$$S_{l, \text{merged}} = \frac{S_{l, j_1} + S_{l, j_2}}{2} \quad (9)$$

② 自交互: 合并后社团内部交互强度为:

$$S_{\text{merged}, \text{merged}} = \frac{S_{j_1, j_1} + 2S_{j_1, j_2} + S_{j_2, j_2}}{4} \quad (10)$$

③ 维度缩减: 删除原 j_1 和 j_2 行/列, 插入合并后的行/列, 得到新矩阵 $S' \in \mathbb{R}^{(k-1) \times (k-1)}$ 。

当 $\Delta k > 1$ 时, 重复上述合并过程。

4.2.4 迭代终止判定

迭代终止采用多条件联合判定策略, 满足以下条件之一即终止。

1) 目标函数稳定性。目标函数稳定性通过双目标函数 Q' 的滑动窗口标准差来评估:

$$\sigma_Q = \sqrt{\frac{1}{w} \sum_{i=t-w+1}^t (Q'^{(i)} - \bar{Q}')^2} \quad (11)$$

取滑动窗口长度 $w=5$, 若 $\sigma_Q < \epsilon$ (阈值 $\epsilon=0.001$), 则迭代结束。

2) 划分一致性。通过连续划分的标准化互信息来评估, 若 $NMI(C^{(t)}, C^{(t-1)}) > 0.95$ (超过 95% 节点归属未显著变化), 则迭代结束。

3) 辅助判别条件。设置最大迭代次数 T_{\max} , 当 $t \geq T_{\max}$ 时, 强制终止。

4.3 社团检测结果生成

Adp-NMTF 通过处理隶属度矩阵 U 生成检测结果, 将软隶属度映射为社团划分。本文采用基于节点隶属度排序的 Top-K 方法: 为每个节点保留隶属度最高的前 topK 个社团, 通过调整参数 topK, 可以支持非重叠 (topK=1) 或重叠 (topK ≥ 2) 社团检测。具体实现过程如算法 1 所示。

算法 1 社团检测结果生成

输入: 社团隶属度矩阵 $U \in \mathbb{R}^{n \times k^*}$, k^* 是最终社团数; 参数 topK, 表示每个节点选择前 topK 个社团

输出: 社团划分 C , $C[i]$ 是节点 i 所属 topK 个社团列表

算法描述

1. 初始化空字典 C : $C \leftarrow \{\}$

2. 对于每个节点 $i = 1 \rightarrow n$:

提取节点 i 的隶属度向量:

$$\mathbf{u}_i = [U_{i1}, U_{i2}, \dots, U_{ik^*}]$$

对 \mathbf{u}_i 降序排序, 得到排序后的索引列表 sorted_indices

取前 topK 个索引:

$$\text{selected} \leftarrow \text{sorted_indices}[0: \text{topK}]$$

存入结果: $C[i] \leftarrow \text{selected}$

3. 返回 C

算法对每个节点 i 的隶属度向量 $U_{i\cdot}$ 按降序排序, 得到索引序列 j_1, j_2, \dots, j_k^* , 满足 $U_{ij_1} \geq U_{ij_2} \geq \dots \geq U_{ij_k^*}$, 保留前 topK 个社团作为节点 i 的社团归属集合 $C_i = j_1, j_2, \dots, j_{\text{topK}}$ 。

5 扩展 NMTF 模型

为提升社团检测的准确性与模型鲁棒性, Adp-NMTF 对算法核心 NMTF 模型进行扩展, 引入多正则项综合约束网络局部结构、节点隶属稀疏性及社团间层次关系, 增强模型表达能力, 实现更精准的复杂网络社团检测。

5.1 多正则化 NMTF 模型设计

基于无向网络 $G = (V, E)$ 的对称性, 本文定义 NMTF 模型为 $\tilde{A} \approx USU^T$, $U \in \mathbb{R}_+^{n \times k}$, $S \in \mathbb{R}^{k \times k}$ 。这里的 \tilde{A} 为网络特征矩阵, 其基本形式为邻接矩阵 A , 可以根据需要纳入更多网络结构信息; U 为节点-社团隶属度矩阵 (U_{ic} 表示节点 i 属于社团 c 的概率); S 为社团关联矩阵 (S_{pq} 表示社团 p 与 q 的关联强度)。

为提高社团检测的准确性,扩展了 NMTF 模型的目标函数(损失函数),定义如下:

$$\mathcal{L} = \underbrace{\frac{1}{2} \|\tilde{\mathbf{A}} - \mathbf{USU}^T\|_F^2}_{\text{重构误差}} + \underbrace{\alpha \cdot \text{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U})}_{\text{图正则项}} + \underbrace{\beta \cdot \|\mathbf{U}\|_1}_{\text{稀疏项}} + \underbrace{\gamma \cdot \|\mathbf{S} \odot (\mathbf{I} - \mathbf{D})\|_F^2}_{\text{社团间连接项}} \quad (12)$$

其中,第一项 $\|\tilde{\mathbf{A}} - \mathbf{USU}^T\|_F^2$ 用于衡量重构误差,反映分解结果对原始网络结构的拟合程度;第二项 $\alpha \cdot \text{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U})$ 是图拉普拉斯正则项,该项强制相邻节点具有相似的社团隶属度分布,保持网络局部结构的平滑性;第三项 $\beta \cdot \|\mathbf{U}\|_1$ 是稀疏性约束项,可促使每个节点仅显著隶属于少数社团,从而提高划分的清晰度;第四项 $\gamma \cdot \|\mathbf{S} \odot (\mathbf{I} - \mathbf{D})\|_F^2$ 是社团间连接约束项,通过惩罚关联矩阵 \mathbf{S} 的非对角元素(社团间连接),间接鼓励社团内部连接。正则化参数 α, β, γ 根据网络特性设置。

5.2 目标函数求解

扩展 NMTF 模型的优化问题为:

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{S}} \mathcal{L} = & \min_{\mathbf{U}, \mathbf{S}} \left(\frac{1}{2} \|\tilde{\mathbf{A}} - \mathbf{USU}^T\|_F^2 + \alpha \cdot \text{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) + \right. \\ & \left. \beta \cdot \|\mathbf{U}\|_1 + \gamma \cdot \|\mathbf{S} \odot (\mathbf{I} - \mathbf{D})\|_F^2 \right) \\ \text{s. t. } & \mathbf{U} \geq 0, \mathbf{S} \geq 0, \mathbf{S} = \mathbf{S}^T \end{aligned} \quad (13)$$

采用乘法更新规则(Multiplicative Update Rule, MUR)进行交替优化。首先计算目标函数 \mathcal{L} 关于 \mathbf{U} 和 \mathbf{S} 的梯度,然后合成总梯度。

$$1) \text{ 重构误差项 } \mathcal{L}_{\text{rec}} = \frac{1}{2} \|\tilde{\mathbf{A}} - \mathbf{USU}^T\|_F^2$$

对 \mathbf{U} 的下降梯度: $\frac{\partial \mathcal{L}_{\text{rec}}}{\partial \mathbf{U}} = -\tilde{\mathbf{A}}\mathbf{US} + \mathbf{USU}^T\mathbf{US}$; 对 \mathbf{S} 的下降

$$\text{梯度: } \frac{\partial \mathcal{L}_{\text{rec}}}{\partial \mathbf{S}} = \mathbf{U}^T \mathbf{USU}^T \mathbf{U} - \mathbf{U}^T \tilde{\mathbf{A}} \mathbf{U}.$$

$$2) \text{ 图拉普拉斯正则项 } \mathcal{L}_{\text{lap}} = \alpha \cdot \text{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U})$$

因为 \mathbf{L} 对称,所以对 \mathbf{U} 的梯度: $\frac{\partial}{\partial \mathbf{U}} \alpha \cdot \text{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) = \alpha(\mathbf{L} +$

$\mathbf{L}^T)\mathbf{U} = 2\alpha\mathbf{L}\mathbf{U}$; 而没有对 \mathbf{S} 的梯度。

$$3) \text{ 稀疏正则项 } \mathcal{L}_{\text{sas}} = \beta \cdot \|\mathbf{U}\|_1$$

对 \mathbf{U} 的梯度: $\frac{\partial \mathcal{L}_{\text{sas}}}{\partial \mathbf{U}} = \beta \cdot \text{sign}(\mathbf{U})$, 这里 \mathbf{U} 为非负,

$\text{sign}(\mathbf{U}) = 1$; 没有对 \mathbf{S} 的梯度。

$$4) \text{ 社团交互正则项 } \mathcal{L}_{\text{com}} = \gamma \cdot \|\mathbf{U} \odot (\mathbf{I} - \mathbf{D})\|_F^2$$

社团交互正则项没有对 \mathbf{U} 的梯度; 对 \mathbf{S} 的梯度: $\frac{\partial \mathcal{L}_{\text{com}}}{\partial \mathbf{S}} =$

$2\gamma\mathbf{S} \odot (\mathbf{I} - \mathbf{D})$ 。

5) 合成总梯度

对 \mathbf{U} 的总梯度: $\frac{\partial \mathcal{L}}{\partial \mathbf{U}} = \tilde{\mathbf{A}}\mathbf{US} - \mathbf{USU}^T\mathbf{US} + 2\alpha\mathbf{L}\mathbf{U} + \beta \cdot \mathbf{1}$; 对 \mathbf{S}

的总梯度: $\frac{\partial \mathcal{L}}{\partial \mathbf{S}} = \mathbf{U}^T \tilde{\mathbf{A}} \mathbf{U} - \mathbf{U}^T \mathbf{USU}^T \mathbf{U} + 2\gamma \cdot (\mathbf{S} \odot (\mathbf{I} - \mathbf{D}))$ 。

由此得到 \mathbf{U} 和 \mathbf{S} 的乘法更新规则:

$$\mathbf{U} \leftarrow \mathbf{U} \odot \frac{[\tilde{\mathbf{A}}\mathbf{US}]}{[\mathbf{USU}^T\mathbf{US} + 2\alpha\mathbf{L}\mathbf{U} + \beta]} \quad (14)$$

$$\mathbf{S} \leftarrow \mathbf{S} \odot \frac{[\mathbf{U}^T \tilde{\mathbf{A}} \mathbf{U}]}{[\mathbf{U}^T \mathbf{USU}^T \mathbf{U} + 2\gamma(\mathbf{S} \odot (\mathbf{I} - \mathbf{D}))]} \quad (15)$$

其中, α, β, γ 为超参数。实验取值: $\alpha \in [0.1, 1]$, 控制平滑性强; $\beta \in [0.1, 0.5]$, 控制稀疏性强; $\gamma \in [0.1, 0.5]$, 抑制社团间交互强度。

6 复杂度分析

Adp-NMTF 算法含初始化、自适应迭代、结果生成 3 个步骤,各阶段复杂度如下。

初始化阶段: Louvain 算法的复杂度为 $O(n \log n)$, 硬划分转软隶属度矩阵复杂度为 $O(nk)$, 总复杂度为 $O(n \log n + nk)$, 因 $k \ll n$, 故主导项为 $O(n \log n)$ 。

自适应迭代阶段: NMTF 分解单次迭代矩阵乘法复杂度为 $O(nk^2)$; 目标评估中, 模块度 Q 的计算复杂度为 $O(m+n)$, 平均轮廓系数的计算复杂度为 $O(n^2)$, 故单次评估的复杂度为 $O(m+n+n^2)$; k 值动态调整的单次复杂度为 $O(nk+k^2)$; 迭代最多执行 T_{max} 次。所以总复杂度为 $O(T_{\text{max}} \cdot (n^2 + nk^2 + m))$ (忽略低阶项, 保留了主导项)。

结果生成阶段: 采用 Top-K 算法进行社团划分, 对每个节点的隶属度排序, 复杂度为 $O(nk \log k)$ 。当 $k \approx \sqrt{n}$ 时, 该步骤复杂度为 $O(n^{3/2} \log n)$ 。

综上, Adp-NMTF 算法的总时间复杂度近似为 $O(T_{\text{max}} \cdot n^2)$, 与同类算法的复杂度相当^[22-24]。

7 实验及结果分析

本章对 Adp-NMTF 社团检测的性能进行实验测试与对比分析。实验环境为一台配备 Intel Xeon E5-2680v4 处理器(2.4GHz, 14 核)、128 GB 内存的服务器, 算法基于 Python 3.8 实现, 依赖库包括 NumPy, SciPy, NetworkX 等。

7.1 数据集

实验采用合成网络数据集与真实网络数据集对 Adp-NMTF 及对比基线方法进行测试。

7.1.1 合成网络数据集

合成网络数据集通过 LFR 基准生成器构建, 生成两组共 10 个具有不同规模和混合参数的合成网络, 参数配置如表 2 所列。第一组包含 5 个网络, 节点数固定为 $n=3000$, 通过调整混合参数 μ 控制节点与社团外节点的连接比例。混合参数 μ 在区间 $[0.2, 0.4]$ 内以步长 0.05 变化, μ 值越大, 网络社团结构越模糊, 检测难度越高。第二组为不同规模的网络, 节点数 n 从 1000 开始, 以 1000 为步长递增至 5000, 节点度分布和社团大小范围相应调整, 且 $\mu=0.25$ 。

表 2 生成人工网络的 LFR 参数设置

Table 2 Parameter settings of LFR benchmark for generating artificial networks

| LFR 参数 | DataGroup1 | DataGroup2 |
|-------------------------|------------|------------|
| 节点数 n | 3000 | 1000~5000 |
| 平均度数 d | 4 | 4 |
| 最大度数 d_{max} | 15 | 15 |
| 最小社团大小 C_{min} | 50 | $n/50$ |
| 最大社团大小 C_{max} | 100 | $n/20$ |
| 幂律指数 (β) | 2 | 2 |
| 混合参数 μ | 0.2~0.4 | 0.25 |

7.1.2 真实网络数据集

实验选用8个广泛使用的真实网络数据集,详细信息如表3所列。其中6个数据集带有真实社团标签,2个无真实标签。

表3 真实网络数据集信息

Table 3 Details of real-world network datasets

| 数据集 | 节点数 | 边数 | 社团数 |
|------------|------|------|-----|
| Dolphins | 62 | 159 | — |
| Facebook | 4038 | 8834 | — |
| Football | 115 | 613 | 12 |
| Polbooks | 105 | 441 | 3 |
| Cornell | 195 | 304 | 5 |
| Texas | 187 | 328 | 5 |
| Washington | 230 | 446 | 5 |
| Wisconsin | 265 | 530 | 5 |

7.2 基线方法

实验选取了两类主流基线方法进行比较。

1) 基于 NMF/NMTF 的改进模型,包括 NMF^[17], SNMF^[18], NMTF^[9], SCNMF^[20] 和 HPNMF^[22] 5 个基线方法。这类方法均需预设社团数 k ,主要用于与 Adp-NMTF 比较社团划分质量,验证 Adp-NMTF 对 NMTF 模型改进的优势及自适应社团检测性能。

2) 自适应社团检测方法,包括 Louvain^[27], LPA^[28] 和 Infomap^[4] 3 个经典社团检测方法。这类方法能自适应确定 k ,

用于验证 Adp-NMTF 在自动确定社团数方面的竞争力,即能否自适应确定合理的社团数 k 且具有良好的社团划分性能。其中,Louvain 算法调用 python-louvain 库实现,LPA 算法调用 NetworkX 库实现,Infomap 调用其官方库实现。

7.3 社团检测性能评估

对于有真实标签的合成网络和真实网络,采用模块度 (Modularity, Q)、标准化互信息 (Normalized Mutual Information, NMI) 和调整兰德指数 (Adjusted Rand Index, ARI) 3 个广泛使用的评价指标;对于无真实标签网络,采用模块度 (Q) 和平均轮廓系数 (Average Silhouette Coefficient, ASC) 作为评价指标。Adp-NMTF 和基线方法的参数均按其默认设置或推荐值设置。

7.3.1 合成网络实验结果

在两组合成网络上进行对比实验。第一组网络(固定规模,变化 μ)的结果如表4所列。随着混合参数 μ 的增大,网络社团结构逐渐模糊,检测难度增加,所有方法的性能均随之下降。测试结果表明,Adp-NMTF 在大多数情况下性能表现最优。例如,当 $\mu=0.3$ 时,Adp-NMTF 的 Q , NMI 和 ARI 相比表现次优的 SNCMF 分别提升了约 2.2%(0.55241 vs. 0.54066),3.5%(0.86999 vs. 0.83443) 和 9.2%(0.78165 vs. 0.71596)。第二组网络(变化规模,固定 $\mu=0.25$)的结果如图2所示,Adp-NMTF 在所有网络中均获得了更优的性能。

表4 第一组合成网络数据集测试结果比较

Table 4 Comparison of test results on the first set of synthetic network datasets

| 数据集 | 指标 | NMF | SNMF | NMTF | HPNMF | SNCMF | Adp-NMTF |
|------------------|-----|---------|---------|---------|---------|----------------|----------------|
| D1 $\mu=0.2$ | Q | 0.60534 | 0.61969 | 0.60518 | 0.66631 | 0.66412 | 0.66960 |
| | NMI | 0.89065 | 0.86860 | 0.81898 | 0.94115 | 0.93706 | 0.94141 |
| | ARI | 0.82592 | 0.76019 | 0.68936 | 0.86173 | 0.90695 | 0.90886 |
| D2 $\mu=0.25$ | Q | 0.50936 | 0.52430 | 0.51420 | 0.57907 | 0.58237 | 0.59056 |
| | NMI | 0.81479 | 0.68698 | 0.74284 | 0.87735 | 0.88711 | 0.90242 |
| | ARI | 0.73277 | 0.50728 | 0.58898 | 0.78847 | 0.79712 | 0.84526 |
| D3 $\mu=0.3$ | Q | 0.43817 | 0.46216 | 0.46931 | 0.52345 | 0.54066 | 0.55241 |
| | NMI | 0.76032 | 0.71030 | 0.70164 | 0.81361 | 0.83443 | 0.86999 |
| | ARI | 0.65084 | 0.54245 | 0.52697 | 0.69323 | 0.71596 | 0.78165 |
| D4 $\mu=0.35$ | Q | 0.33122 | 0.36998 | 0.37964 | 0.43122 | 0.43332 | 0.44265 |
| | NMI | 0.61264 | 0.55571 | 0.54570 | 0.69107 | 0.69364 | 0.70524 |
| | ARI | 0.46870 | 0.36550 | 0.34713 | 0.55653 | 0.56084 | 0.56188 |
| D5 $\mu=0.4$ | Q | 0.22658 | 0.29858 | 0.32639 | 0.33494 | 0.35262 | 0.36327 |
| | NMI | 0.44036 | 0.38613 | 0.42346 | 0.53694 | 0.53351 | 0.56233 |
| | ARI | 0.27634 | 0.18842 | 0.23497 | 0.32362 | 0.36240 | 0.36214 |

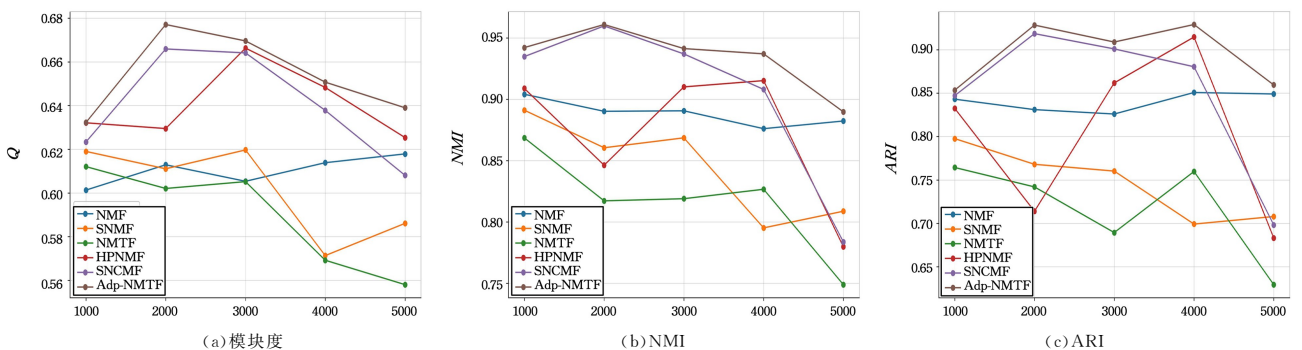


图2 第二组合成数据集测试结果

Fig. 2 Test results of the second synthetic datasets

7.3.2 真实网络实验结果

由于 NMF/NMTF 类基线方法需预设 k , 因此实验选取 4 个有真实标签的网络, 并将 k 设置为真实社团数, 实验结果如图 3 所示。

测试表明, Adp-NMTF 在大多数指标和数据集上表现

最优, 尤其在模块度(Q)和标准化互信息(NMI)上优势显著, 说明其在捕捉社团结构和匹配真实标签方面具有较强能力。在 Cornell 数据集上, Adp-NMTF 的 ARI 略低于 HPNMF 和 SNCMF(0.85 vs 0.86+), 表明其在特定网络结构下仍有优化空间。

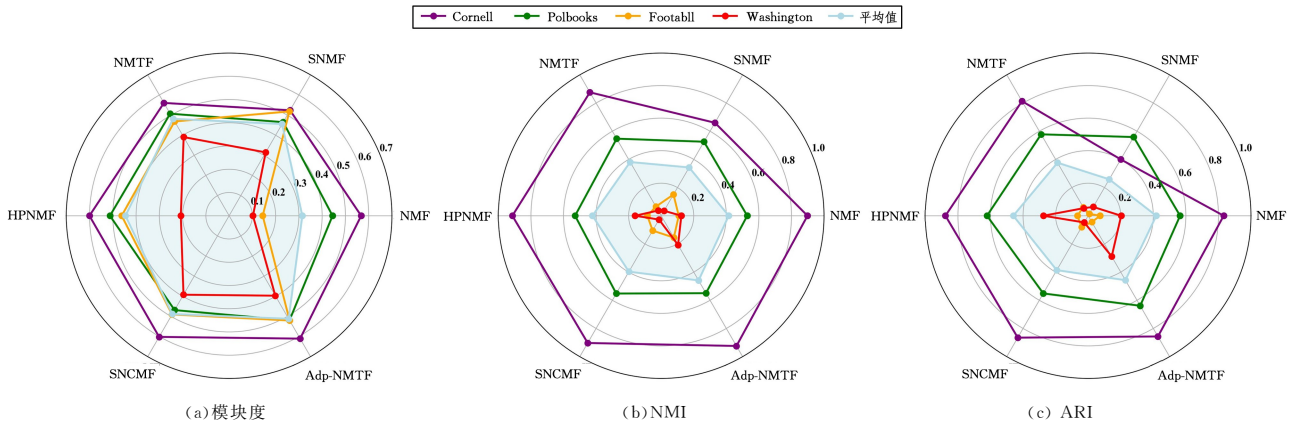


图 3 真实数据集测试结果

Fig. 3 Test results of ground-truth datasets

7.4 社团数量自适应确定评估

为评估 Adp-NMTF 自适应确定社团数 k 的有效性, 在 6 个真实网络上进行实验, 与无需预设 k 的基线方法(Louvain, LPA, Infomap)进行对比, 结果如表 5 所列。

结果表明, Adp-NMTF 能够有效确定社团数量 k , 且获得

最好的社团划分效果。Adp-NMTF 在所有 6 个数据集上的模块度(Q)均高于或与最优方法相当; 对于平均轮廓系数(ASC), Adp-NMTF 在 Dolphins, FaceBook, Teax, Wisconsin 上最优, 在 Football 和 Polbooks 上次于最优方法(Infomap 和 Louvain)。

表 5 自适应确定社团数测试

Table 5 Adaptive determination of the number of communities

| 数据集 | 指标 | Louvain | LPA | Infomap | Adp-NMTF |
|-----------|---------|----------------|---------|----------------|----------------|
| dolphins | 社团数 k | 5 | 6 | 5 | 6 |
| | 模块度 | 0.51766 | 0.49858 | 0.50852 | 0.51868 |
| | 平均轮廓系数 | 0.65557 | 0.57068 | 0.71049 | 0.71346 |
| FaceBook | 社团数 k | 16 | 44 | 76 | 28 |
| | 模块度 | 0.73495 | 0.73684 | 0.81377 | 0.82995 |
| | 平均轮廓系数 | 0.83380 | 0.84095 | 0.77536 | 0.86915 |
| Football | 社团数 k | 9 | 9 | 11 | 12 |
| | 模块度 | 0.60219 | 0.55092 | 0.60306 | 0.61052 |
| | 平均轮廓系数 | 0.76131 | 0.63821 | 0.77567 | 0.77264 |
| polbooks | 社团数 k | 6 | 4 | 5 | 3 |
| | 模块度 | 0.50258 | 0.48114 | 0.52590 | 0.54943 |
| | 平均轮廓系数 | 0.83844 | 0.77796 | 0.75108 | 0.83239 |
| Texas | 社团数 k | 15 | 24 | 26 | 24 |
| | 模块度 | 0.55312 | 0.33895 | 0.53464 | 0.59420 |
| | 平均轮廓系数 | 0.84871 | 0.84292 | 0.85047 | 0.87505 |
| Wisconsin | 社团数 k | 16 | 46 | 35 | 24 |
| | 模块度 | 0.63597 | 0.56727 | 0.59865 | 0.66124 |
| | 平均轮廓系数 | 0.85554 | 0.72434 | 0.82113 | 0.86414 |

7.5 运行效率与收敛性

比较 Adp-NMTF 与 HPNMF 以及 SNCMF 的运行时间, 结果如表 6 所列(表中时间为 10 次独立运行的平均值)。由于 Adp-NMTF 需要自适应搜索最优 k 并多次执行 NMTF, 其运行时间高于 HPNMF 和 SNCMF, 但得益于热启动策略加速了单次 NMTF 收敛, 实际运行时间在可接受范围内。对于中小型网络(Dolphin, Football, Polbooks, Wisconsin), Adp-NMTF 的运行时间在 1~5 秒级别, 满足离线分析

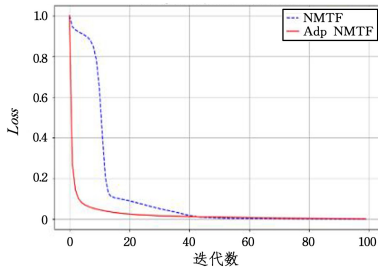
需求; 对于 Facebook 网络, Adp-NMTF(270s)虽慢于 SNCMF(47s)和 HPNMF(157s), 但避免了 SNCMF /HPNMF 所需的 k 值调优过程(通常需尝试多个 k 值), 其端到端效率更具优势。

图 4 展示了 Adp-NMTF 中采用热启动策略(Louvain 初始化+迭代间传递)相比于随机初始化, 对单次 NMTF 收敛速度的提升效果。以 Polbooks 和 Football 数据集为例, 热启动策略显著加快了 NMTF 的收敛速度。

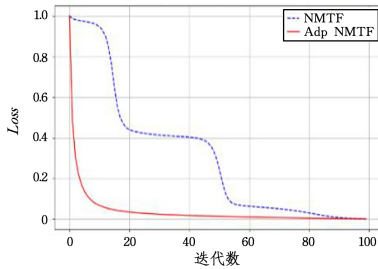
表 6 运行时间比较

Table 6 Comparison of execution time

| | HPNMF | SNCMF | Adp-NMTF |
|-----------|-----------|----------|-----------|
| Dolphin | 0.34903 | 0.50996 | 1.22431 |
| Facebook | 156.73462 | 47.19032 | 270.35672 |
| Football | 1.31163 | 1.67185 | 4.13420 |
| polbooks | 1.42584 | 1.24306 | 1.91287 |
| Teax | 1.47259 | 1.11538 | 5.42406 |
| Wisconsin | 4.27709 | 1.46572 | 3.25610 |



(a) Polbooks



(b) Football

图 4 热启动策略对 NMTF 收敛速度的影响

Fig. 4 Influence of warm-start strategy on NMTF convergence speed

结束语 本文针对非负矩阵三因子分解(NMTF)在网络社团检测中的关键挑战,提出了自适应算法 Adp-NMTF,实现社团数量自适应确定与模型泛化能力提升。实验表明,该算法借助动态反馈调整机制和多约束联合优化,在合成与真实网络上的检测性能均优于基线方法。然而,算法在超参数自适应设置及稀疏大规模网络高效处理上仍需深入研究。未来将探索动态正则化权重分配、稀疏网络优化技术、多模态信息融合及并行化加速,以拓展其在动态网络和超大规模网络中的应用。

参考文献

[1] FORTUNATO S, NEWMAN M E J. 20 years of network community detection[J]. *Nature Physics*, 2022, 18(8): 848-850.

[2] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821-7826.

[3] CHOUDHARY C, SINGH A. Community detection algorithms for recommendation systems: techniques and metrics[J]. *Computing*, 2022, 105(2): 1-37.

[4] JIN D, YU X, LI H, et al. A survey of community detection approaches: From statistical modeling to deep learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(2):

1149-1170.

[5] FORTUNATO S, BARTHÉLEMY M. Resolution limit in community detection[J]. *Proceedings of the National Academy of Sciences*, 2007, 104(1): 36-41.

[6] VON LUXBURG U. A tutorial on spectral clustering[J]. *Statistics and Computing*, 2007, 17(4): 395-416.

[7] ROSVALL M, BERGSTROM C T. Maps of random walks on complex networks reveal community structure[J]. *Proceedings of the National Academy of Sciences*, 2008, 105(4): 1118-1123.

[8] ZHANG Z, CUI P, ZHU W. Deep learning on graphs: A survey [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(1): 249-270.

[9] DING C, LI T, PENG W, et al. Orthogonal nonnegative matrix tri-factorizations for clustering[C] // *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2006: 126-135.

[10] YANG Y, YU S, PAN B, et al. Community Detection in Multiplex Networks Using Orthogonal Non-Negative Matrix Tri-Factorization Based on Graph Regularization and Diversity [J]. *Mathematics*, 2024, 12(8): 1124.

[11] YANG L, ZHANG L. Community Detection Based on Co-regularized Nonnegative Matrix Tri-Factorization in Multi-view Social Networks[C] // *2018 IEEE International Conference on Big Data and Smart Computing*. 2018: 98-105.

[12] LI Y, JIA C, YU J. Survey on community detection algorithms using nonnegative matrix factorization model [J]. *Journal of Frontiers of Computer Science and Technology*, 2016, 10(1): 1-13.

[13] LI Y, LIU X, ZHANG H, et al. Contrastive deep nonnegative matrix factorization for community detection[C] // *2024 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2024: 6725-6729.

[14] LI X, YU W, XU G, et al. MSDA-NMF: A multilayer complex system model integrating deep autoencoder and NMF[J]. *Mathematics*, 2022, 10(15): 2750.

[15] DING C, HE X, SIMON H D. On the equivalence of nonnegative matrix factorization and spectral clustering[C] // *Proceedings of the 2005 SIAM International Conference on Data Mining*. 2005: 126-135.

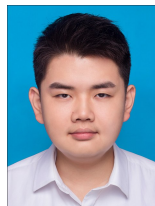
[16] MA X, ZHANG Y, LIU J, et al. Semi-supervised clustering algorithm for community structure detection in complex networks [J]. *Physica A: Statistical Mechanics and its Applications*, 2009, 389(1): 187-197.

[17] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization[J]. *Advances in Neural Information Processing Systems*, 2001, 13: 556-562.

[18] KUANG D, DING C, PARK H. Symmetric nonnegative matrix factorization for graph clustering[C] // *Proceedings of the SIAM International Conference on Data Mining*. 2012: 126-135.

[19] SHELDON B, ROBERTS S. Overlapping community detection using Bayesian non-negative matrix factorization [J]. *Physical Review E*, 2011, 83(6): 066114.

- [20] LIU Z, YUAN G, LUO X. Symmetry and nonnegativity-constrained matrix factorization for community detection[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(9): 1691-1693.
- [21] GUAN J, CHEN B, HUANG X. Community detection via multi-hop nonnegative matrix factorization[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(7): 10033-10044.
- [22] YE F, CHEN C, WEN Z, et al. Homophily preserving community detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31: 2903-2915.
- [23] SONG Y, LI M, LUO X, et al. Improved symmetric and non-negative matrix factorization models for undirected, sparse and large-scaled networks: A triple factorization-based approach[J]. IEEE Transactions on Industrial Informatics, 2020, 16(5): 3006-3017.
- [24] LI Z, YANG J, WANG L, et al. A semi-orthogonal nonnegative matrix tri-factorization algorithm for overlapping community detection[J]. Statistical Papers, 2024, 65(6): 1-19.
- [25] ZHANG J, WANG F, ZHOU J. Community detection based on nonnegative matrix tri-factorization for multiplex social networks[J]. Journal of Complex Networks, 2024, 12(2): 1-16.
- [26] NEWMAN M E J. Community detection in networks: Modularity optimization and maximum likelihood are equivalent[J]. Physical Review E, 2016, 94(5): 052315.
- [27] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008(10): P10008.
- [28] RAGHAVAN U N, ALBERT R, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks[J]. Physical Review E, 2007, 76(3): 036106.



LI Xilong, born in 1997, postgraduate. His main research interest is cyberspace situational awareness.



LIU Yan, born in 1979, Ph. D, professor. Her main research interests include cyberspace situational awareness and cybersecurity.

(责任编辑:何杨)