

基于层次结构嵌入的动态社区检测

朱瑞, 叶亚琴, 李圣文, 汤子健, 肖玥

引用本文

朱瑞, 叶亚琴, 李圣文, 汤子健, 肖玥. [基于层次结构嵌入的动态社区检测](#) [J]. 计算机科学, 2025, 52(8): 127-135.

ZHU Rui, YE Yaqin, LI Shengwen, TANG Zijian, XIAO Yue. [Dynamic Community Detection with Hierarchical Modularity Optimization](#) [J]. Computer Science, 2025, 52(8): 127-135.

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)

[视觉表征学习综述](#)

Review of Visual Representation Learning

计算机科学, 2024, 51(11): 112-132. <https://doi.org/10.11896/jsjcx.231100089>

[锚社区时序网络图生成算法](#)

Generation Algorithm of Temporal Networks with Anchor Communities

计算机科学, 2024, 51(1): 113-123. <https://doi.org/10.11896/jsjcx.231000153>

[跨模态目标重识别研究综述](#)

Survey on Cross-modality Object Re-identification Research

计算机科学, 2024, 51(1): 13-25. <https://doi.org/10.11896/jsjcx.yg20240103>

[基于异构网络表征学习的作者学术行为预测](#)

Author's Academic Behavior Prediction Based on Heterogeneous Network Representation Learning

计算机科学, 2022, 49(9): 76-82. <https://doi.org/10.11896/jsjcx.210900078>

[基于可变形图卷积的点云表征学习](#)

Deformable Graph Convolutional Networks Based Point Cloud Representation Learning

计算机科学, 2022, 49(8): 273-278. <https://doi.org/10.11896/jsjcx.210900023>

基于层次结构嵌入的动态社区检测

朱瑞¹ 叶亚琴^{1,2} 李圣文¹ 汤子健¹ 肖玥¹

1 中国地质大学(武汉)计算机学院 武汉 430078

2 智能地学信息处理湖北省重点实验室 武汉 430078

(1202110914@cug.edu.cn)

摘要 作为理解复杂网络内在模式和组织结构的有力工具,动态社区检测揭示了网络中相互密集连接的节点集合的演化过程,是社会科学、城市规划等领域的一项基本任务。近年来,动态社区检测领域涌现了大量基于表征学习的方法。这些方法通过融合网络拓扑结构和演化特征,将节点映射到低维连续向量空间中,实现了对节点相似度和差异性的准确度量。然而,现有表征学习方法未能充分考虑节点的长距离信息,导致节点表征未能全面捕捉网络的全局特征。对此,提出了一种基于层次结构嵌入的动态社区检测算法 DHM。具体而言,DHM 基于网络的多粒度特性生成层次结构,并通过设计自底向上和自顶向下的消息传递机制,将不同层次的节点组织关系嵌入至节点表征。在人工网络数据集和真实网络数据集上进行的实验结果显示,DHM 在标准互信息、调整兰德指数和模块度 3 个指标上优于现有的动态社区检测算法,可以较好地完成时序网络下的社区检测任务。

关键词: 动态社区检测;时序网络;模块度优化;层次结构;表征学习

中图分类号 TP181

Dynamic Community Detection with Hierarchical Modularity Optimization

ZHU Rui¹, YE Yaqin^{1,2}, LI Shengwen¹, TANG Zijian¹ and XIAO Yue¹

1 School of Computer Science, China University of Geosciences, Wuhan 430078, China

2 Hubei Key Laboratory of Intelligent Geo-Information Processing, Wuhan 430078, China

Abstract Serving as a powerful tool for understanding intrinsic patterns and organizational structures within complex networks, dynamic community detection unveils the evolutionary process of densely connected sets of nodes, standing as a fundamental task in disciplines such as social sciences and urban planning. In recent years, various methods based on representation learning have been applied to the field of dynamic community detection. These methods map structured nodes into low-dimensional continuous latent space by integrating network topology and evolution characteristics, achieving reliable measurement of node similarity and difference. However, existing representation learning methods inadequately consider nodes' long-range information, falling short of capturing global structural features. To address this issue, this paper proposes DHM, which combines modularity optimization and hierarchical structure embedding to capture long-range node interactions. Specifically, DHM generates hierarchical organization based on the network's multi-granularity nature and embeds different levels of node relationships into node representations through bottom-up and top-down message-passing mechanisms. Experimental results on synthetic and real-world network datasets demonstrate that DHM outperforms existing dynamic community detection algorithms in terms of normalized mutual information, adjusted Rand index, and modularity, and can effectively detect communities in temporal networks.

Keywords Dynamic community detection, Dynamic networks, Modularity optimization, Hierarchical structure, Representation learning

到稿日期:2024-06-17 返修日期:2024-08-31

基金项目:国家自然科学基金(42071382);智能地学信息处理湖北省重点实验室开放研究项目(KLIGIP-2022-B14);桐兴(湖北)投资咨询有限公司合作项目(2021116490)

This work was supported by the National Natural Science Foundation of China(42071382), Open Research Project of Hubei Key Laboratory of Intelligent Geo-Information Processing(KLIGIP-2022-B14) and Tongxing(Hubei) Investment Consulting Co. Project(2021116490).

通信作者:叶亚琴(yeyaqin@cug.edu.cn)

1 引言

信息技术记录了人类活动所衍生的数据,进而创造出多种类型的网络。例如,研究机构之间的信息共享创造了万维网^[1],人际间的沟通产生了社交网络^[2],研究者的学术交流构成了引文网络^[3]。对这些复杂网络进行深入分析,不仅能够揭示隐藏在海量数据背后的模式和规律,还能够发掘有价值的知识,为指导和改进人类活动提供重要的决策支持^[4]。

复杂网络的一个重要特性是社区结构,即网络本身由若干个子网络(社区)构成^[5]。社区内节点的关联性较强,社区间节点的关联性则较弱。基于此,社区结构能够揭示节点在网络中的行为,可以用来理解网络的组织或功能^[6]。例如,在蛋白质相互作用网络中,社区将细胞内具有相同特定功能的蛋白质分组,可以帮助研究员更好地了解蛋白质在细胞中的作用^[7];在社交网络中,社区表示具有相似兴趣和行为的人群^[8],能够帮助平台提升用户个性化的体验。

现实世界中的社区往往具有时间维度,例如,在社交网络中,新用户的加入和老用户的离开会对网络结构带来扰动,而社区结构也会因这些扰动而产生变化^[9]。为了适应现实网络的动态演化,设计出有效的动态社区检测方法具有重要的研究意义^[10]。

随着人工智能领域的不断发展,动态社区检测领域涌现出一些表征学习模型^[11],它们致力于捕捉网络的复杂模式和演化过程,为强连接节点生成紧密嵌入,并通过分析嵌入在潜在空间中的邻近性来识别社区。然而,现有表示方法未能充分考虑节点的长距离信息,导致节点表征未能全面捕捉网络的全局特征,影响了动态社区检测的准确性。

为了解决上述问题,本文提出基于层次结构嵌入的动态社区检测模型(Dynamic community detection with Hierarchical Modularity optimization, DHM)。具体而言, DHM 在模块度优化的基础上,设计层次编码模块以捕捉节点之间更远距离的相互作用。进一步讲, DHM 基于网络的多粒度特性生成多层的网络组织,并通过设计自底向上和自顶向下的消息传递机制,将不同层次的节点组织关系嵌入至节点表征。DHM 旨在为节点生成更全面的嵌入向量,以提高动态社区检测的性能。本文联合模块度优化和层次结构嵌入两个维度对表示学习予以聚类优化,并设计联合的损失函数用于模型训练。本文在多个人工和真实网络数据集上进行了实验,结果证明了所提模型的有效性。

本文的主要贡献如下:

- 1) 提出了一种动态社区检测网络嵌入模型,该模型能够在联合框架中捕捉网络拓扑结构的动态变化、实现模块度优化并感知节点的远距离信息。
- 2) 开发了一个层次结构嵌入模块,该模块旨在为节点生成更全面的嵌入向量,以增强节点表征的判别性。
- 3) 在 6 个合成网络数据集与 8 个真实网络数据集上验证了所提模型的有效性。

2 相关工作

自 21 世纪以来,动态社区检测一直是计算机领域中的基

础性研究课题之一^[12]。本文将现有的检测方法根据求解方式的不同划分为两个主要类别:传统检测算法和表征学习模型。下文将从上述两个方面总结分析国内外动态社区检测的相关工作。

2.1 传统动态社区检测算法

传统社区检测算法将社区检测问题形式化为一个明确定义的问题,并通过制定策略来显式求解最优划分。本文基于算法对时间维度的不同处理方式,将现有检测算法划分为 3 个子类别:基于匹配、基于增量更新,以及基于演化建模的动态社区检测算法。

1) 基于匹配的方法不显式地考虑网络时序信息,而是将静态算法直接应用于动态网络。这类方法大致分为两个阶段:第一阶段在每个快照上独立地使用静态方法挖掘社区结构;第二阶段进行匹配,即判断相邻快照上的两个社区是否为同一个动态社区。匹配阶段的关键在于判断两个社区的相似性是否大于预先定义的阈值。若相似性大于阈值,则认定这两个社区为同一个动态社区;反之,则不认为。Hopcroft 等^[10]最早利用凝聚聚类算法^[13]识别社区,并通过计算相邻网络快照之间社区的相似度来检查社区随时间的演变。Tajeuna 等^[14]以向量形式表示一个社区与所有其他检测到的社区之间的共同关系,并引入 Jaccard 系数^[15]来比较整个观察期间的社区。Jia 等^[16]使用 Leiden 算法^[17]得到社区结果,然后根据 Jaccard 系数量化社区相似性以判定动态事件的发生。

2) 基于增量分析的算法显式地考虑相邻快照的拓扑变化。这类方法以前一个快照的社区划分为基准,并根据预定义的拓扑变化事件,增量地更新社区结构。QCA^[18]预定义了 4 种事件类型,即新增节点、新增连边、删除节点和删除连边,以自适应方式检测和追踪社区结构的频繁变化。DynaMo^[19]考虑了边的权重变化,通过迭代添加优先级最高的节点或边,并基于模块度最大化得到有效的社区划分。BBTA^[20]使用骨干边来反映社区内关键的边,使用桥边来描述社区之间的关键连接,并利用网络变化率来评估网络的演化。

3) 基于启发式演化算法的方法引入时间平滑框架来量化两个连续快照之间的相似性。FacetNet^[21]使用随机块模型生成社区,联合 Dirichlet 分布的概率模型来描述社区的演化,并引入代价函数来衡量某一时刻社区结构的质量。Fu 等^[22]联合模块密度函数以及图划分的否定平均关联函数以评估动态网络中的社团结构。Pisces^[23]将进化谱聚类扩展到动态社团检测,并采用全局平滑策略以提高算法的鲁棒性。

2.2 基于表征学习的动态社区检测方法

与传统检测算法不同,表征学习方法通过将高维网络数据映射到低维向量空间,使得原始网络中相近的节点在表示空间中具有相似的嵌入,以求解社区检测问题。鉴于不同模型对网络拓扑结构处理方式的差异,本文将这些模型划分为 3 个主要类别:基于非负矩阵分解的动态社区检测方法、基于自编码器的动态社区检测方法,以及基于图神经网络的动态社区检测模型。

1) 基于非负矩阵分解(Nonnegative Matrix Factorization, NMF)的动态社区检测使用邻接矩阵建模网络拓扑结构,并结合演化算法建模拓扑结构的时序特征。sE-NMF^[24]基于演

化的非负矩阵分解框架,并基于数据的先验信息提出了一种半监督的动态社区检测模型。jLDEC^[25]将特征学习和图聚类整合到一个目标函数中,并从边的角度定义聚类精度和漂移来量化社区的动态特性。在jLDEC的基础上,jLMDC^[26]进一步利用前一个快照的顶点特征与当前快照的拓扑结构关系来量化单个节点的动态,并在优化过程中仅更新动态顶点的特征,以提高算法的时间效率。

2)基于自编码器(Autoencoder, AE)的动态社区检测,通过拓扑重构方式,生成捕捉社区结构的低维表征。基于Encoder-Decoder架构,DynGEM^[27]通过启发式策略增量地增加神经网络的大小,为节点数量不断增长的动态网络生成稳定的节点嵌入。Dyngraph2vec^[28]进一步引入LSTM学习动态网络的时序特征,以捕捉网络演化的连续性变化。DANE-CD^[29]放宽了模块度最大化中硬社区分配约束,它基于低维表征计算节点-社区隶属矩阵,并以可微的方式优化软模块度以求解社区划分。

3)基于图神经网络(Graph Neural Network, GNN)技术的动态社区检测,通过特征传播和邻居聚合的方式,捕捉了节点的邻域特征和上下文信息。EvolveGCN^[30]通过GCN学习节点的空间连接模式,并使用RNN建模GCN的演化参数以捕捉网络的时序特性。DGCN^[31]采用数据增强方式构建负样本,并以互信息最大化为目标,保证在原始网络中相近的节点在嵌入空间中同样接近。在DGCN的基础上,ODDGI^[32]在信息聚合时考虑了节点的二阶相似性,并引入时间平滑项约束嵌入的稳定性。

尽管上述相关工作取得了一定进展,但它们通常仅考虑节点的局部拓扑结构,忽略了节点间的远距离依赖关系,缺乏将长距离信息提取与社区检测相结合的统一过程。

3 问题与定义

本章将定义动态网络中的社区检测问题。

定义 1(动态网络) 动态网络定义为 $\mathbf{G} = \{G^1, G^2, \dots, G^t, \dots, G^T\}$, 其中 T 表示 \mathbf{G} 包含的时间步数, 元素 G^t 是时间步 $t (1 \leq t \leq T)$ 处的观测图。每个网络快照表示为 $G^t = (V^t, E^t)$, 其中 $V^t = \{v_1, v_2, \dots, v_N\}$ 是 \mathbf{G} 的一个共享节点集, E^t 是 G^t 的边集。 $\mathbf{A} = \{A^1, A^2, \dots, A^t, \dots, A^T\}$ 代表 \mathbf{G} 的邻接矩阵序列, 其中 $A_{(i,j)}^t = 1$ 表示节点 $v_i \in V_t$ 和节点 $v_j \in V_t$ 之间存在一条边, 反之 $A_{(i,j)}^t = 0$ 。

定义 2(动态网络表示学习) 给定动态网络 \mathbf{G} , 表征学习为其生成一组节点表示矩阵集合 $\mathbf{Z} = \{Z^1, Z^2, \dots, Z^t, \dots, Z^T\}$, 其中 $Z^t \in \mathbb{R}^{N \times D}$ 且 $D \ll N$ 。

定义 3(动态网络社区检测) 动态网络社区检测将每个快照 G^t 划分为 K 个彼此不相交的组: $C^t = \{C_1^t, C_2^t, \dots, C_K^t\}$, 即 $\bigcup_{k=1}^K C_k^t = V^t$ 且 $C_i^t \cap C_j^t = \emptyset, \forall i \neq j$ 。

4 DHM 模型

本章介绍基于层次结构嵌入的动态社区检测模型。图 1 展示了 DHM 的架构, 该方法在模块度优化的基础上, 引入了层次结构嵌入模块。编码器将结构信息和时序信息映射至潜在空间; 解码器将模块度最大化问题转换为可微的矩阵重构问题; 层次结构嵌入模块通过自底向上和自顶向下的消息传递机制, 捕捉节点之间更远距离的相互作用。

DHM 的执行流程包括 5 个关键步骤: 1) 基于网络中节点的邻域关系生成一系列不同粒度的网络, 其中粗粒度节点是由网络中一些具有高密度边的细粒度节点抽象而来; 2) 将 Encoder 生成的嵌入矩阵作为底层网络的输入, 并在所有粒度的网络执行特征聚合, 得到所有粒度的初始节点表示; 3) 通过自底向上的消息传递, 更新每层的节点表征; 4) 通过自顶向下的消息传递, 聚合不同粒度的节点表示, 得到融合的网络节点表示; 5) 综合层次损失和重构损失, 训练得到最终的网络节点表示。

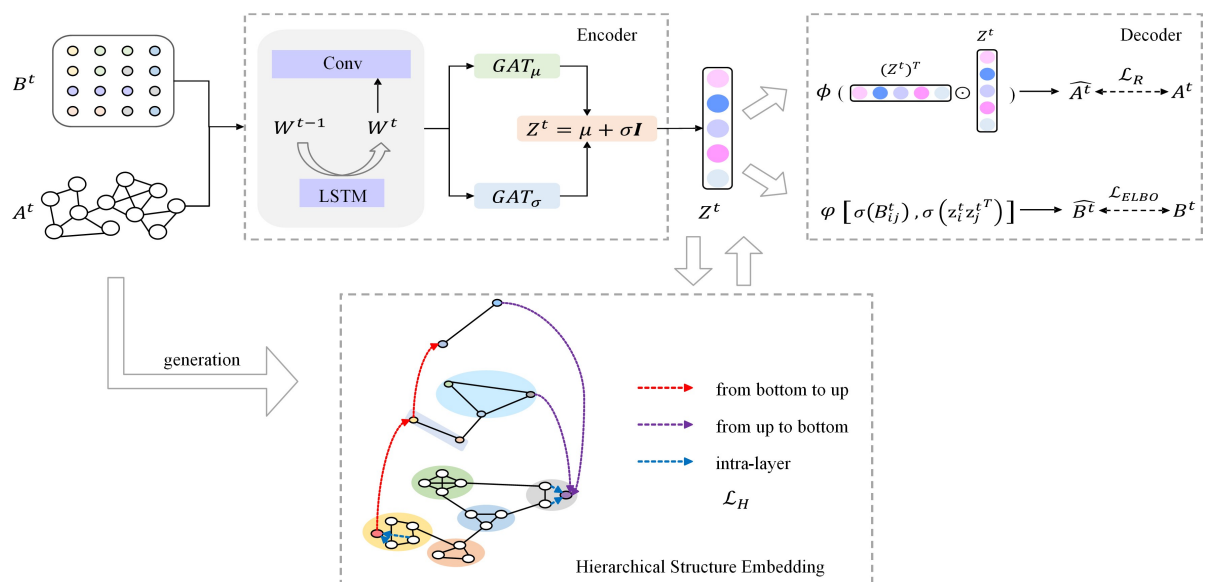


图 1 DHM 模型框架图

Fig. 1 Overall architecture of DHM

DHM 旨在借助层次结构捕捉节点的远程依赖关系, 通

过联合模块度优化, 综合考虑空间中不同粒度的节点信息,

以生成丰富和准确的节点表示。

4.1 编码器

给定邻接矩阵 \mathbf{A} 和计算得到的模块度矩阵 \mathbf{B} , 编码器的目标是生成嵌入矩阵 \mathbf{Z} 。基于变分图自编码器 (Variational Graph Autoencoder, VGAE)^[33] 的设计理念, DHM 的编码过程定义为:

$$q(\mathbf{Z} | \mathbf{B}, \mathbf{A}') = \prod_{i=1}^N q(z_i^t | \mathbf{B}', \mathbf{A}') \quad (1)$$

其中, $q(z_i^t | \mathbf{B}', \mathbf{A}')$ 是节点 i 后验分布的变分近似, 具体表示为:

$$q(z_i^t | \mathbf{B}', \mathbf{A}') = \mathcal{N}(z_i^t | \mu_i, \text{diag}(\sigma_i^2)) \quad (2)$$

具体而言, 嵌入矩阵 \mathbf{Z} 通过重新参数化技巧计算得到:

$$\mathbf{Z} = \mu + \sigma \mathbf{1} \quad (3)$$

进一步讲, 为了捕获拓扑动态并保留局部结构特征, 本文选择具有演化权重的 GCN 层将节点映射为嵌入 \mathbf{H} :

$$\mathbf{H} = \mathcal{O}(L' \mathbf{B}' \mathbf{W}_0') \quad (4)$$

$$\mathbf{W}_0' = LSTM(\mathbf{W}_0'^{-1}) \quad (5)$$

其中, \mathcal{O} 表示激活函数, $L' = \mathbf{D}^{-\frac{1}{2}} (\mathbf{A}' + \mathbf{I}) \mathbf{D}^{-\frac{1}{2}}$ 为拉普拉斯矩阵。然后选择两个 GCN 层分别对 \mathbf{H} 的均值 μ 和方差 σ 进行建模, 分别表示为:

$$\mu = \tanh(L' \mathbf{H}' \mathbf{W}_1) \quad (6)$$

$$\log \sigma = \tanh(L' \mathbf{H}' \mathbf{W}_2) \quad (7)$$

其中, \mathbf{W}_1 和 \mathbf{W}_2 代表可学习的参数矩阵。

4.2 解码器

给定嵌入矩阵 \mathbf{Z} , 解码器一方面通过重构邻接矩阵确保嵌入空间能够有效地捕捉和保留原始图的结构信息。另一方面, 通过重构模块度矩阵 \mathbf{B} 来实现模块度最大化, 以促进节点在社区内的紧密连接并增强社区间的分离度。具体而言, DHM 利用 \mathbf{Z} 和 $(\mathbf{Z}')^T$ 的内积来重构原始的邻接矩阵 \mathbf{A} :

$$\hat{\mathbf{A}} = \text{sigmoid}(\mathbf{Z} (\mathbf{Z}')^T) \quad (8)$$

其相应的重构损失为:

$$\mathcal{L}_R = \text{BinaryCrossEntropy}(\hat{\mathbf{A}}, \mathbf{A}') \quad (9)$$

最小化 \mathcal{L}_R 保证了嵌入过程能够有效地捕获局部拓扑结构信息。

基于 Eckart 和 Young 矩阵重构定理^[34], 模块度最大化和模块度的矩阵重构具有等价性。为了实现模块度优化, DHM 旨在最大化下列条件概率:

$$p(\hat{\mathbf{B}} | \mathbf{Z}) = \prod_{i=1}^N \prod_{j=1}^N p(\hat{B}_{ij}' | z_i^t, z_j^t) \quad (10)$$

其中, \hat{B}_{ij}' 为 B_{ij}' 的重构值, z_i^t 和 z_j^t 是节点 i 和 j 由式(3)计算而来的节点表征向量。

根据先前的工作^[35], 式(10)的变分下界 $\mathcal{L}_{\text{ELBO}}(\phi, \theta)$ 定义为:

$$\mathcal{L}_{\text{ELBO}}(\phi, \theta) = \mathbb{E}_{q_{\phi}(\mathbf{Z}' | \mathbf{B}', \mathbf{A}')} [\log p_{\theta}(\hat{\mathbf{B}} | \mathbf{Z}')] - \text{KL}[q_{\phi}(\mathbf{Z}' | \mathbf{B}', \mathbf{A}') \parallel p(\mathbf{Z}')] \quad (11)$$

其中, 第一个分量表示 \mathbf{B}' 和 $\hat{\mathbf{B}}$ 之间的重建损失, 第二个分量表示 KL 散度, $(\phi, \theta) \in \{\mathbf{W}_0', \mathbf{W}_1, \mathbf{W}_2\}$ 为参数空间。然后, 取 $p(\mathbf{Z}') = \prod_i p(z_i^t) = \prod_i \mathcal{N}(z_i^t | 0, \mathbf{I})$, 那么 $\log p(\hat{B}_{ij}' | z_i^t, z_j^t)$ 可以

被进一步表示为:

$$\log p(\hat{B}_{ij}' | z_i^t, z_j^t) = \sigma(B_{ij}') \log(\sigma(z_i^t z_j^{tT})) + (1 - \sigma(B_{ij}')) \log(1 - \sigma(z_i^t z_j^{tT})) \quad (12)$$

最大化 $\mathcal{L}_{\text{ELBO}}(\phi, \theta)$ 能够确保动态嵌入过程有效地捕捉模块度信息, 生成面向动态社区检测任务的节点表征。

4.3 层次结构嵌入模块

层次结构嵌入模块将层次结构中蕴含的节点远程依赖关系嵌入表征学习过程。

4.3.1 层次结构生成

该步骤在保留原始网络结构特性的前提下, 通过递归地合并节点以缩小网络规模, 从而获得不同粒度的网络, 并根据合并关系构建网络的层次结构。具体而言, 给定一个网络快照 $G = (V, E)$, 其节点集合 V 可以由 L 个不同层次的超级节点组成, 即 $G_H = \{G_1, G_2, \dots, G_L\}$, 其中 G_{l-1} ($2 \leq l \leq L$) 的密集互连节点组合成 G_l 的超级节点。

4.3.2 自底向上消息传递

基于前文生成的层次序列 $\{G_1, G_2, \dots, G_L\}$, 对 G_l ($l > 1$) 的节点表征进行独立初始化:

$$\mathbf{Z}_l = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}}' \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}^l \mathbf{W}) \quad (13)$$

得到初始的表征序列 $\{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_L\}$, 其中, \mathbf{Z}_1 为编码器的输出。

然后, 如图 2(a) 所示, 使用 G_{l-1} ($l \geq 2$) 中的节点表示更新层次结构中 G_l 的节点表示, 进行自底向上的消息传递:

$$z_i^l = \frac{1}{|P_i| + 1} (\sum_{j \in P_i} z_j^{l-1} + z_i^l) \quad (14)$$

其中, P_i 为 G_l 中的超级节点 i 在 G_{l-1} 中对应的节点集合。经过自底向上的消息传递, 得到的表征序列为 $\{\mathbf{Z}_1', \mathbf{Z}_2', \dots, \mathbf{Z}_L'\}$ 。

4.3.3 层内消息传递

如图 2(b) 所示, 层内消息传递将自底向上地对更新后的节点表征再一次进行特征传播, 以确保每个节点都能考虑到最新的邻居节点的特征和状态:

$$\mathbf{Z}_l'' = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}}' \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{Z}_{(l-1)}' \mathbf{W}) \quad (15)$$

经过层内的消息传递, 得到的表征序列为 $\{\mathbf{Z}_1'', \mathbf{Z}_2'', \dots, \mathbf{Z}_L''\}$ 。

4.3.4 自顶向下消息融合

如图 2(c) 所示, 自顶向下的消息融合旨在将粗粒度的节点信息叠加到原始的节点表征。首先, 计算每个粗粒度的节点表征与原始节点表征的注意力系数:

$$\alpha_i = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}(z_{0i}'') \parallel \mathbf{W}(z_{li}'')]))}{\sum_{l=0}^L \exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}(z_{0i}'') \parallel \mathbf{W}(z_{li}'')]))} \quad (16)$$

其中, z_{li}'' 表示第 l 层节点 i 的表征。然后, 基于注意力系数, 融合得到底层节点的表征:

$$z_{0i}''' = \sigma(\sum_{l=0}^L \alpha_i z_{li}'') \quad (17)$$

经过自顶向下的消息传递, 得到的表征序列为 $\{\mathbf{Z}_1''', \mathbf{Z}_2''', \dots, \mathbf{Z}_L'''\}$ 。

通过上述 3 个步骤, 能够在高层进行消息共享, 并将信息传播到低层节点, 以嵌入节点之间的长距离信息。消息传递的策略选择将在 5.5 节中详细探讨。

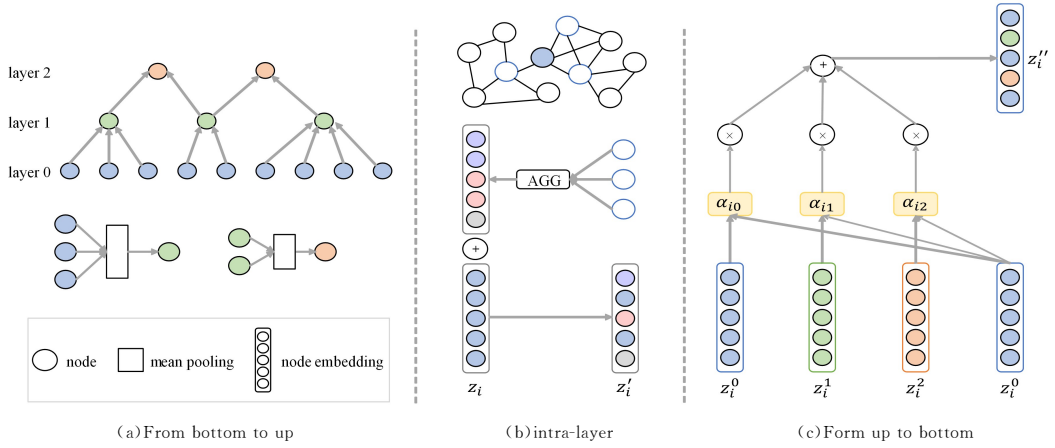


图 2 层次结构嵌入模块

Fig. 2 Hierarchical structure embedding module

5 实验

5.1 数据集

为了评估 DHM 的性能,本文在 6 个人工数据集和 8 个真实数据集上进行了实验。

1)人工数据集由设计良好的图生成器生成的网络组成,这些网络具有真实的拓扑特征和预设的社区结构,被广泛用于社区发现任务^[36]。本文选择 GN^[37]作为静态基准数据集,选择 Greene^[38](又名 SYN-EVENT)作为动态基准数据集。其中,GN 包括 SYN-FIX 和 SYN-VAR 两类数据集。SYN-FIX 数据集在每个时间步的社区数量是固定的,SYN-VAR 数据集在每个时间步的社区数量是可变的。Greene 基准进一步考虑了社区的生命周期:该基准从一个具有真实社区的静态合成图出发,并随机置换若干的节点成员以模拟节点-社区网络演化事件。根据社区的变化行为分为 4 个基于事件的数据集:Birthdeath,Expand contract,Hide 和 Mergesplit。

2)真实数据集用于验证所提出的方法在现实网络中的应用效果。本文选取 3 个真值已知的数据集 InVS15^[39],Primary^[39]和 High School^[39];以及 5 个真值未知的数据集 Enron^[40],SFHH^[39],Email-EU^[41],Voles^[42]和 fb^[42]。

14 个数据集的统计信息如表 1 所列,其中, $N, \bar{M}, \tau, \bar{K}$ 以及 GT 分别表示节点个数、平均边数、快照个数、平均社区个数以及是否有标签真值(若有标签真值,取值为‘Y’,反之为‘U’)。

表 1 数据集的统计信息

Table 1 Statistics of selected datasets

| Dataset | N | \bar{M} | τ | \bar{K} | GT |
|-----------------|------|-----------|--------|-----------|------|
| SYN-FIX | 256 | 2560 | 10 | 4 | Y |
| SYN-VAR | 256 | 2560 | 10 | 6 | Y |
| Birthdeath | 1000 | 3324 | 10 | 33 | Y |
| Expand contract | 1000 | 3849 | 10 | 33 | Y |
| Hide | 1000 | 3623 | 10 | 30 | Y |
| Mergesplit | 1000 | 3932 | 10 | 33 | Y |
| InVS15 | 219 | 3274 | 14 | 12 | Y |
| Primary School | 242 | 2635 | 6 | 11 | Y |
| High School | 327 | 1493 | 9 | 7 | Y |
| Enron | 151 | 313 | 12 | — | U |
| SFHH | 403 | 2156 | 6 | — | U |
| Email-EU | 986 | 3493 | 18 | — | U |
| Voles | 1686 | 759 | 7 | — | U |
| fb | 1899 | 4331 | 6 | — | U |

5.2 基线

为验证所提模型的有效性,本文选择了 8 个基线,包括 4 个传统算法 FacetNet^[21],Pisces^[23],ECD^[43]和 Dyperm^[44],以及 4 个基于表征学习的方法 TIMERS^[45],Dyngraph2vec^[28],EvolveGCN^[30]和 DGCN^[31]。其中 Dyngraph2vec 包含 3 种变体: DynAE, DynAERNN 和 DynRNN。

值得注意的是,对于基于嵌入的方法,在真值未知的数据集上,最大模块度下 k -means 的 k 值被确定为社区的数量。

5.3 评价指标

本文采用 3 种广泛使用的指标来评估 DHM 模型的通用性和有效性,包括标准化互信息(NMI)^[46]、调整兰德指数(ARI)^[47]和模块度(Q)^[37]。

1)标准互信息(Normalized Mutual Information, NMI)。在信息论中,互信息(Mutual Information)衡量了两个随机变量之间的关联程度。给定一个预测划分 \hat{C} 和一个真实划分 C' ,它们的互信息计算式如下:

$$MI(\hat{C}, C') = H(\hat{C}) + H(C') - H(\hat{C}, C') \quad (18)$$

其中, $H(\cdot)$ 为熵函数。标准互信息是互信息的一种标准化形式,用于解决数据集大小可能导致的偏差问题。标准互信息计算式如下:

$$NMI(\hat{C}, C') = \frac{2MI(\hat{C}, C')}{H(\hat{C}) + H(C')} \quad (19)$$

2)调整兰德指数(Adjusted Rand Index, ARI)。兰德指数(Rand Index, RI)测量了两个数据分区之间相似对的数量与不相似对的数量之和的比率。ARI 是 RI 的一种修正,它考虑了由于随机性引起的指标分数的偏差。给定一个预测划分 \hat{C} 和一个真实划分 C' ,ARI 定义为:

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}} \quad (20)$$

其中, n_{ij} 表示 \hat{C} 的第 i 个社区与 C' 的第 j 个社区的公共元素个数, $a_i = \sum_j n_{ij}$ 并且 $a_j = \sum_i n_{ij}$ 。

3) 模块度 (Modularity, Q)。模块度指数 Q 量化了同一社区中节点相对于不同社区中节点之间连接的密集程度, 其定义如下:

$$Q = \frac{1}{2m^t} \sum_{i,j} \left[A_{ij} - \frac{k_i^t k_j^t}{2m^t} \delta(C_i^t, C_j^t) \right] \quad (21)$$

其中, m^t 为 t 时刻的总边数; A_{ij} 为邻接矩阵的元素, 取值为 1 或 0 表示节点 i 和 j 有或无连接边; k_i^t 为节点 i 的度, 表示节点 i 的邻居个数; δ 为节点的社区隶属度函数, 当 i 和 j 属于同一个社区时, $\delta(C_i^t, C_j^t) = 1$, 反之为 0。

对于上述指标, 得分越高表示检测性能越好。

5.4 实验设置

实验在 Ubuntu 20.04 环境进行, 使用单个 NVIDIA RTX 3080 GPU, CUDA 12.0。Python 环境版本为 3.6.13, 由 torch, sklearn 和 networkx 包支持。采用 Adam 作为优化器进行训练, 初始学习率为 0.01。

在 DHM 模型中, 超参数 L, β 以及 D 分别控制了层次结构的深度、层次嵌入模块, 以及嵌入维数对模型的贡献度。经过实验调式, 表 2 列出了它们在本文所有数据集下的具体取值。

表 2 参数设置

Table 2 Hyperparameter setting

| Dataset | L | β | D |
|-----------------|-----|---------|-----|
| SYN-FIX | 2 | 1 | 8 |
| SYN-VAR | 2 | 1 | 8 |
| Birthdeath | 3 | 1.25 | 16 |
| Expand contract | 3 | 1.25 | 16 |
| Hide | 3 | 1 | 16 |
| Mergesplit | 3 | 1.5 | 16 |
| InVS15 | 2 | 1 | 8 |
| Primary | 2 | 1 | 16 |
| HighSchool | 2 | 1 | 16 |
| Enron | 2 | 1 | 8 |
| SFHH | 2 | 1.25 | 16 |
| Email-EU | 2 | 1 | 16 |
| Voles | 3 | 1 | 16 |
| fb | 3 | 1.25 | 32 |

表 3 在 6 个人工数据集上的平均得分

Table 3 Average score on six synthetic datasets

| Method | SYNFIX | | SYNVAR | | Birthdeath | | Expand contract | | Hide | | Mergesplit | |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|--------------|--------------|--------------|--------------|--------------|
| | NMI | ARI | NMI | ARI | NMI | ARI | NMI | ARI | NMI | ARI | NMI | ARI |
| FacetNet | 0.921 | 0.931 | 0.750 | 0.626 | 0.943 | 0.874 | 0.937 | 0.842 | 0.950 | 0.886 | 0.908 | 0.747 |
| Pisces | 0.998 | 0.998 | 0.992 | 0.982 | 0.918 | 0.638 | 0.946 | 0.742 | 0.932 | 0.649 | 0.971 | 0.880 |
| ECD | 0.732 | 0.589 | 0.817 | 0.675 | 0.879 | 0.576 | 0.983 | 0.972 | 0.937 | 0.802 | 0.999 | 0.998 |
| Dyperm | 0.679 | 0.723 | 0.643 | 0.592 | 0.430 | 0.144 | 0.675 | 0.072 | 0.465 | 0.171 | 0.360 | 0.158 |
| TIMERS | 0.881 | 0.904 | 0.789 | 0.698 | 0.646 | 0.465 | 0.681 | 0.516 | 0.673 | 0.456 | 0.680 | 0.540 |
| DynAE | 0.739 | 0.738 | 0.881 | 0.866 | 0.675 | 0.290 | 0.599 | 0.231 | 0.596 | 0.251 | 0.552 | 0.227 |
| DynAERNN | 0.960 | 0.971 | 0.968 | 0.972 | 0.800 | 0.402 | 0.860 | 0.612 | 0.838 | 0.570 | 0.866 | 0.687 |
| DynRNN | 0.016 | 0.003 | 0.067 | 0.026 | 0.458 | 0.213 | 0.512 | 0.250 | 0.474 | 0.238 | 0.478 | 0.244 |
| EvolveGCN | 0.973 | 0.980 | 0.989 | 0.991 | 0.976 | 0.925 | 0.968 | 0.895 | 0.952 | 0.839 | 0.966 | 0.912 |
| DGCN | 0.989 | <u>0.992</u> | <u>0.992</u> | <u>0.994</u> | 0.982 | 0.921 | 0.990 | 0.963 | <u>0.983</u> | 0.904 | 0.963 | 0.874 |
| DHM | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.994 | 0.997 | 0.992 | 0.999 | 0.999 | <u>0.993</u> | <u>0.983</u> |

表 4 在 8 个现实数据集上的平均得分

Table 4 Average score on eight real-world datasets

| Method | InVS15 | | Primary | | High School | | Enron | SFHH | Email-EU | Voles | fb |
|----------|--------------|--------------|--------------|--------------|-------------|-------|--------------|-------|----------|-------|-------|
| | NMI | ARI | NMI | ARI | NMI | ARI | Q | Q | Q | Q | Q |
| FacetNet | 0.331 | 0.168 | <u>0.824</u> | <u>0.685</u> | 0.421 | 0.237 | 0.429 | 0.263 | 0.453 | 0.373 | 0.241 |
| Pisces | 0.437 | <u>0.191</u> | 0.808 | <u>0.685</u> | 0.647 | 0.435 | 0.443 | 0.142 | 0.342 | * | * |
| ECD | <u>0.483</u> | 0.089 | 0.598 | 0.330 | 0.692 | 0.543 | 0.617 | 0.268 | 0.428 | * | * |

5.5 结果分析

1) 在人工数据集上的实验结果。

表 3 列出了 DHM 以及所有基线在 6 个人工数据集上的平均得分。与所有基线相比, DHM 取得了最佳的结果, 这得益于 DHM 在利用时序模型捕获网络演化的同时, 高度保留了网络的空间结构。

在传统基线中, Pisces 在 GN 基准上的表现最好, 这是因为 Pisces 在谱聚类的基础上引入了特征向量平滑, 以促进历史和未来结果的结合。FacetNet 在 Greene 基准上表现最好, 这是因为 FacetNet 关注相邻快照之间的差异, 能够适应社区的演化事件。Dyperm 的综合表现最差, 这是因为 Dyperm 关注局部的拓扑变化, 而在整体网络结构的演化中可能忽略了全局的关系调整, 导致对网络整体动态性的捕捉不够全面。

在表征学习基线中, DGCN 表现最好, 这是因为 DGCN 在局部的特征传播和信息聚合的基础上, 结合了互信息最大化来强化节点之间的相关性。在基于自编码器架构的模型中, DynAERNN 表现最好, 这是因为 DynAERNN 结合了自编码器和神经网络的优势, 同时捕捉了网络的空间和时间信息。相比之下, DynAE 和 DynRNN 的表现甚至不如基于矩阵分解的 TIMERS, 原因在于它单方面地关注空间或时间信息, 不能准确建模网络的演化特征。此外, 尽管 DGCN 在一定程度上捕捉了节点之间的远程依赖关系, 但在各个数据集上表现出比 DHM 相对较差的性能。

DHM 通过将紧密相连的独立节点合并为层次结构来扩大原始网络中节点的感受野, 然后通过构造层次损失来显式地融合节点之间的远程信息, 对于复杂的时空交互具有更强的建模能力。

2) 在现实数据集上的结果

表 4 列出了 DHM 以及所有基线在 8 个小规模现实数据集的平均得分。与所有基线相比, DHM 在这些数据集上均表现出更为优异的结果。这体现了 DHM 具有有效捕捉真实网络演化特征的能力, 同时高度保留了现实网络节点之间的依赖关系。

(续表)

| Method | InVS15 | | Primary | | High School | | Enron | SFHH | Email-EU | Voles | fb |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|-------|--------------|--------------|--------------|--------------|
| | NMI | ARI | NMI | ARI | NMI | ARI | Q | Q | Q | Q | Q |
| Dyperm | 0.459 | 0.029 | 0.530 | 0.230 | 0.706 | 0.540 | 0.139 | 0.058 | 0.040 | 0.118 | 0.182 |
| TIMERS | 0.415 | 0.140 | 0.775 | 0.659 | 0.524 | 0.226 | 0.283 | 0.209 | 0.284 | 0.449 | 0.086 |
| DynAE | 0.394 | 0.141 | 0.767 | 0.642 | 0.557 | 0.307 | 0.284 | 0.074 | 0.216 | 0.062 | 0.096 |
| DynAERNN | 0.357 | 0.129 | 0.795 | 0.683 | 0.705 | 0.550 | 0.416 | 0.069 | 0.340 | 0.091 | 0.121 |
| DynRNN | 0.141 | 0.016 | 0.277 | 0.106 | 0.529 | 0.357 | 0.086 | 0.017 | 0.218 | 0.050 | 0.107 |
| EvolveGCN | 0.415 | 0.151 | 0.794 | 0.697 | 0.618 | 0.362 | 0.510 | 0.294 | 0.480 | 0.597 | 0.267 |
| DGCN | 0.316 | 0.117 | 0.733 | 0.576 | 0.689 | 0.593 | 0.605 | 0.138 | 0.536 | 0.516 | 0.131 |
| DHM | 0.503 | 0.311 | 0.843 | 0.746 | 0.771 | 0.679 | 0.583 | 0.362 | 0.574 | 0.626 | 0.348 |

在传统基线中, FacetNet 的综合表现最好, 而 Pisces 和 ECD 由于时空复杂度过高, 无法求解动态社区检测问题。ECD 在 Enron 数据上依旧超过了所有方法, 这证明了传统演化算法的有效性。在嵌入基线中, EvolveGCN 的综合表现最好。TIMERS 的整体表现与 DynAERNN 相差无几, 但与 EvolveGCN 存在一定差距, 突显了 GNN 模型在编码拓扑结构方面的有效性。此外, 与 DHM 相比, EvolveGCN 的局部嵌入过程忽略了节点的长程信息。DHM 通过联合模块度最大化和层次结构嵌入, 从全局角度约束了节点的交互, 使得生成的嵌入能更好地保留网络的社区特性。

5.6 消融实验

为了更深入地研究层次结构嵌入模块的性能, 本文定义了以下几种变体进行验证。

- 1)HM: 从 DHM 中移除编码器中的 LSTM 模块。
- 2)DM: 从 DHM 中移除层次结构嵌入模块。
- 3)DH: 从 DHM 中移除编码器模块和解码器模块。

表 5 列出了消融实验的结果, 其中 DHM 和 DM 的表现差异证明了层次嵌入模块对最终性能产生了积极影响。此外, 3 个变体中 DM 的表现最好, 结合 DHM 和 DH 的表现差异来看, 层次嵌入模块能够辅助模块度的优化目标, 共同致力于提高节点表征的判别性。值得一提的是, 即便忽略了网络快照之间的时序关联, HM 也能够获得不错的结果, 能够适用于静态社区检测任务。

表 5 DHM 在 Expand contract 和 Email-EU 上的消融结果

Table 5 Ablation results of DHM on Expand contract and Email-EU

| Dataset | Metric | HM | DM | DH | DHM |
|-----------------|--------|-------|-------|-------|--------------|
| Expand contract | NMI | 0.961 | 0.982 | 0.974 | 0.998 |
| | ARI | 0.915 | 0.950 | 0.932 | 0.992 |
| Email-EU | Q | 0.473 | 0.498 | 0.486 | 0.574 |

5.7 层间消息传递策略分析

DHM 设计了自底向上和自顶向下的层间消息通道, 以便为节点叠加更丰富的结构信息。因此, 本节旨在探讨不同的层间消息传递策略对模型表现的影响。具体而言, 本节选取了平均池化 (Mean pooling, MP)、图卷积 (GCN) 以及自注意力 (Self-attention, SA) 3 种策略, 并选择在 SYNFIX 和 Enron 两个数据集上进行验证。

表 6 列出了在不同的层间消息传递策略下模型的表现结果, 从中可以观察到, DHM 的最佳策略组合是平均池化和自注意力的结合。在自底向上的过程中, 不同层间的尺度信息被连续地叠加至顶层。在这种情况下, 平均池化的方法能够保持相对稳定, 不会因超级节点对应的节点集合大小而产生较大的波动。而在自顶向下的过程中, 不同层间的尺度信息

被分别叠加至底层。在这种情况下, 自注意力机制更能够考虑到不同尺度之间的差异, 从而更适合此时的信息整合。

表 6 层间消息传递策略分析

Table 6 Analysis of inter-layer message passing strategy

| from bottom to top | from top to bottom | SYNFIX | | Enron |
|--------------------|--------------------|--------------|--------------|--------------|
| | | NMI | ARI | Q |
| MP | GCN | 0.482 | 0.443 | 0.364 |
| | MP | 0.981 | 0.976 | 0.576 |
| | SA | 1.000 | 1.000 | 0.583 |
| GCN | MP | 0.885 | 0.834 | 0.534 |
| | GCN | 0.971 | 0.956 | 0.572 |
| | SA | 0.996 | 0.997 | 0.566 |
| SA | MP | 0.798 | 0.794 | 0.471 |
| | GCN | 0.991 | 0.990 | 0.569 |
| | SA | 0.995 | 0.993 | 0.573 |

结束语 本文提出了一种基于表征学习的动态网络社区检测模型 DHM。该模型联合模块度优化和层次结构嵌入技术, 通过自底向上和自顶向下的消息传递过程, 为节点嵌入融合了长距离信息。在多个人工和现实网络数据集上的实验结果验证了所提模型的有效性。通过分析 DHM 在不同规模网络下的检测结果发现, 节点的长距离信息对大规模网络的检测性能有较大帮助, 这项研究有助于更好地理解和分析真实世界网络的演化社区。

本文仅针对非重叠社区场景下的动态社区检测问题, 而现实世界中的节点通常同时属于多个社区。如何设计一种兼顾非重叠和重叠社区的动态检测方法是未来的研究方向。

参考文献

- [1] DOURISBOURE Y, GERACI F, PELLEGRINI M. Extraction and classification of dense communities in the web[C]// Proceedings of the 16th International Conference on World Wide Web. 2007:461-470.
- [2] LIN Y R, CHI Y, ZHU S, et al. Analyzing communities and their evolutions in dynamic social networks[J]. ACM Transactions on Knowledge Discovery from Data, 2009, 3(2): 1-31.
- [3] SINGHC K, JOLAD S. Structure and evolution of Indian physics co-authorship networks[J]. Scientometrics, 2019, 118(2): 385-406.
- [4] ROSSETTI G, CAZABET R. Community discovery in dynamic networks: a survey[J]. ACM Computing Surveys, 2018, 51(2): 1-37.
- [5] MA R X, DENG G S. Research of Dynamic Community Discovery Based on Role Assorted Thoughts[J]. Chinese Journal of Computer Science, 2012, 39(9): 60-63, 70.
- [6] WANG B B, XIN J C, CHEN J Y, et al. Motif-aware Adaptive

- Cross-layer Random Walk Community Detection[J]. Chinese Journal of Computer Science, 2024, 51(6): 128-134.
- [7] CHEN J, YUAN B. Detecting functional modules in the yeast protein-protein interaction network[J]. *Bioinformatics*, 2006, 22(18): 2283-2290.
- [8] DU N, WU B, PEI X, et al. Community detection in large-scale social networks[C]// *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*. 2007: 16-25.
- [9] PAN Y, WANG S H, ZHANG L, et al. Survey of Community Detection in Complex Network[J]. *Chinese Journal of Computer Science*, 2022, 49(S2): 208-218.
- [10] HOPCROFT J, KHAN O, KULIS B, et al. Tracking evolving communities in large linked networks[J]. *Proceedings of the National Academy of Sciences*, 2004, 101(suppl_1): 5249-5253.
- [11] LIT, WANG W, JIAO P, et al. Exploring temporal community structure via network embedding[J]. *IEEE Transactions on Cybernetics*, 2023, 53(11): 7021-7033.
- [12] GUIMERAR, MOSSA S, TURTSCHI A, et al. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles[J]. *Proceedings of the National Academy of Sciences*, 2005, 102(22): 7794-7799.
- [13] LUKASOVÁ A. Hierarchical agglomerative clustering procedure[J]. *Pattern Recognition*, 1979, 11(5/6): 365-381.
- [14] TAJEUNAE G, BOUGUENSA M, WANG S. Tracking the evolution of community structures in time-evolving social networks[C]// *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2015: 1-10.
- [15] LEVANDOWSKYM, WINTER D. Distance between sets[J]. *Nature*, 1971, 234(5323): 34-35.
- [16] JIAT, CAI C, LI X, et al. Dynamical community detection and spatiotemporal analysis in multilayer spatial interaction networks using trajectory data[J]. *International Journal of Geographical Information Science*, 2022, 36(9): 1719-1740.
- [17] TRAAGV A, WALTMAN L, VAN ECK N J. From Louvain to Leiden: guaranteeing well-connected communities[J]. *Scientific Reports*, 2019, 9(1): 5233.
- [18] NGUYEN P, DINH T N, XUAN Y, et al. Adaptive algorithms for detecting community structure in dynamic social networks[C]// *2011 Proceedings IEEE INFOCOM*. IEEE, 2011: 2282-2290.
- [19] ZHUANG D, CHANG J M, LI M. DynaMo: Dynamic community detection by incrementally maximizing modularity[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 33(5): 1934-1945.
- [20] LONGH, LI X, LIU X W, et al. BBTA: Detecting communities incrementally from dynamic networks based on tracking of backbones and bridges[J]. *Applied Intelligence*, 2023, 53(1): 1084-1100.
- [21] LIN Y, CHI Y, ZHU S, et al. Facetnet: a framework for analyzing communities and their evolutions in dynamic networks[C]// *Proceedings of the 17th International Conference on World Wide Web*. 2008: 685-694.
- [22] FU L D, NIE J J. Dynamic Community Detection Based on Evolutionary Spectral Method[J]. *Chinese Journal of Computer Science*, 2018, 45(2): 171-174.
- [23] LIUF, CHOI D, XIE L, et al. Global spectral clustering in dynamic networks[J]. *Proceedings of the National Academy of Sciences*, 2018, 115(5): 927-932.
- [24] MA X, DONG D. Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(5): 1045-1058.
- [25] LI D, LIN Q, MA X. Identification of dynamic community in temporal network via joint learning graph representation and nonnegative matrix factorization[J]. *Neurocomputing*, 2021, 435: 77-90.
- [26] LI D, MA X, GONG M. Joint learning of feature extraction and clustering for large-scale temporal networks[J]. *IEEE Transactions on Cybernetics*, 2023, 53(3): 1653-1666.
- [27] GOYA LP, KAMRA N, HE X, et al. Dyngem: Deep embedding method for dynamic graphs[J]. *arXiv*: 1805.11273, 2018.
- [28] GOYAL P, CHHETRI S R, CANEDO A. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning[J]. *Knowledge-Based Systems*, 2020, 187: 104816.
- [29] PU S, ZHAO W D. Community Detection Algorithm for Dynamic Academic Network[J]. *Chinese Journal of Computer Science*, 2022, 49(1): 89-94.
- [30] PAREJA A, DOMENICONI G, CHEN J, et al. Evolvegen: Evolving graph convolutional networks for dynamic graphs[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020: 5363-5370.
- [31] GAO C, ZHU J, ZHANG F, et al. A novel representation learning for dynamic graphs based on graph convolutional networks[J]. *IEEE Transactions on Cybernetics*, 2023, 53(6): 3599-3612.
- [32] LIU H, HE L, ZHANG F, et al. Dynamic community detection over evolving networks based on the optimized deep graph infomax[J]. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2022, 32(5): 053119.
- [33] KIPF T N, WELLMING M. Variational graph auto-encoders[C]// *Neural Information Processing Systems Workshop on Bayesian Deep Learning*. 2016.
- [34] ECKART C, YOUNG G. The approximation of one matrix by another of lower rank[J]. *Psychometrika*, 1936, 1: 211-218.
- [35] QIU C, HUANG Z, XU W, et al. VGAER: graph neural network reconstruction based community detection[J]. *arXiv*: 2201.04066, 2022.
- [36] ROSSETTI G. RDyn: graph benchmark handling community dynamics[J]. *Journal of Complex Networks*, 2017, 5(6): 893-912.
- [37] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. *Physical Review E*, 2004, 69(2): 026113.
- [38] GREENE D, DOYLE D, CUNNINGHAM P. Tracking the evolution of communities in dynamic social networks[C]// *2010 International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2010: 176-183.
- [39] GÉNOIS M, BARRAT A. Can co-location be used as a proxy for

face-to-face contacts? [J]. EPJ Data Science, 2018, 7(1): 1-18.

- [40] SUN J, FALOUTSOS C, PAPANITRIOU S, et al. Graphscope: parameter-free mining of large time-evolving graphs [C]// Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2007: 687-696.
- [41] PARANJAPEA, BENSON A R, LESKOVEC J. Motifs in temporal networks [C]// Proceedings of the 10th ACM International Conference on Web Search and Data Mining. 2017: 601-610.
- [42] ROSSI R, AHMED N. The network data repository with interactive graph analytics and visualization [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2015.
- [43] LIUF, WU J, ZHOU C, et al. Evolutionary community detection in dynamic social networks [C]// 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019: 1-7.
- [44] AGARWALP, VERMA R, AGARWAL A, et al. DyPerm: Maximizing permanence for dynamic community detection [C]// Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer, 2018: 437-449.
- [45] ZHANG Z, CUI P, PEI J, et al. Timers: Error-bounded svd restart on dynamic networks [C]// Proceedings of the AAAI Con-

ference on Artificial Intelligence. 2018.

- [46] ESTÉVEZ P A, TESMER M, PEREZ C A, et al. Normalized mutual information feature selection [J]. IEEE Transactions on Neural Networks, 2009, 20(2): 189-201.
- [47] HUBERT L, ARABIE P. Comparing partitions [J]. Journal of Classification, 1985, 2: 193-218.



ZHU Rui, born in 2000, postgraduate. Her main research interests include spatial-temporal data mining and network analysis.



YE Yaqin, born in 1979, Ph.D, associate professor, postgraduate supervisor, is a member of CCF (No. 42951M). Her main research interests include intelligent computing of urban multimodal trajectories and spatial reasoning.

(责任编辑:喻黎)

【征集通知】第六期“CCF 科普阅读推荐图书”

“CCF 科普阅读推荐图书”征集活动以普及计算机知识、弘扬科学精神为核心,旨在拓宽公众视野、推动计算机知识普及,满足人民群众日益增长的知识需求,提升全社会科学文化素养。

诚邀各位科普作者和图书组织编写单位对符合要求的图书作品进行申报。

图书精品推荐结果将于 2025 年 9 月下旬 CCF 科普月活动闭幕式上正式发布,并进行 CCF 科普图书集中展示。入选图书,推荐参与“科普月优秀图书分享活动”,择优向中国科协推荐典赞·2025 科普中国——科普作品奖项,并成为 CCF 科普基地、CCF 走进中小学选用的科普图书。

一、申请条件和要求

1) 图书作品征集范围为国内近两年公开出版发行的科普图书,包括著作、编选作品、翻译作品、编译作品、画册;计算机类教材、职业教育类图书。

2) 图书作品应符合国家《出版管理条例》及《图书质量管理条例》所规定的相关要求;知识产权清晰,符合著作权法的有关规定;具有普及计算机知识、倡导计算思维、传播科学思想、弘扬科学精神的深厚内涵,并有较高的科学性和趣味性,语言生动流畅、形式活泼、内容丰富、通俗易懂,为公众所喜闻乐见;反映时代要求,具有较高的社会认同度。

二、申请方式和时间

即日起至 2025 年 8 月 20 日(星期三)24:00 前,图书作者或图书组织编写单位下载并填写《CCF 科普阅读推荐图书申请表》,并将扫描件发送至邮箱 pop_sci@ccf.org.cn。

纸质版申请表和样书(2 本)请邮寄至如下地址:

北京市海淀区中关村科学院南路 6 号中国计算机学会 王颖 18515230996

三、联系方式

CCF 秘书处

联系人:王颖

联系电话:010-62562503-24,18515230996

CCF 科学普及工作委员会

联系人:李林晓

联系电话:16619927190



(扫码下载申请表)