

结合拓扑势与信任度调整的重叠社区发现算法

李晓红,王闪闪,周学铭,宿云

(西北师范大学 计算机科学与工程学院,兰州 730070)

摘要:现实世界中的复杂系统可建模为复杂网络,探究复杂网络中的社区发现算法对于分析复杂网络的拓扑结构和层次结构具有重要作用。早期研究通常将网络中的节点局限在一个社区中,但随着研究的深入发现社区结构呈现重叠特性。针对现有重叠社区发现算法存在划分社区结构不稳定、忽略节点交互和属性等问题,提出一种基于网络拓扑势与信任度调整的重叠社区发现算法。融合节点的属性和结构特征计算节点的拓扑势,依据节点的拓扑势选取核心节点。从核心节点出发构建初始社区群,计算各个社区间的调整信任度,实现社区的合并与再调整,从而识别重叠社区。在多个人工模拟网络和真实网络数据集上的实验结果表明,与基于贪婪派系扩张、种子扩张等的重叠社区发现算法相比,该算法将扩展模块度最高提升至0.719,能有效识别社区结构及重叠节点,提升重叠社区检测性能。

关键词:重叠社区发现;节点属性;拓扑势;核心节点;信任度调整

开放科学(资源服务)标志码(OSID):



中文引用格式:李晓红,王闪闪,周学铭,等.结合拓扑势与信任度调整的重叠社区发现算法[J].计算机工程,2022,48(5):53-58.

英文引用格式:LI X H, WANG S S, ZHOU X M, et al. Overlapping community discovery algorithm combining topology potential and trust adjustment[J]. Computer Engineering, 2022, 48(5): 53-58.

Overlapping Community Discovery Algorithm Combining Topology Potential and Trust Adjustment

LI Xiaohong, WANG Shanshan, ZHOU Xueming, SU Yun

(School of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China)

[Abstract] All types of complex systems in the real world can be modeled as complex networks. The study of community discovery algorithms in complex networks plays an important role in analyzing the topology and hierarchy of complex networks. Early studies limited the nodes in the network to one community. With more extensive research, it was determined that community structures also display overlapping characteristics. This paper focuses on the defects of existing algorithms, which consist of an unstable structure of divided communities and ignore the node interaction and attributes. In this study, the proposed overlapping community discovery algorithm combines topology potential and trust adjustment. The topological potential is first computed using the attribute value and structural characteristics, and the core node is selected based on the topological potential. Next, starting from the core node, initial communities centered on the core node are formed. The communities are subsequently merged and readjusted by calculating the adjustment trust between each community to enable the detection of overlapping communities. The experimental results on several artificial simulated networks and real network datasets show that compared to overlapping community discovery algorithms based on Greedy Factional Expansion (GFE), Two Expansions of Seeds (TES), etc., the proposed algorithm increases the expansion module degree by up to 0.719, which allows effective identification of community structure and overlapping nodes, improving overlapping community detection performance.

[Key words] overlapping community discovery; node attribute; topology potential; core node; trust adjustment

DOI: 10.19678/j.issn.1000-3428.0060725

基金项目:国家自然科学基金(61862058, 61967013);甘肃省高等学校创新创业基金(2020B-089);甘肃省科技计划项目(20JR5RA518);甘肃省自然科学基金(20JR10RA076)。

作者简介:李晓红(1978—),女,副教授、硕士,主研方向为数据挖掘、机器学习、智能信息处理;王闪闪,硕士研究生;周学铭,学士;宿云,副教授、博士。

收稿日期:2021-01-27 **修回日期:**2021-05-08 **E-mail:** xiaohongli@nwnu.edu.cn

0 概述

复杂网络是指具有自组织、自相似、吸引子、小世界、无标度特征中部分或全部性质的网络,如现实生活中的交通网、社交网、基因调控网等。社区是指具有相同或相似特性的节点构成的集合,同一社区内部节点之间链接的概率比不同社区间节点的链接概率高得多,这反映了复杂网络中个体行为的局部特征及其相互之间的关联关系。对于社区发现的研究有助于揭示网络重要结构和功能信息,在过去的几十年中,研究人员设计了大量的社区发现算法,但随着研究的深入发现社区结构还表现出重叠特性,例如社交网络中有一些节点同时归属于不同社区,且这种同时属于多个社区的节点又是信息传播和网络演变中的关键节点。因此,区分稳定的重叠社区并设计高效的发现算法是亟需解决的问题,受到研究人员的广泛关注。

目前,重叠社区发现算法主要包括基于派系过滤、基于局部扩张及优化、基于线图或边社区3类。基于派系过滤的重叠社区发现算法^[1-2]是一种基于网络拓扑结构的社区发现算法,从网络特性出发,认为社区由多个K派系组成,每个派系都是一个全连通网络,网络中的每一个节点可以划分到不同的K派系中。EVANS^[3]提出的Clique Graph借鉴了派系过滤的思想,通过在网络中建立派系图来研究社区结构。卢志刚等^[4]提出一种基于贪婪派系扩张(Greedy Fractional Expansion, GFE)的重叠社区发现算法。该算法根据企业社会化网络中极大派系间的链接强度将原始网络图转换成最大派系图,在最大化适应度函数的条件下贪婪扩张最大派系图中的种子派系进行社区发现。但是,基于派系过滤的重叠社区发现算法受限于网络中缺失完全子图,自由参数较多,不同参数设定对结果影响较大,并且经常产生较大的时间复杂度。WEN等^[5]提出一种基于最大团的多目标进化算法(MOEA)检测重叠社区。该算法用原始图的一组最大团作为节点定义,两个最大团可共享原始图的相同节点。MOEA以类似于非重叠社区检测的方式处理重叠社区检测问题。基于局部扩张及优化的重叠社区发现算法通常将种子节点作为初始社区,通过不断优化质量函数扩展社区,最终得出社区划分结果^[6-7]。代表算法为局部扩展算法(Local Fitness Method, LFM)^[8],LFM随机选取一个节点进行扩展,通过迭代局部社区得到重叠社区。CHANG等^[9]提出一种新的重叠社区发现算法ENFI,该算法利用网络的微观特征,通过计算朋友亲密度提取本地社区并形成网络的重叠社区。WILDER等^[10]用随机游走算法为每个节点找出一个子图,同时计算出该节点的权重以及正比于权重的概率,并以此判定初始节点,该算法的时间复杂度较高。但是以上算法具有不确定性和向外扩展性,形成的网络模块容易存在不稳定性 and 漂移性。基于线图或边社区的重叠社区发现算法计算过程复杂,

不适用于小规模边社区。AHN等^[11]根据链接网络非重叠与重叠的转化思想,对网络中的链接进行层次聚类提出LINK算法,该算法揭示了这种边社区在真实网络中的普遍存在性。除了以上3类重叠社区发现算法之外,研究人员还提出了一些其他的改进算法^[12-14],均取得了不错的效果。

本文借鉴物理学势能中的拓扑势思想,提出一种基于拓扑势与信任度调整的重叠社区发现算法。由于网络中的每个节点周围存在一个作用场,场中的任何节点都将受到其他节点的联合作用,因此利用拓扑势选取核心节点,然后从核心节点出发构建初始社区群,并且充分利用社区间的共享边越多社区越相容、节点间频繁的交互和联系促使社区发生融合这两个特性,通过调整信任度进行社区合并与调整。

1 相关工作

假设复杂网络表示为有向图 $G=(V, E, A)$,其中:顶点集合 $V=\{v_1, v_2, \dots, v_i, \dots, v_n\}$, n 表示节点数; $A_i=\{a_{i1}, a_{i2}, \dots, a_{im}\}$ 表示节点 v_i 的属性集合, $m=|A|$ 表示网络属性数。

1.1 改进的节点拓扑势

在物理学中,场表示物体间的相互作用,这种相互作用并不是接触才有的,势表示特定场中的质点从一点移动到另一点时产生的功。势和场的分布对应物质粒子之间由位置确定的势能分布^[15-16],并且这种势能分布与粒子之间的距离成递减关系,随着距离的增长势能下降,直至衰减为0。类比于势场,网络中的每个节点周围存在着一个作用场,位于场中的任何节点都将受到其他节点的联合作用。通过路径描述节点之间的联系,利用路径长度描述节点间相互作用力的强弱。因此,节点间的相互作用与节点属性及节点间的距离密切相关,并且每个节点的影响力会随着节点间路径长度的增长而衰减。采用高斯势函数描述这种相互作用,将复杂网络中节点 v_i 的拓扑势定义为某个范围内其他节点在该节点处产生能量的累加和,如式(1)所示:

$$\varphi(v_i) = \sum_{j \in \Gamma(v_i)} m(v_j) \times e^{-\left(\frac{d(v_i, v_j)}{\sigma}\right)^2} \quad (1)$$

其中: $\Gamma(v_i)$ 表示节点 v_i 影响某个范围内的节点集合;常数 $\sigma \in (0, +\infty)$ 用来控制节点的作用范围; $d(v_i, v_j)$ 表示节点 v_j 到节点 v_i 的最短路径长度($i \neq j$)。

首先,为了突出各个节点固有属性的差异, $m(v_j)$ 的值由节点属性值决定^[17]。假定节点 v_i 的属性值为 $\{w_{i1}, w_{i2}, \dots, w_{ij}, \dots, w_{im}\}$,则节点 v_i 的属性重要性表示为 $m(v_i) = \sum_{j=1}^m w_{ij}$ 。其次,由于节点在网络中所处的位置及其链接关系存在差异,使得各个节点在网络中具有不同的重要性。为了突出这种差异,引入位置权重 $w_{\text{weight}}(v_j)$,表示距离核心节点越远,位置权重越小。

$w_{\text{weight}}(v_j)$ 计算公式如下:

$$w_{\text{weight}}(v_j) = \frac{|\Gamma(v_i)|}{d(v_i, v_j)} \quad (2)$$

同时,假设式(1)的拓扑势满足规范化条件,则 $\forall v_i \in V$ 的拓扑势可重写为:

$$\varphi(v_i) = \frac{1}{M} \sum_{j \in \Gamma(v_i)} w_{\text{weight}}(j) \times m(v_j) \times e^{-\left(\frac{d(v_i, v_j)}{\sigma}\right)^2} \quad (3)$$

其中: $M = \sum_{v_i \in V} \varphi(v_i)$ 。在 $d(v_i, v_j)$ 确定的情况下,节点分布越稀疏, $\varphi(v)$ 越小,反之,节点分布越密集, $\varphi(v)$ 越大。因此,拓扑势间接地反映了网络密集程度。

1.2 信任值调整

从网络结构的角度来看,两个节点 v_i 与 v_j 之间拥有的公共邻居越多,它们越相似。扩展至社区,同样地,如果两个社区共享边较多,就意味着社区间重叠的部分越多,社区就越相容。用于比较有限集合之间的相似性与差异性的 Jaccard 系数^[18]可描述这种结构上的重叠程度。除此之外,网络中的节点不是独立存在的,它们之间具有不同亲疏的交互和联系。社区间的信任解析为彼此所含节点之间的认同,而这种认同实质上是一个互动问题,互动会改变它们之间的关系,使它们发生交叉甚至融合,因此,采用节点间相似度来体现这种互动。综上,本文基于社区间的共享边越多社区越相容及节点间频繁的交互和联系会促使社区发生融合这两个特性判断两个社区能否合并,称此过程为调整信任度。调整信任度的计算公式如下:

$$A_{\text{AT}}(C_i, C_j) = \lambda N_{\text{NE}}(C_i, C_j) + (1 - \lambda) T(C_i, C_j) \quad (4)$$

其中: λ 为调节参数 ($0 < \lambda < 1$), 用于控制分析社区间共享边数和社区间信任度的占比; $N_{\text{NE}}(C_i, C_j)$ 表示两社区共享的边数; $T(C_i, C_j)$ 表示社区 C_i 与 C_j 的信任度^[19], 为两社区内所有节点间信任度的和。 $N_{\text{NE}}(C_i, C_j)$ 和 $T(C_i, C_j)$ 的计算公式如下:

$$N_{\text{NE}}(C_i, C_j) = \frac{\sum_{v_i \in C_i, v_j \in C_j} |N(v_i) \cap N(v_j)|}{|N(C_i \cap C_j)|} \quad (5)$$

$$T(C_i, C_j) = \sum_{v_i \in C_i, v_j \in C_j} T(v_i, v_j) \quad (6)$$

其中: $N(v_i)$ 为节点 v_i 依附的边构成的集合; $N(C_i, C_j)$ 表示社区 C_i 与 C_j 合并后所包含的边。

节点间的信任度采用基于节点属性的相似度来度量,计算公式如下:

$$T(v_i, v_j) = \text{sim}(v_i, v_j) = \frac{\mathbf{a}_i \cdot \mathbf{a}_j}{\|\mathbf{a}_i\| \cdot \|\mathbf{a}_j\|} \quad (7)$$

其中: \mathbf{a}_i 、 \mathbf{a}_j 分别为节点 v_i 、 v_j 的属性向量。

2 重叠社区发现算法

本文提出的基于节点拓扑势和信任值调整的重叠社区发现算法本质上是发现核心节点并以该核心

节点为中心进行扩展形成社区的过程。为方便描述,将本文算法简称为 CTPT 算法。CTPT 算法流程如图 1 所示,其中 q 表示最终合并社区的个数。

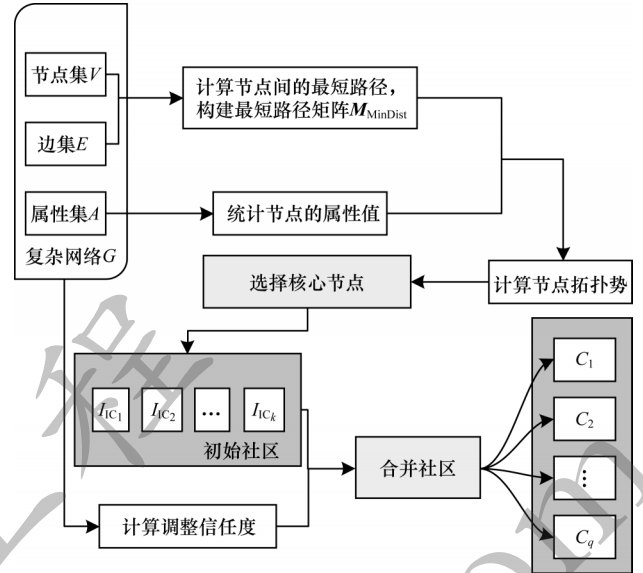


图1 CTPT 算法流程

Fig.1 Procedure of CTPT algorithm

2.1 核心节点选取

首先,计算网络 G 中任意一对节点 v_i 和 v_j 的最短路径长度 $d(v_i, v_j)$, 构建最短路径矩阵 M_{MinDist} 。然后,利用 M_{MinDist} 提供的路径信息,生成每个节点 v_i 的作用节点集合 $\Gamma(v_i)$ 。接下来,按照式(3)迭代地计算每个节点的拓扑势 $\varphi(v_i)$, 并以此为判断依据选取 K 个 $\varphi(v_i)$ 最大的节点作为核心节点,保存在 Top 数组中。核心节点选取流程如图 2 所示。

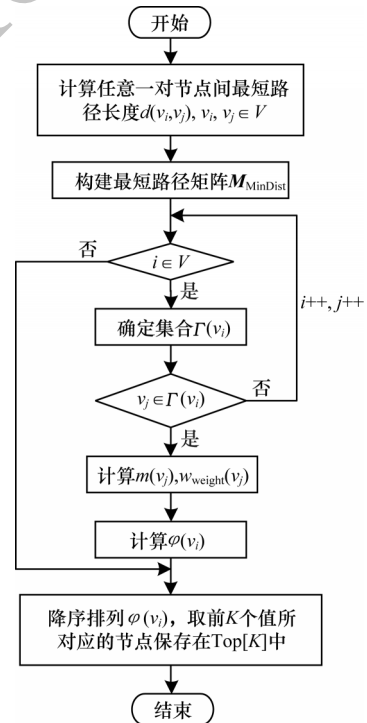


图2 核心节点选取流程

Fig.2 Procedure of core node selection

2.2 社区形成与再调整

在初始化时,将 $\text{Top}[K]$ 中的每一个核心节点视为一个社区。首先,从单节点社区出发,依次取出 V 中的节点 v_i 并尝试加入社区 I_{IC_k} ,计算模块度函数 $Q^{[20]}$ 。如果 Q 增大,则将该节点加入社区,否则尝试将节点 v_i 加入下一个社区 $I_{IC_{k+1}}$,其中, $Q = \sum (e_{mm} - a_m^2)$, e_{mm} 表示社区 m 内部所有节点之间的边的集合, a_m 表示所有连接到社区 m 的边数, $a_m = \sum_n e_{mn}$, e_{mn} 表示社区 m 和社区 n 之间的边数。然后,通过调整信任度对已形成的初始社区群落进行合并与再调整,上述步骤完成后形成最终社区。未加入任何社区的节点被视为离群点,并将其剔除。社区形成与合并流程如图3所示。

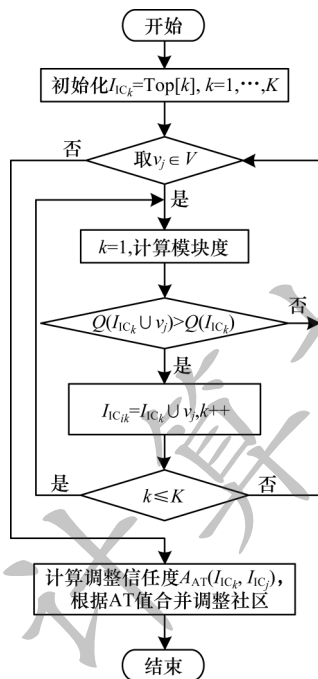


图3 社区发现与合并流程

Fig.3 Procedure of community discovery and merging

3 实验结果与分析

3.1 实验设置

3.1.1 数据集

为评价CTPT算法性能,实验采用两类数据集:1)LFR人工网络数据集,由经典的LFR基准程序生成,可以模拟具有重叠社区结构的网络数据集,主要用于测试社区检测算法的优劣;2)真实网络数据集,来源于斯坦福大学的大型网络数据集网站(<http://memetracker.org/data/index.html>)。表1给出了人工网络数据集Net1、Net2和Net3的参数信息,其中, O_n 表示重叠节点数, μ 表示LFR网络中的混合系数,其值越大表明所生成的网络拓扑结构越复杂。表2给出了真实网络数据集ego-Facebook、Musae-twitch和Web-flickr的参数信息。ego-Facebook数据集由facebook的朋友列表组成,是facebook上用户间的社交网络。Musae-twitch数据集是使用某种语言进行流式传输的游戏玩家的Twitch用户-用户网络,该网络中的节点是用户本身,链接是他们之间的好友关

系。Web-flickr是用户分享图片和视频的社交网络,该网络中每一个节点都是flickr中的用户,每一条边都是用户之间的好友关系。另外,每一个节点都有标签,用于标识用户的兴趣小组。

表1 LFR人工网络数据集参数设置

Table 1 Parameter setting of LFR artificial network dataset

网络	顶点数	平均度	边数	最小社区数	最大社区数	O_n	μ
Net1	3 000	25.25	37 883	30	60	300	0.3
Net2	12 000	43.40	260 527	50	100	1 000	0.3
Net3	50 000	37.90	949 603	100	200	4 000	0.3

表2 真实网络数据集参数设置

Table 2 Parameter setting of real network dataset

网络	顶点数	平均度	边数
ego-Facebook	4 039	43.7	88 234
Musae-twitch	34 118	25.1	429 113
Web-flickr	80 513	73.3	5 899 882

为评价算法性能,通常需要选取度量社区划分好坏的性能指标。扩展模块度是最常用的指标之一,计算公式如下:

$$E_{EQ} = \frac{1}{2m} \sum_k \sum_{v_i \in C_k, v_j \in C_k} \frac{1}{C_k} \left[M - \frac{d_{v_i} d_{v_j}}{2m} \right] \quad (8)$$

其中: d_{v_i} 是节点 v_i 的度。EQ值越大,表明社区结构上存在的重叠越明显,当所有节点都属于同一个社区时EQ值等于0。

指标归一化互信息(Normalized Mutual Information, NMI)^[21]用于度量检测到的社区与真值之间的距离,计算公式如下:

$$N_{NMI}(X|Y) = 1 - \frac{1}{2} [H(X|Y) + H(Y|X)] \quad (9)$$

其中: $H(X|Y)$ 表示 X 对 Y 的规范条件熵; $H(Y|X)$ 表示 Y 对 X 的规范条件熵,该值越大,意味着算法所发现社区与网络真实社区相吻合的程度越高。

3.1.2 对比算法

将本文提出的CTPT算法与以下4种算法进行对比:

1)基于隶属度传播(Membership-Degree Propagation, MDP)的重叠社区划分算法^[22]:以种子节点的基本特征为依据构建网络节点之间的隶属度传播模型,将种子节点的社团隶属度传播至非种子节点进行社区发现。

2)面向属性网络的重叠社区划分算法(overlapping community discovery Algorithm for Attributed Networks, ANA)^[23]:利用节点的密集度和间隔度搜索局部密度中心,并将其作为社区中心,通过计算非中心节点的社区隶属度实现重叠社区的划分。

3)基于GFE的重叠社区划分算法^[4]:根据网络中极大派系间的链接强度将原始网络图转换成最大派系图,在最大化适应度函数的条件下,贪婪扩张最大派系图中的种子派系进行社区发现。

4)基于种子扩张(Two Expansions of Seeds, TES)的重叠社区划分算法^[24]:首先根据网络节点的拓扑特征形成初始社区,然后采用重力度再次进行社区的合

并和扩展。

3.2 结果分析

3.2.1 LFR 人工生成网络数据集上的结果分析

图 4 给出了在不同规模的 LFR 人工生成网络数据集上 5 种重叠社区发现算法随重叠社区数量变化对应 NMI 的变化规律。CTPT 算法考虑不同结构和属性的影响,分别设置调整信任度的调节参数 λ 为 0.4、0.5、0.6,实验结果显示 $\lambda = 0.5$ 时效果最好。

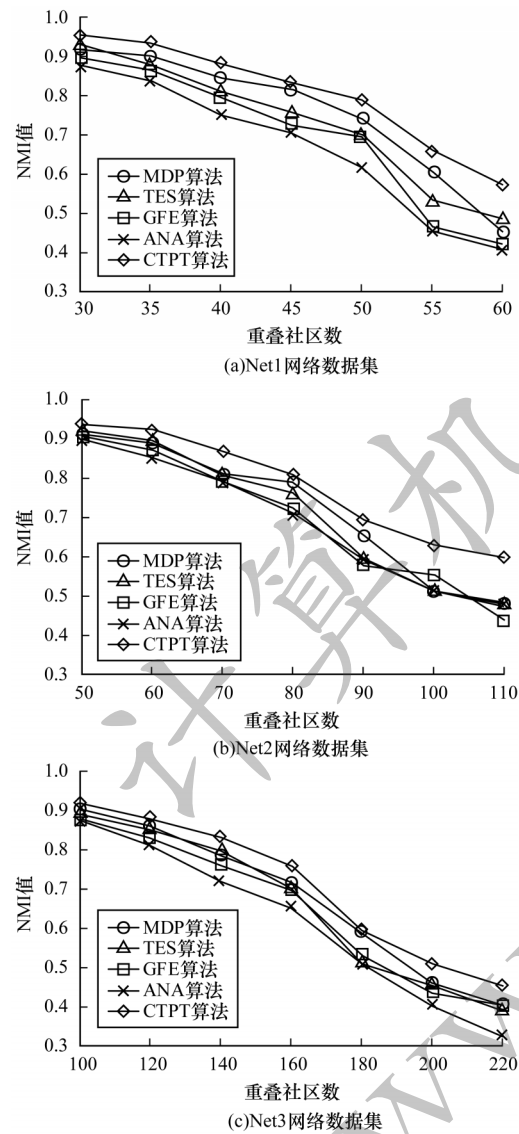


图 4 不同重叠社区数量对算法 NMI 值的影响
Fig.4 Influence of the number of different overlapping communities on the NMI value of the algorithm

通过观察图 4 中 NMI 变化曲线可以发现:随着重叠社区数量的增加,5 种算法在 3 组数据集上的 NMI 值不断减小;随着网络规模的增大,5 种算法的效率均有所下降,但 CTPT 算法的社区划分效果相对较好,并且具有较好的稳定性。其中,ANA 算法在 Net3 网络数据集上表现较差,NMI 值下降较快,比其他算法效率差。同时,对比 CTPT 算法在 3 个网络数据集上的效果后发现,CTPT 算法在 Net3 网络数据

集上得到的 NMI 值低于 Net1 和 Net2 网络数据集,意味着网络越复杂,CTPT 算法的执行效果越差,这是下一步需要改进之处。由此可见,在大规模复杂网络数据集的重叠社区检测方面,CTPT 算法具有较好的检测效果和稳定性。

3.2.2 真实网络数据集上的结果分析

在真实网络数据集上采用扩展模块度进行重叠社区发现算法的性能评估。表 3、表 4 给出了 CTPT 算法在 3 个真实网络数据集上实验结果,分别选取社区数量 $|C|$ 在不同值的情况下,运行 15 次 CTPT 算法所得的 EQ 均值作为最终结果。

表 3 真实网络数据集上社区发现的 EQ 值($|C|=20$)

Table 3 EQ value of community discovery on real network dataset($ C =20$)					
网络数据集	MDP 算法	TES 算法	GFE 算法	ANA 算法	CTPT 算法
ego-Facebook	0.434	0.463	0.417	0.396	0.512
Musae-twitch	0.487	0.539	0.461	0.445	0.605
Web-flickr	0.533	0.564	0.465	0.434	0.661

表 4 真实网络数据集上社区发现的 EQ 值($|C|=30$)

Table 4 EQ value of community discovery on real network dataset($ C =30$)					
网络数据集	MDP 算法	TES 算法	GFE 算法	ANA 算法	CTPT 算法
ego-Facebook	0.503	0.572	0.489	0.467	0.637
Musae-twitch	0.394	0.548	0.429	0.441	0.573
Web-flickr	0.598	0.627	0.602	0.563	0.719

由于 CTPT 算法充分考虑了社区形成过程中各种因素的共同作用,一方面引入拓扑势,考虑了网络结构(最短路径长度)不同导致的节点重要性不同,另一方面多次融合节点属性,兼顾了节点交互对社区的影响,因此在真实网络数据集上性能表现良好,实验结果整体优于对比算法。TES 算法与 CTPT 算法的第一阶段均为使用网络的节点拓扑特征得到初始社区,在 3 个网络上的 EQ 值仅次于 CTPT 算法,表现较好,尤其是在 Musae-twitch 网络数据集上,在最好情况下 EQ 值仅相差 0.025。ANA 算法实现时会产生较小的链接社区,阻碍了社区的形成,导致难以获得较好的性能,因此在 3 个数据集上的实验结果均是最差的。

4 结束语

本文充分考虑社区形成过程中各种因素的共同作用,提出基于网络拓扑势与信任度调整的重叠社区发现算法。由于节点在网络中所处位置及其链接关系存在差异,使得节点重要性不同,通过网络拓扑势来体现这一特征,并且每个节点又具有独立的固有属性,因此将属性的影响力叠加在拓扑势的计算过程中。在社区合并阶段,将节点间的信任度作为

初始社区能否合并的依据,形成最终的社区划分。在不同数据集上的实验结果验证了本文算法相对于同类算法的优越性。在下一阶段的研究中将尝试在核心节点的选取中加入注意力机制来体现节点重要性,利用图嵌入技术优化生成的属性向量,提升重叠社区发现算法的实现效率,并最终将其应用于实际网络分析及网络推荐系统。

参考文献

- [1] PALLA G, DERÉNYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. *Nature*, 2005, 435(7043): 814-818.
- [2] 阎艳, 黄智兴, 邱玉辉. 一种基于派系过滤的社区进化发现研究[J]. *重庆师范大学学报(自然科学版)*, 2009, 26(2): 90-93.
YAN Y, HUANG Z X, QIU Y H. A research on community evolution discovery based on clique percolation[J]. *Journal of Chongqing Normal University(Nature Science)*, 2009, 26(2): 90-93. (in Chinese)
- [3] EVANS T S. Clique graphs and overlapping communities[EB/OL]. [2020-12-04]. <https://arxiv.org/pdf/1009.0638.pdf>.
- [4] 卢志刚, 吴露. ESN中基于贪婪派系扩张的重叠社区发现[J]. *计算机工程*, 2019, 45(7): 32-40.
LU Z G, WU L. Overlapping community discovery based on greedy factional expansion in ESN[J]. *Computer Engineering*, 2019, 45(7): 32-40. (in Chinese)
- [5] WEN X Y, CHEN W N, LIN Y, et al. A maximal clique based multiobjective evolutionary algorithm for overlapping community detection[J]. *IEEE Transactions on Evolutionary Computation*, 2017, 21(3): 363-377.
- [6] 桑睿彤. 基于扩张思想的局部社区发现方法研究[D]. 北京: 中国石油大学(北京), 2020.
SANG R T. Research on local community detection method based on expansion thought[D]. Beijing: China University of Petroleum(Beijing), 2020. (in Chinese)
- [7] 李有红, 王学军, 谌裕勇, 等. 一种融合邻边属性的个人社交网络社区发现算法[J]. *计算机工程*, 2021, 47(7): 81-87.
LI Y H, WANG X J, CHEN Y Y, et al. A community discovery algorithm fused with adjacent edge attribute for personal social network[J]. *Computer Engineering*, 2021, 47(7): 81-87. (in Chinese)
- [8] LANCICHINETTI A, FORTUNATO S, RADICCHI F. Benchmark graphs for testing community detection algorithms[EB/OL]. [2020-12-04]. <https://arxiv.org/abs/0805.4770>.
- [9] CHANG F R, ZHANG B F, LI H Y, et al. Discovering overlapping communities in ego-nets using friend intimacy[J]. *Journal of Intelligent & Fuzzy Systems*, 2019, 36(6): 5167-5175.
- [10] WILDER B, IMMORLICA N, RICE E, et al. Influence maximization with an unknown network by exploiting community structure[EB/OL]. [2020-12-04]. https://www.researchgate.net/publication/345514410_5_-_Influence_Maximization_with_Unknown_Network_Structure.
- [11] AHN Y Y, BAGROW J P, LEHMANN S. Link communities reveal multiscale complexity in networks[J]. *Nature*, 2010, 466(7307): 761-764.
- [12] GUI Q, DENG R, CHENG X H, et al. A new method for overlapping community detection based on complete subgraph and label propagation[C]//*Proceedings of the 3rd International Conference on Intelligent Information Processing*. New York, USA: ACM Press, 2018: 127-134.
- [13] LEI Y, ZHOU Y, SHI J. Overlapping communities detection of social network based on hybrid C-means clustering algorithm[J]. *Sustainable Cities and Society*, 2019, 47: 101436.
- [14] WANG Y Y, BU Z, YANG H, et al. An effective and scalable overlapping community detection approach: integrating social identity model and game theory[J]. *Applied Mathematics and Computation*, 2021, 390: 125601.
- [15] 韩祺祎, 任梦吟, 文红. 基于拓扑势的P2P社区推荐信任模型[J]. *电子与信息学报*, 2015, 37(6): 1279-1284.
HAN Q Y, REN M Y, WEN H. Topological potential based recommendation trust model for P2P communities system[J]. *Journal of Electronics & Information Technology*, 2015, 37(6): 1279-1284. (in Chinese)
- [16] 吴振宇, 胡军, 李德毅. 社会标注系统幂律特性分析[J]. *复杂系统与复杂性科学*, 2014, 11(2): 5-16.
WU Z Y, HU J, LI D Y. Analysis of the power law characteristics in social tagging systems[J]. *Complex Systems and Complexity Science*, 2014, 11(2): 5-16. (in Chinese)
- [17] 王梦迪, 付顺顺. 基于节点属性的重叠社区发现算法改进[J]. *通信技术*, 2018, 51(1): 128-133.
WANG M D, FU S S. Modified overlapping community detection algorithm based on node attributes[J]. *Communications Technology*, 2018, 51(1): 128-133. (in Chinese)
- [18] NIWATTANAKUL S, SINGTHONGCHAI J, NAENUDORN E, et al. Using of Jaccard coefficient for keywords similarity[EB/OL]. [2020-12-04]. https://www.researchgate.net/publication/317248581_Using_of_Jaccard_Coefficient_for_Keywords_Similarity.
- [19] SHENG J F, ZHU J F, WANG Y Y. Identifying influential nodes of complex networks based on trust-value[J]. *Journal of Chinese Computer Systems*, 2019, 40(11): 2337-2342.
- [20] NEWMAN M, GIRVAN M. Finding and evaluating community structure in networks[J]. 2004, 69(2): 26113.
- [21] SUN P G. Weighting links based on edge centrality for community detection[J]. *Physica A: Statistical Mechanics and Its Applications*, 2014, 394: 346-357.
- [22] ZHANG H Y, CHEN X W, LI J, et al. Fuzzy community detection via modularity guided membership-degree propagation[J]. *Pattern Recognition Letters*, 2016, 70: 66-72.
- [23] 杜航原, 裴希亚, 王文剑. 面向属性网络的重叠社区发现算法[J]. *计算机应用*, 2019, 39(11): 3151-3157.
DU H Y, PEI X Y, WANG W J. Overlapping community detection algorithm for attributed networks[J]. *Journal of Computer Applications*, 2019, 39(11): 3151-3157. (in Chinese)
- [24] LI Y, HE J, WU Y X, et al. Overlapping community discovery method based on two expansions of seeds[J]. *Symmetry*, 2020, 13(1): 1-18.