



计算机科学

COMPUTER SCIENCE

面向双层网络的 EWCC 社区发现算法

唐春阳, 肖玉芝, 赵海兴, 冶忠林, 张娜

引用本文

唐春阳, 肖玉芝, 赵海兴, 冶忠林, 张娜. [面向双层网络的 EWCC 社区发现算法](#)[J]. 计算机科学, 2022, 49(4): 49-55.

TANG Chun-yang, XIAO Yu-zhi, ZHAO Hai-xing, YE Zhong-lin, ZHANG Na. [EWCC Community Discovery Algorithm for Two-Layer Network](#)[J]. Computer Science, 2022, 49(4): 49-55.

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)

[基于节点相似性和网络嵌入的复杂网络社区发现算法](#)

Complex Network Community Detection Algorithm Based on Node Similarity and Network Embedding
计算机科学, 2022, 49(3): 121-128. <https://doi.org/10.11896/jsjcx.210200009>

[基于子图结构的局部社区发现算法](#)

Overlapping Community Detection Algorithm Based on Subgraph Structure
计算机科学, 2021, 48(9): 244-250. <https://doi.org/10.11896/jsjcx.201100010>

[基于重连机制的复杂网络鲁棒性分析](#)

Robustness Analysis of Complex Network Based on Rewiring Mechanism
计算机科学, 2021, 48(7): 130-136. <https://doi.org/10.11896/jsjcx.201000108>

[基于矩阵分解的属性网络嵌入和社区发现算法](#)

Attributed Network Embedding Based on Matrix Factorization and Community Detection
计算机科学, 2021, 48(12): 204-211. <https://doi.org/10.11896/jsjcx.210300060>

[图神经网络社区发现研究综述](#)

Survey of Graph Neural Network in Community Detection
计算机科学, 2021, 48(11A): 11-16. <https://doi.org/10.11896/jsjcx.210500151>

面向双层网络的 EWCC 社区发现算法

唐春阳 肖玉芝 赵海兴 冶忠林 张 娜

青海师范大学计算机学院 西宁 810016

省部共建藏语智能信息处理及应用国家重点实验室 西宁 810008

藏文信息处理教育部重点实验室 西宁 810008

青海省藏文信息处理与机器翻译重点实验室 西宁 810008

(Tangcyqh@163.com)

摘要 针对关系型网络的社区发现问题,考虑节点间相互作用的强弱程度和信息渗流机理,创新性地提出了一种基于边权重和连通分支(Edge Weight and Connected Component,EWCC)的社区发现算法。为了验证算法的有效性,首先,构建了5种具有相互作用的双层网络模型,通过分析层间节点作用的强弱程度对网络拓扑结构的影响,确定了5种双层网络模型下生成的30个数据集;其次,选用真实数据集分别与GN算法和KL算法在模块度、算法复杂度和社区划分数目评价准则上进行了对比,实验结果表明EWCC算法的准确性较高;然后,结合数值仿真得出,随着层间作用关系减弱,模块度值和社区数目成反比,并且当双层网络层间节点关系较弱时,社区划分效果较好;最后,作为算法的应用,利用实证数据构建了“用户-APP”的双层网络并进行了社区划分。

关键词: 关系型网络;社区发现;双层网络;边权重;连通分支

中图法分类号 TP312

EWCC Community Discovery Algorithm for Two-Layer Network

TANG Chun-yang, XIAO Yu-zhi, ZHAO Hai-xing, YE Zhong-lin and ZHANG Na

College of Computer, Qinghai Normal University, Xining 810016, China

The Provincial and Ministerial Joint Construction of the State Key Laboratory of Tibetan Intelligent Information Processing and Application, Xining 810008, China

Key Laboratory of Tibetan Information Processing, Ministry of Education, Xining 810008, China

Tibetan Information Processing and Machine Translation Key Laboratory of Qinghai Province, Xining 810008, China

Abstract Aiming at the problem of community discovery in relational networks, considering the strength of interaction between nodes and information seepage mechanism, an edge weight and connected component (EWCC) community discovery algorithm based on edge weight and connected branches is innovatively proposed. In order to verify effectiveness of the algorithm, firstly, five kinds of interactive two-layer network models are constructed. By analyzing influence of interaction degree of nodes between layers on the network topology, 30 data sets generated under five kinds of two-layer network models are determined. Secondly, the real data set is selected to compare with GN algorithm and KL algorithm in the evaluation criteria of modularity, algorithm complexity and community division number. Experimental results show that EWCC algorithm has high accuracy. Then, the numerical simulation shows that with the weakening of interaction relationship between layers, the module degree is inversely proportional to number of communities, and the community division effect is better when node relationship between layers is weaker. Finally, as an application of the algorithm, the “user-APP” two-layer network is constructed based on empirical data, and the community is divided.

Keywords Relational network, Community discovery, Two-Layer network, Edge weight, Connected branch

1 引言

近年来,复杂网络已逐渐成为许多学科领域的研究热点。

无标度、小世界和社区结构等是复杂网络的基础特性,其中社区结构指在同一社区内的节点之间连接紧凑,在不同社区的节点间连接稀疏^[1]。社区具有模块度、组、群等含义,例如

到稿日期:2021-08-31 返修日期:2021-12-12

基金项目:国家自然科学基金(61763041);青海省科技项目(2020-GX-112);青海师范大学自然科学中青年科研基金(2020QZR007)

This work was supported by the National Natural Science Foundation of China(61763041), Science and Technology Department of Qinghai Province(2020-GX-112) and Middle-Youth Program of Natural Science Foundation of Qinghai Normal University(2020QZR007).

通信作者:赵海兴(h. x. zhao@163.com)

万维网中同一社区内讨论的是共同话题,引文网络中的社区代表的是同一主题相关的论文等。

目前国内外关于社区发现的算法较多^[2-4],如 Newman^[5]等提出的 GN 算法,Xie 等^[6]提出的标签传播算法 LPA,Kernighan 等^[7]提出的 KL 算法等。这些社区发现算法大多被应用于节点关系单一的网络。现实世界中的网络结构更为复杂,不同类节点之间交错连接,即节点间可能存在多种类型的关系,每种关系构成一层网络。例如,社交媒体网络中的用户不仅存在朋友关系,还存在评论、转发关系,并且评论行为等对朋友关系有逆向影响。因此,面向多层关系型网络的社区发现问题已经逐渐成为复杂网络新的研究课题。

关系型网络社区发现的研究方法主要分为两类:基于聚合的方法和基于扩展的方法^[8]。基于聚合的方法又分为划分聚合和网络聚合。划分聚合是在每层网络上进行社区划分,然后将结果进行聚合;网络聚合指将多层网络聚合成单层网络,然后在单层网络上使用传统的社区发现算法进行社团划分;基于扩展的方法是将单一网络社区发现扩展到多维型网络上。现有的多层网络社区发现方法主要包括:Hmimida 等^[9]提出的将单层网络社区发现方法 Licod^[10]扩展到多层网络的方法;Alimadadi 等^[11]提出的 MLPA 算法,用于检测多层网络中的社区结构;Interdonato 等^[12]提出的 ML-LCD 方法,将局部社区发现扩展到多层网络;Kuncheva 等^[13]提出的在多层网络上进行跨层随机游走检测社区结构的方法等。

双层关系型网络的社团挖掘算法倾向于利用实证数据集的特征或相似度矩阵将两层节点关系映射成单层节点关系,从而提出挖掘算法,并在实证数据集中进行验证。特别地,在映射过程中,层间节点容易丢失原有的强弱关系。因此,本文以双层网络拓扑结构为切入点,分析了网络拓扑结构变化时的社团形成规律,充分考虑了双层网络生成机制,即层间/层内节点之间的强弱关系以及信息交互扩散问题,提出了基于边权重和连通分支的 EWCC 社区发现算法,并与经典社区发现算法进行了对比分析。本文将节点强弱关系定义为边权重。首先,根据层间随机耦合机理构建了 5 种概率模型,深入分析了相应网络的度特性,以便确定社区发现的实验数据集;然后,提出了 EWCC 社区发现算法,并在真实网络数据集上验证了算法的有效性;最后,在人工合成网络数据集和“用户-APP”实证网络数据集上将 EWCC 算法和经典 GN 算法以及 KL 算法进行了对比。结果表明,本文算法具有更高的准确率和更低的时间复杂度。

2 双层网络模型演化

2.1 基本定义

双层网络 L 由网络层集合 g 和层间连边 Q 组成。 $g = \{G_\alpha; \alpha \in \{1, 2\}\}$, $G_\alpha = (X_\alpha, E_\alpha)$ 是双层网络的网络层, $Q = \{E_{\alpha\beta} \in X_\alpha \times X_\beta; \alpha, \beta \in \{1, 2\}, \alpha \neq \beta\}$ 为不同网络层 G_α 和 G_β 节点间连边的集合。 E_α 的组成元素为双层网络的层内连边, $E_{\alpha\beta} (\alpha \neq \beta)$ 的组成元素为双层网络的层间连边。其中,每个网络层 G_α 的邻接矩阵为:

$$a_{ij}^\alpha = \begin{cases} 1, & \text{if } (x_i^\alpha, x_j^\alpha) \in E_\alpha \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

对应的 $E_{\alpha\beta}$ 的层间邻接矩阵为:

$$a_{ij}^{\alpha\beta} = \begin{cases} 1, & \text{if } (x_i^\alpha, x_j^\beta) \in E_{\alpha\beta} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

双层网络 L 的投影网络 (Projection Network) 为图 $proj(L) = (X_L, E_L)$, 其中

$$X_L = \bigcup_{\alpha=1}^L X_\alpha \text{ 且 } E_L = \left(\bigcup_{\alpha=1}^L E_\alpha \right) \cup \left(\bigcup_{\alpha, \beta=1, \alpha \neq \beta}^L E_{\alpha\beta} \right) \quad (3)$$

投影网络的邻接矩阵用 A_L 表示。

根据双层网络的定义,图 1 给出了双层网络的拓扑结构和超邻接矩阵。



图 1 双层网络模型图

Fig. 1 Diagram of two-layer network model

2.2 双层网络模型构建

本节充分考虑了双层网络中层间节点相互作用的强弱关系,结合网络层内连接概率和层间连接概率的不确定性,构建了 5 种概率网络模型,分别记为 BE 网络、WE_s 网络、WE_D 网络、EE_s 网络和 EE_D 网络,并分析了网络拓扑指标对网络性能的影响,为后续分析双层网络社区结构的形成机理奠定了基础。

根据双层网络的定义将双层网络记为 $L(N, p, p_{\alpha\beta})$, 其中 N 为整体网络规模, p 为网络层内连接概率, $p_{\alpha\beta}$ 为网络层 G_α 和网络层 G_β 间的连接概率。层内生成算法分别采用无标度网络生成算法^[14]、小世界网络生成算法^[15]和随机网络生成算法^[16], 5 种概率网络模型的演化过程如下。

2.2.1 BE 网络模型的演化过程

(1) 网络层内连边。网络层 G_α 和网络层 G_β 均采用无标度网络生成算法构建,网络层的规模为 n 。

(2) 网络层间连边。在网络层 G_α 和网络层 G_β 中选择未被选择过的节点对,生成随机数 $m(0, 1)$, 如果 $m < p_{\alpha\beta}$, 那么在这对节点之间添加链路,直到层间连边数为 $n \times n \times p_{\alpha\beta}$ 时,算法结束。将形成的双层网络简称为 BE 网络,记为 $G_{BE}(N, p, p_{\alpha\beta})$, 网络规模为 $N = 2n$ 。

2.2.2 WE 网络模型的演化过程

(1) 网络层内连边。网络层 G_α 和网络层 G_β 均采用小世界网络生成算法构建,网络层的规模为 n 。

(2) 网络层间连边。在网络层 G_α 和网络层 G_β 中选择未被选择过的节点对,生成随机数 $m(0, 1)$, 如果 $m < p_{\alpha\beta}$, 那么在这对节点之间添加链路,直到层间连边数为 $n \times n \times p_{\alpha\beta}$ 时,算法结束。将形成的网络简称为 WE 网络,记为 $G_{WE}(N, p, p_{\alpha\beta})$, 网络规模为 $N = 2n$ 。

(3) 当 $p = p_{\alpha\beta}$ 时,将形成的双层网络简称为 WE_s 网络。

(4) 当 $p + p_{\alpha\beta} = 1$ 时,将形成的双层网络简称为 WE_D 网络。

2.2.3 EE网络模型的演化过程

(1)网络层内连边。网络层 G_α 和网络层 G_β 均采用随机网络生成算法构建,网络层的规模为 n 。

(2)网络层间连边。在网络层 G_α 和网络层 G_β 中选择未被选择过的节点对,生成随机数 $m(0,1)$,如果 $m < p_{\alpha\beta}$,那么在这对节点之间添加链路,直到层间连边数为 $n \times n \times p_{\alpha\beta}$ 时,算法结束。将形成的网络简称为EE网络,记为 $G_{EE}(N, p, p_{\alpha\beta})$,网络规模为 $N=2n$ 。

(3)当 $p = p_{\alpha\beta}$ 时,将形成的双层网络简称为 EE_S 网络。

(4)当 $p + p_{\alpha\beta} = 1$ 时,将形成的双层网络简称为 EE_D 网络。

2.3 双层网络特性分析

基于不同生成机制演化的网络结构具有一定的差异性。度作为衡量网络结构的基本拓扑指标,其在双层网络中呈现的度特性和在单层网络中呈现的结果是否一致,值得深入研究。因此,本节通过数值仿真分析了5种双层网络的度特性,仿真中约定5种双层网络的规模 N 皆为300。

度分布 $P(k)$ 表示网络中随机选出的一个节点的度为 k 的概率。由于 $P(k)$ 是一个概率,因此其必须满足归一化约束,即:

$$\sum_{k=0}^{\infty} P_k = 1 \quad (4)$$

对于有 N 个节点的网络而言,其度分布可以表示为:

$$P(k) = \frac{N_k}{N} \quad (5)$$

其中, N_k 指度为 k 的节点个数。

2.3.1 BE网络度特性分析

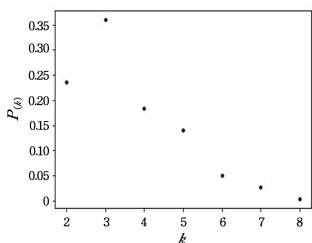
在BE网络模型中,根据无标度算法,每次引入一个新的节点和 m 条边,实验中设置 $m=2$,网络层间连接概率 $p_{\alpha\beta}$ 的取值范围为 $[0,1]$ 。BE网络在部分层间连接概率下的网络参数如表1所列。

表1 BE网络参数

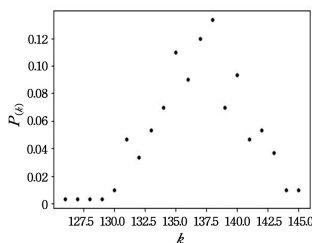
Table 1 BE network parameters

网络类型	网络参数			度分布特性	
	N	$p_{\alpha\beta}$	M		
BE网络	300	0.01	817	5.44	幂律
		0.1	2842	18.94	泊松
		0.3	7342	48.94	泊松
		0.5	11842	78.94	泊松
		0.7	4254	28.36	泊松
		0.9	20842	138.94	泊松

表1中, N 表示网络总节点数, $p_{\alpha\beta}$ 表示网络层间连接



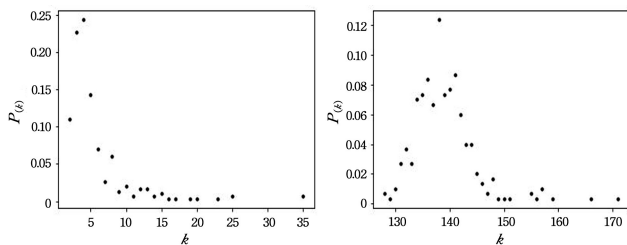
(a) $p=0.01, p_{\alpha\beta}=0.01$



(b) $p=0.9, p_{\alpha\beta}=0.9$

概率, M 表示网络总边数, $\langle k \rangle$ 表示整个网络的平均度。后文中的 $N, p_{\alpha\beta}, M$ 和 $\langle k \rangle$ 均与此处的含义相同。

图2给出了表1中当 $p_{\alpha\beta}$ 的取值为0.01和0.9时,BE网络的度分布情况。



(a) $m=2, p_{\alpha\beta}=0.01$

(b) $m=2, p_{\alpha\beta}=0.9$

图2 BE网络度分布

Fig. 2 BE network degree distribution

由表1和图2得出:

(1)当 $p_{\alpha\beta}$ 趋于0.01时,网络呈幂律分布;当 $p_{\alpha\beta}$ 趋于0.9时,网络呈泊松分布。

(2)层间随机连接概率值越小,层间节点间的相互关系较弱时,网络呈幂律分布。

2.3.2 基于相同层内/层间连接概率的网络度特性分析

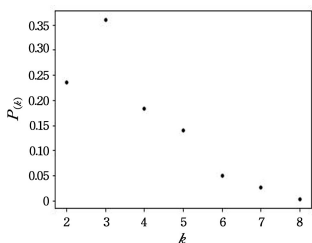
根据网络生成机制,双层网络涉及层内连接概率 p 和层间连接概率 $p_{\alpha\beta}$ 。因此,本小节考虑了层内随机依赖程度和层间随机依赖程度相同,即 $p = p_{\alpha\beta} = [0,1]$ 时,部分连接概率下 WE_S 网络和 EE_S 网络的参数情况,具体如表2所列。

表2 WE_S 网络和 EE_S 网络的参数

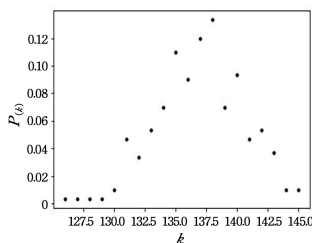
Table 2 Parameters of WE_S and EE_S networks

网络类型	网络参数				度分布特性	
	N	p	$p_{\alpha\beta}$	M		
WE_S 网络	300	0.01	0.01	525	3.5	泊松
		0.1	0.1	2550	17	泊松
		0.3	0.3	7050	47	泊松
		0.5	0.5	11550	77	泊松
		0.7	0.7	16050	107	泊松
		0.9	0.9	20550	137	泊松
EE_S 网络	300	0.01	0.01	403	2.68	泊松
		0.1	0.1	4538	30.25	泊松
		0.3	0.3	13390	89.26	泊松
		0.5	0.5	22456	149.70	泊松
		0.7	0.7	31184	207.89	泊松
		0.9	0.9	40268	268.45	泊松

图3给出了表2中 p 和 $p_{\alpha\beta}$ 的取值相同且分别为0.01和0.9时, WE_S 网络和 EE_S 网络的度分布情形。



(c) $p=0.01, p_{\alpha\beta}=0.01$



(d) $p=0.9, p_{\alpha\beta}=0.9$

图3 WE_S 网络和 EE_S 网络的度分布

Fig. 3 Degree distribution of WE_S and EE_S networks

结合表 2 和图 3 得出, WE_S 网络和 EE_S 网络均呈泊松分布; 层间节点间相互关系的强弱对网络的平均度有较大影响, 关系越强, 平均度越大。

2.3.3 基于不同层内/层间连接概率的网络度特性分析

本小节考虑了层内随机依赖程度和层间随机依赖程度不同即 $p + p_{\alpha\beta} = 1$ 时, 部分连接概率下 WE_D 网络和 EE_D 网络的参数设置, 具体情况如表 3 所列。

表 3 WE_D 网络和 EE_D 网络参数

Table 3 Parameters of WE_D and EE_D networks

网络类型	网络参数					度分布特性
	N	p	$p_{\alpha\beta}$	M	$\langle k \rangle$	
WE_D 网络	300	0.01	0.99	22575	150.48	泊松
		0.1	0.9	20550	137	泊松
		0.3	0.7	16050	107	泊松
		0.5	0.5	11550	77	泊松
		0.7	0.3	7050	47	泊松
		0.9	0.1	2550	17	泊松
EE_D 网络	300	0.01	0.99	22493	149.95	泊松
		0.1	0.9	22552	150.34	泊松
		0.3	0.7	22452	149.68	泊松
		0.5	0.5	22456	149.70	泊松
		0.7	0.3	22434	149.56	泊松
		0.9	0.1	22404	149.36	泊松

从表 3 可以看出: WE_D 网络和 EE_D 网络均呈现出单峰特性; 当层间依赖值较大时, WE_D 网络和 EE_D 网络的平均度相对较大, 且层间节点间相互关系的强弱对 EE_D 网络的平均度影响较小, 对 WE_D 网络的影响较大。

综上所述, 在关系型双层网络中, 生成机制和配置参数的不同, 使得网络度分布呈现出不同的特性。在关系型双层网络的社区发现中, 重点考虑层内节点和层间节点的依赖程度。一般来说, 强的依赖程度能促进社区快速形成, 并且在双层网络之间按照社团结构快速传播信息。因此, 下文依据节点的强弱关系以及在连通分支中信息传播的优势, 提出了基于边权重和连通分支的社区发现算法, 并将算法在双层网络实验数据集集中进行验证, 目的在于检测拓扑结构变化的网络中的社区, 并为多层相互关联的网络社区发现提供一定的指导思想。

3 EWCC 算法框架

3.1 边权重

用 w_{ij} 表示节点 i 和 j 之间连边的权重, 边权重可以定义为边两端节点度的乘积^[17]。

$$D(e) = w_{ij} = k_i k_j \quad (6)$$

其中, k_i 和 k_j 分别表示节点 i 和节点 j 的度。

3.2 连通分支

若图 G 中每一对不同的顶点 u, v 都有一条 (u, v) 道路, 则称图 G 是连通的。以 $u \equiv v$ 表示顶点 u 和 v 是连通的, 那么这种顶点间的连通关系是一个等价关系。因此, 等价关系 $u \equiv v$ 确定了顶点集 v 的一个分类, 把 v 分成非空子集 v_1, v_2, \dots, v_k , 当且仅当两个顶点 u 和 v 属于同一子集 v_i 时, 它们才是连通的, 子图 $G[v_1], G[v_2], \dots, G[v_k]$ 被称为 G 的连通分支, 连通分支数记为 $k(G)$ 。

3.3 EWCC 社区发现算法的描述

基于分裂的层次聚类思想, 将连通分支数看作社区个数, 提出了基于边权重和连通分支的 EWCC 社区发现算法。该算法的流程图如图 5 所示。

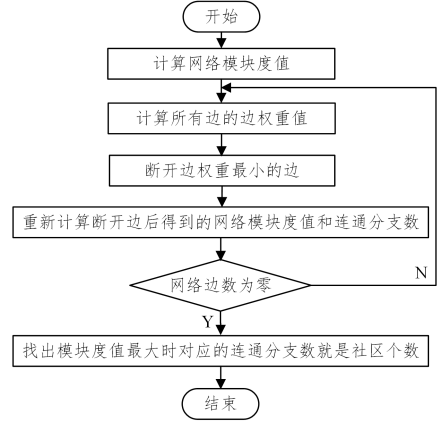


图 5 EWCC 算法流程图

Fig. 5 Flow chart of EWCC algorithm

EWCC 算法的伪代码如算法 1 所示, 主要步骤如下:

- (1) 初始时, 将整个网络看作一个社区并计算网络的模块度值;
- (2) 计算网络中所有边的边权重值, 把边权重小的边断开, 得到新网络, 计算新网络的模块度和连通分支数;
- (3) 重复步骤(2)直到网络中的边数为零, 算法结束, 找出模块度最大时对应的连通分支数即为社区个数。

算法 1 EWCC 社区发现算法

输入: 具有 M 条边的双层网络 L 的邻接矩阵 $A(L)$

输出: 模块度最大值 Q , 连通分支数 C

1. 将 L 的邻接矩阵 $A(L)$ 还原成一个网络, 初始时整个网络可看作一个社区;
2. IF $\text{MIN}(EW)$; // 得到网络中边权重最小的边并断开
3. do $E - 1$; // 将网络的边数减少 1
4. GET C ; // 社区个数 C
5. IF $C = C + 1$; // 连通分支数增加 1
6. GET Q // 获得此时的模块度 Q 值
7. $qList.add(Q)$; // 维护所有的模块度
8. END IF;
9. FOR EW IN $G.edge_weights$; // 遍历所有边的边权重值
10. GET $\text{MIN}(EW)$; // 得到边权重的最小值
11. END FOR;
12. until $E = 0$; // 网络 G 的总边数减少到 0
13. END IF;
14. FOR Q IN $qList$; // 遍历所有的模块度值
15. GET $\text{MAX}(Q)$; // 得到模块度的最大值
16. END FOR;

3.4 时间复杂度分析

对于具有 M 条边和 N 个节点的网络而言, 在 EWCC 算法中, 初始计算 M 条边的边权重值, 时间复杂度为 $O(M)$ 。遍历后将网络中边权重最小的边断开得到新网络, 接着遍历新网络中剩余边的边权重值, 最多需要遍历 M 次。因此, 算法的时间复杂度为 $O(M^2)$ 。

GN 算法的时间复杂度为 $O(MN^2)$, KL 算法的时间复杂度为 $O(N^2 \log N)$, 因此, EWCC 算法有一定的优势。

4 算法评价标准

模块度是衡量社区划分准确率的一个重要指标, 模块度越大社区划分质量越好, 模块度越小社区划分质量越差。模块度的定义如下^[18]:

$$Q = \frac{1}{2m} \sum (A_{u,v} - \frac{d_u \times d_v}{2m}) \delta(C_u, C_v) \quad (7)$$

其中, m 为网络边数; d_u, d_v 分别为节点 u 和节点 v 的度; $A_{u,v}$ 是网络邻接矩阵元素, 若节点 u 和节点 v 存在连边, 则 $A_{u,v} = 1$, 否则为 0; C_u 表示节点 u 所属的社区, C_v 表示节点 v 所属的社区, 当 $C_u = C_v$ 时, $\delta(C_u, C_v) = 1$, 说明节点 u 和 v 属于同一个社区; 模块度 Q 的取值范围为 $[-0.5, 1)$ 。

5 仿真实验与结果分析

5.1 数据集描述

为了验证算法的真实性和有效性, 在实验中采用了三类数据集进行测试, 包括经典网络数据集、人工合成网络数据集和“用户-APP”实证数据集。3 类数据集的具体信息如下。

(1) 经典网络数据集

经典网络数据集如表 4 所列。

表 4 经典网络数据集

Table 4 Classic network datasets

名称	节点	边	平均度	数据集描述
Karate club	34	78	4.58	空手道俱乐部网络
Football	115	613	10.66	美国大学生足球网

(2) 人工合成网络数据集

人工合成网络数据集选取了 2.2 节中构建的 BE 网络、WE_s 网络、EE_s 网络、WE_D 网络和 EE_D 网络分别在不同层间随机耦合概率下生成的 30 个数据集, 其具体数据信息包括表 1 列出的 BE 网络在部分层间连接概率下生成的 6 个数据集, 表 2 列出的 WE_s 网络和 EE_s 网络在层内和层间连接概率同时生成的 12 个数据集, 表 3 列出的 WE_D 网络和 EE_D 网络在层内和层间连接概率不同时生成的 12 个数据集。

(3) “用户-APP”实证数据集

作为算法的应用, 本小节采用了某运营商提供的用户及用户 APP 流量使用情况, 构建了“用户-APP”双层网络。其中, 用户层包含 1 147 个用户, APP 层包含传统视频和短视频两大类, 传统视频包括爱奇艺、腾讯、优酷等, 短视频包括抖音、快手、西瓜等。其网络模型如图 6 所示。

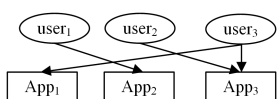


图 6 “用户-APP”双层网络模型图

Fig. 6 Diagram of “user-APP” two-layer network model

图 6 中, 双层网络的两个子网络层内没有连边, 层间按照用户使用 APP 的行为进行连边。为了构造用户与用户间的关系, 结合用户使用 APP 流量值, 利用随机概率 p 刻画用户层内节点的连边关系, 即将“用户-APP”层间关系映射成

用户层内关系, 并生成最终的网路模型 G 。网络 G 包含了 1 174 个节点和 8 331 条边, 网络平均聚类系数为 0.061, 网络直径为 7。

5.2 社区发现实验结果

5.2.1 经典网络社区划分

为了验证 EWCC 算法的真实性和有效性, 在 Karate club 数据集和 Football 数据集上将所提算法与 GN 算法和 KL 算法进行了对比实验, 结果如表 5 和表 6 所列。

表 5 各算法在 Karate club 数据集上的结果的比较

Table 5 Comparison of results of each algorithm on Karate club dataset

算法	指标	
	模块度	社区个数
GN	0.396 860 61	5
KL	0.397 435 89	2
EWCC	0.400 969 52	2

表 6 各算法在 Football 数据集的结果的比较

Table 6 Comparison of results of each algorithm on Football dataset

算法	指标	
	模块度	社区个数
GN	0.590 446 52	10
KL	0.603 161 78	2
EWCC	0.636 283 95	12

分析表 5 和表 6 的实验结果可知, 在 Karate club 和 Football 数据集上, EWCC 算法的模块度值均优于 GN 算法和 KL 算法, 且能正确划分出社区个数; 在 Football 数据集上 EWCC 算法的社区划分数量较多; 结果表明本文算法具有较好的社区划分性能。

5.2.2 人工合成网络社区划分

本小节旨在探索拓扑结构变化在关系型双层网络中的社区形成规律。

(1) BE 网络的社区划分

使用 GN, KL 和 EWCC 算法对 BE 网络进行社区划分, 结果如表 7 所列。

表 7 各算法在 BE 网络上的结果的比较

Table 7 Comparison of results of each algorithm on BE network

$p_{\alpha\beta}$	GN	KL	EWCC	EWCC 社区个数
0.01	0.406 464 37	0.410 695 90	0.662 668 77	190
0.1	0.186 327 18	0.184 325 75	0.280 837 50	258
0.3	0.099 182 01	0.098 536 37	0.160 325 00	281
0.5	0.065 652 91	0.065 993 81	0.115 052 74	257
0.7	0.044 003 47	0.042 055 12	0.106 113 03	288
0.9	0.021 976 59	0.022 333 99	0.088 428 66	285

由表 7 可以得出:

1) 整体而言, 采用 EWCC 算法划分社区时, 模块度相对较大, 说明 EWCC 算法的性能优于 GN 算法和 KL 算法。

2) 随着 BE 网络层间连接概率逐渐增大, GN, KL 和 EWCC 算法的模块度逐渐减小; 当层间连接概率越大时, 使用 EWCC 算法划分得到的社区个数越多, 划分结果越零碎, 划分意义不大。因此, 在 BE 网络呈幂律分布时, 社区划分结果越好。

(2) WE_s网络和 EE_s网络的社区划分

本文基于相互作用网络层间随机连接概率不确定的情况,分析了网络层内和层间连接概率相等时各算法的社区划分性能。

表 8 和表 9 列出了 WE_s网络和 EE_s网络在不同算法下的模块度值和社区个数的变化情况。

表 8 各算法在 WE_s网络上的结果的比较Table 8 Comparison of results of each algorithm on WE_s network

p	$p_{\alpha\beta}$	GN	KL	EWCC	EWCC 社区个数
0.01	0.01	0.56068390	0.56274467	0.57085400	17
0.1	0.1	0.20280807	0.20157447	0.29779404	246
0.3	0.3	0.10387315	0.09828796	0.16230647	254
0.5	0.5	0.06710955	0.06670749	0.12500000	291
0.7	0.7	0.04263084	0.04218448	0.10956790	293
0.9	0.9	0.02177733	0.02138285	0.08906525	286

表 9 各算法在 EE_s网络上的结果的比较Table 9 Comparison of results of each algorithm on EE_s network

p	$p_{\alpha\beta}$	GN	KL	EWCC	EWCC 社区个数
0.01	0.01	0.63020522	0.64163316	0.78780718	254
0.1	0.1	0.14445123	0.14462937	0.27160493	287
0.3	0.3	0.07170516	0.07106200	0.11568887	227
0.5	0.5	0.04303637	0.04291866	0.07451255	252
0.7	0.7	0.02939252	0.04218448	0.10956790	293
0.9	0.9	0.01259126	0.02138285	0.08906525	286

由表 8、表 9 可以得出:

1)整体而言,采用 EWCC 算法划分社区时,模块度相对较大,说明 EWCC 算法的性能优于 GN 算法和 KL 算法。

2)随着 WE_s网络和 EE_s网络的层间连接概率逐渐减小,GN, KL 和 EWCC 算法的模块度逐渐增大,且 EWCC 算法划分得到的社区个数整体上逐渐减少;反之,当网络层间连接概率增大时,EWCC 算法的模块度逐渐减小,划分得到的社区个数整体上逐渐增多,社区划分结果较为零碎。因此,在 WE_s网络和 EE_s网络层间连接概率较小时,社区划分效果越好。

3)对比表 8 和表 9 的实验结果可知,使用 EWCC 算法划分社区时,在基于小世界算法生成的 WE_s网络中,当层间连接概率较小时,社区个数迅速减少。

(3) WE_D网络和 EE_D网络的社区划分

下文分析了基于不同层内/层间连接概率下各算法的社区划分性能。表 10 和表 11 列出了 WE_D网络和 EE_D网络在不同层内/层内连接概率下,EWCC 算法与 GN 算法和 KL 算法在模块度指标和社区数目上的实验对比结果。

表 10 各算法在 WE_D网络上的结果的比较Table 10 Comparison of results of each algorithm on WE_D network

p	$p_{\alpha\beta}$	GN	KL	EWCC	EWCC 社区个数
0.01	0.99	0.01031966	0.01035938	0.09183673	288
0.1	0.9	0.02215378	0.02190977	0.12204081	290
0.3	0.7	0.04334748	0.04323163	0.13223140	292
0.5	0.5	0.06710955	0.06670749	0.12500000	291
0.7	0.3	0.10311530	0.09747483	0.16060803	261
0.9	0.1	0.19561284	0.20896670	0.29708636	270

表 11 各算法在 EE_D网络上的结果的比较Table 11 Comparison of results of each algorithm on EE_D network

p	$p_{\alpha\beta}$	GN	KL	EWCC	EWCC 社区个数
0.01	0.99	0.00793221	0.00815821	0.07660590	288
0.1	0.9	0.02914093	0.02569671	0.07529843	287
0.3	0.7	0.04185155	0.04266255	0.07209508	257
0.5	0.5	0.04303637	0.04291866	0.07451255	252
0.7	0.3	0.19911741	0.19911741	0.19978715	12
0.9	0.1	0.39957150	0.39957150	0.39970896	4

由表 10、表 11 可以得出:

(1)整体而言,采用 EWCC 算法划分社区时,模块度相对较大,说明 EWCC 算法的性能优于 GN 算法和 KL 算法。

(2)随着 WE_D网络和 EE_D网络的层间连接概率逐渐减小(层内连接概率逐渐增大),GN, KL 和 EWCC 算法的模块度逐渐增大,且采用 EWCC 算法划分得到的社区个数相对减少;反之,当 EWCC 算法模块度逐渐减小时,划分得到的社区个数逐渐增多。因此,在 WE_D网络和 EE_D网络层间连接概率较小(层内连接概率较大)时,社区划分效果较好。

(3)对比表 10 和表 11 的实验结果可知,使用 EWCC 算法划分社区时,层间随机耦合概率的变化对 WE_D网络的社区个数影响不明显;而在 EE_D网络中,层间连接概率较小时,划分得到的社区个数急剧减少。

综上所述,在算法优势上,EWCC 算法的性能优于 GN 算法和 KL 算法;在度特性对社区划分的影响上,BE 网络呈幂律分布时,社区划分效果较好;在层间节点相互作用的强弱程度对社区划分的影响上,层内/层间节点作用关系相同或不同时,随着层间节点相互作用关系减弱,算法模块度增强;在网络拓扑结构对社区划分的影响上,WE_s网络和 EE_s网络拓扑结构的变化会对社区划分结果产生较大影响,而 WE_D网络和 EE_s网络拓扑结构的变化对社区划分结果的影响相对较小。

5.2.3 “用户-APP”实证网络社区划分

不同社区发现算法在“用户-APP”实证网络上的运行结果如表 12 所列。

表 12 各算法在“用户-APP”实证网络上的结果的比较

Table 12 Comparison of results of each algorithm on “user-APP”

算法	empirical network	
	指标	
	模块度	社区个数
GN	0.39047307	5
KL	0.39144661	2
EWCC	0.39826094	6

由表 12 可以看出,相比两种经典算法,EWCC 算法的社区发现效果有显著提升。该算法成功识别出 6 款 APP(见图 7)。

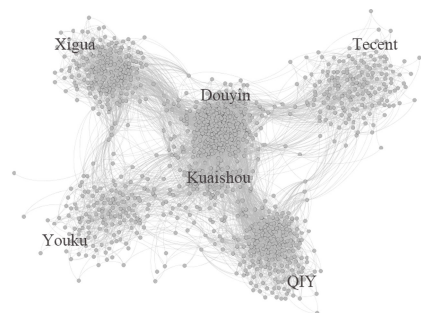


图 7 EWCC 算法在实证网络上的运行结果

Fig. 7 Running results of EWCC algorithm on empirical network

不难看出,西瓜用户更倾向于使用优酷视频,抖音用户更倾向于观看腾讯视频。根据划分结果可制定传统视频类和短视频类相结合的套餐流量包,为用户提供更好的观看体验。

结束语 本文基于关系型双层网络中节点之间的相互作用强度和网络拓扑结构,首先构建了5种相互作用的双层网络模型,提出了EWCC社区发现算法,并通过实验证明了算法的准确性;然后,在人工合成网络数据集上探究了拓扑结构变化中的网络社区结构,发现当双层网络层间节点耦合关系较弱时算法的性能较好;最后,通过建立“用户-APP”模型来进行算法的应用和推广。在局部算法的比较上,本文算法体现出相对较优的划分效果,但是仍存在一些问题。1)网络建模机制,在本文构建的双层网络模型中,层内采用了经典的网络生成算法,而层间仅考虑了随机耦合关系,后续工作将考虑层间多种耦合机制;2)EWCC社区发现算法的改进,在本文提出的算法中,将节点间的强弱关系用边权重定量刻画,在社区划分中具有确定性。实验证明,在多层网络中,层间节点关系复杂,因此后续工作将结合机器学习的方法来确定节点间关系的强弱,充分挖掘隐含关系。

参考文献

- [1] WANG X F, LIU Y B. A survey of community structure algorithms in complex networks[J]. Journal of University of Electronic Science and Technology of China, 2009, 38(5): 537-543.
- [2] XIE J, KELLEY S, SZYMANSKI B K. Overlapping community detection in networks: the state of the art and comparative study [J]. ACM Computing Surveys, 2013, 45(4): 1-35.
- [3] AN X D, ZHANG X Q, CAO F Y. Binary network community discovery algorithm based on edge density propagation[J]. Computer Applications and Software, 2019, 36(3): 243-248, 254.
- [4] ZHANG H, WU Y K, YANG Z Z, et al. Community discovery method based on multi-layer node similarity [J]. Computer Science, 2018, 45(1): 216-222.
- [5] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. Proc. Natl. Acad. Sci., 2001, 99(12): 7821-7826.
- [6] XIE J, SZYMANSKI B K. Community Detection Using a Neighborhood Strength Driven Label Propagation Algorithm [C] // 2021 IEEE Network Science Workshop. West Point, NY, USA: IEEE, 2011: 188-195.
- [7] KERNIGHAN B W, LIN S. An efficient heuristic procedure for partitioning graphs [J]. Bell System Technical Journal, 1970, 49(2): 291-307.
- [8] CHEN K J, CHEN L M, WU T. A review of research on discovery of multi-layer Network communities [J]. Journal of Frontiers of Computer Science & Technology, 2020, 14(11): 1801-1812.

- [9] HMIMIDA M, KANAWATI R. Community Detection in Multiplex Networks: A Seed-centric Approach [J]. Networks & Heterogeneous Media, 2015, 10(1): 71-85.
- [10] YAKOUBI Z, KANAWATI R. LICOD: A leader-driven algorithm for community detection in complex networks [J]. Vietnam Journal of Computer Science, 2014, 1(4): 241-256.
- [11] ALIMADADI F, KHANANGI E, BAGHERI A. Community detection in facebook activity networks and presenting a new multilayer label propagation algorithm for community detection [J]. International Journal of Modern Physics B, 2019, 33(10): 1950089(1)-1950089(21).
- [12] INTERDONATO R, TAGARELLI A, IENCO D, et al. Local Community Detection in Multilayer Networks [J]. Data Mining & Knowledge Discovery, 2017, 31(5): 1444-1479.
- [13] KUNCHEVA Z, MONTANA G. Community detection in multiplex networks using locally adaptive random walks [C] // Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Paris, France: IEEE, 2015: 1308-1315.
- [14] BARABASI A L, ALBERT R. Emergence of Scaling in Random Networks [J]. Science, 1999, 286(5439): 509-512.
- [15] WATTS D J, STROGATZ S H. Collective dynamics of ‘small-world’ network [J]. Nature, 1998, 393(6684): 440-442.
- [16] ERDOS P, RENYI A. On the Evolution of Random Graphs [J]. Publ. Math. Inst. Hung. Acad. Sci., 1960, 5(1): 17-61.
- [17] HOLME P, KIM B J, YOON C N, et al. Attack vulnerability of complex networks [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2002, 65(5): 056109.
- [18] NEWMAN M, GIRVAN M. Finding and evaluating community structure in networks [J]. Physical Review E, 2004, 69(2): 423-433.



TANG Chun-yang, born in 1993, post-graduate, is a member of China Computer Federation. Her main research interests include complex network, multi-layer network theory and application.



ZHAO Hai-xing, born in 1969, Ph. D, professor, Ph. D supervisor. His main research interests include complex network, semantic network and machine translation, database and hypergraph theory, and network reliability.

(责任编辑:喻黎)