

LBSN 中融合多维关系的社区发现方法*

龚卫华¹, 陈彦强¹, 裴小兵², 杨良怀¹



¹(浙江工业大学 计算机科学与技术学院, 浙江 杭州 310023)

²(华中科技大学 软件学院, 湖北 武汉 430074)

通讯作者: 裴小兵, E-mail: xiaobingp@hust.edu.cn

摘要: 如何发现高质量的社区结构对于深刻研究和分析基于位置的社交网络(location-based social networks, 简称 LBSN)这种新型复杂网络具有重要意义,然而,现有的面向社交网络的社区发现方法都无法适用于具有多维异构关系的 LBSN.为此,提出了一种基于联合聚类的用户社区发现方法 Multi-BVD,该方法首先给出了融合用户社交网络与地理位置标签网络中多模实体及其异构关系的社区划分目标函数,然后使用拉格朗日乘子法得到目标函数极小值的迭代更新规则,并运用块值矩阵分解技术来确定最优的社区划分结果.仿真实验结果表明,Multi-BVD 方法能够有效地发现 LBSN 中具有地理特征的用户社区结构,该社区结构在社交关系和地理兴趣标签上都有更优的内聚性,并能更紧密地体现用户社区与地理标签簇间的兴趣关联性.

关键词: 基于位置的社交网络;社区发现;联合聚类;矩阵分解

中图法分类号: TP311

中文引用格式: 龚卫华,陈彦强,裴小兵,杨良怀.LBSN 中融合多维关系的社区发现方法.软件学报,2018,29(4):1163–1176. <http://www.jos.org.cn/1000-9825/5269.htm>

英文引用格式: Gong WH, Chen YQ, Pei XB, Yang LH. Community detection of multi-dimensional relationships in location-based social networks. Ruan Jian Xue Bao/Journal of Software, 2018,29(4):1163–1176 (in Chinese). <http://www.jos.org.cn/1000-9825/5269.htm>

Community Detection of Multi-Dimensional Relationships in Location-Based Social Networks

GONG Wei-Hua¹, CHEN Yan-Qiang¹, PEI Xiao-Bing², YANG Liang-Huai¹

¹(School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

²(School of Software Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: How to detect the high-quality community structures in location based social networks (LBSN) plays a significant role that helps to study and analyze this novel type of composite network comprehensively. However, most of existing community detection methods in social networks still cannot solve the problems of combining the correlations of multi-typed heterogeneous relations in LBSN. To address the issue, this paper proposes a co-clustering method for mining the users' community with multi-dimensional relationships, called Multi-BVD. Firstly, the objective function of clustering community is given to fuse multi-modal entities and their multi-dimensional relationships embedded in users' social network and geo-tagged location network. Then, in order to gain the minimum value of the given function, Lagrange multiplier method is applied to obtain the iterative upgrading rules of matrix variants so that the optimal results of users' communities can be determined by the way of decomposing block matrices. Simulation results show that the

* 基金项目: 浙江省自然科学基金(LY13F020026, LY14F020017, LY14C130005); 中国博士后科学基金(2015M581957); 国家自然科学基金(61571400, 31471416); 浙江省博士后择优资助科研项目(BSH1502019)

Foundation items: Natural Science Foundation of Zhejiang Province, China (LY13F020026, LY14F020017, LY14C130005); China Postdoctoral Science Foundation (2015M581957); National Natural Science Foundation of China (61571400, 31471416); Excellent Postdoctoral Research Projects of Zhejiang Province (BSH1502019)

收稿时间: 2016-05-08; 修改时间: 2016-07-14, 2016-11-03, 2017-01-11; 采用时间: 2017-02-16; jos 在线出版时间: 2017-03-31

CNKI 网络优先出版: 2017-03-31 21:54:38, <http://kns.cnki.net/kcms/detail/11.2560.TP.20170331.2154.005.html>

proposed Multi-BVD can find the community structures with geographical characteristics more effectively and accurately in location based social network. At the same time, the mined non-overlapping community has more cohesive structures in both social relationships and geographical tagged interests, which also can better embody the correlations of interests between users' communities and semantic geo-tagged clusters on locations.

Key words: location-based social networks; community detection; co-clustering; matrix decomposition

近年来,随着社交网络与位置服务的融合与发展,形成了基于位置的社交网络(location-based social networks,简称 LBSN),这种新型网络通过移动用户的位置签到功能,把线上虚拟社会与线下真实世界关联在一起,比传统的社交网络具有更复杂的关系结构特征和更广泛的应用服务,例如 FourSquare、FacebookPlaces 和 Yelp 等社交应用不仅具备传统社交网络的社交功能,还能衍生出多种多样的位置服务,如用户位置签到及与位置相关的信息共享和热点推荐等.由此而形成的复杂社区已成为目前 LBSN 领域最基础性的研究课题,它对于人们深入观察和理解 LBSN 这种多维复合网络的拓扑特征、群体行为演化、链接预测等具有十分重要的意义和价值.

现有的面向社交网络领域的社区发现研究已取得许多成果,归纳起来主要有基于拓扑结构的社区划分^[1-4]和基于语义聚类^[5-9]的社区发现.前者具有代表性的有图分割^[1]、模块度^[2,3]和标签传播^[4]等算法,这些方法仅从外在链接形态划分的社区结构忽略了用户之间的共同兴趣特征,存在结构形式单一和扩展性差等问题,也不能保证社区发现的兴趣内聚性;后者是较为流行的社区发现方式,将节点关系强度与属性信息进行结合发现语义社区,其优点是社区挖掘结果更准确、更凝聚,比较适合发现重叠的社区结构.然而,上述方法都无法直接适用于 LBSN 这类具有多维关系网络中的社区发现,究其原因在于,LBSN 是由多种实体对象及其异构关系构成的复合网络,除了包含用户、地点等多模实体外,还有用户间的社交关系、地理位置关系以及用户-位置签到连接等多种联系.由此可见,如何有效地解决这些多模异构关系的关联性与融合难题是实现 LBSN 中高质量社区发现的关键技术.

为此,本文针对 LBSN 的复合关系网络特征提出一种基于联合聚类的用户社区发现方法——Multi-BVD,该方法首先给出融合了用户-用户社交关系、地理位置-位置相关关系及用户-地理位置签到关系等多维关系的用户社区划分目标函数,然后使用拉格朗日乘子法得到目标函数极小值的迭代更新规则,并运用块值矩阵分解技术来确定最优的社区划分结果.最后实验结果验证了 Multi-BVD 方法能够有效地发现 LBSN 中具有地理特征的用户社区结构,该用户社区在社交关系和地理兴趣标签上都有更优的内聚性.

本文第 1 节介绍 LBSN 社区发现的研究现状.第 2 节给出 LBSN 的网络建模方法,并提出基于联合聚类的用户社区发现算法.第 3 节进行仿真实验及结果分析.最后,第 4 节总结全文.

1 相关工作

目前,针对 LBSN 这种复合网络的社区发现研究还未引起人们足够的重视.早期的一些研究,如文献[10,11]仅考虑用户-位置签到关系,提出了以位置为中心的社区检测方法,并且以非重叠的社区结构为主.文献[12]也对 LBSN 中的用户签到数据运用传统的 Louvain 算法来挖掘位置网络形成的社区结构及位置社区内的用户行为模式.文献[13]提出一种非重叠的社区发现模型——GeoScop,该方法仅依赖于用户签到数据而不考虑社交关系,首先用 DBSCAN 算法产生初始化具有地理位置特征的社交群,然后运用最大期望方法来获得优化的社区结果.为了考察位置因素对社区结构的影响,文献[14,15]采取将用户位置信息和空间距离融合到现有典型的社区检测算法,如模块度、CPM^[16]等中用以挖掘地理关联的非重叠社区.总体而言,上述研究都仅考虑了 LBSN 中的用户签到关系数据,并仍借用传统的社区发现方法来挖掘具有位置特征的社区结构,这些方法存在明显的局限性,如果忽略 LBSN 中多维异构关系的融合,会导致社区发现的结果严重不合理.

最近,一些学者的共识是将 LBSN 建模为多模异构网络,并提出了融合用户、位置、媒体等多模实体及其多维关系的社区发现新方法,以期能够准确、高效地挖掘出更加复杂的社区结构.如 Zhao 等人^[17]首先将 LBSN 建模成多模异构超图结构,超图中顶点有 4 模实体:用户、地点、评论、图片,超边表示实体间的关系,共有 4 种

类型:用户位置签到关系、用户在位置上发表评论关系、用户在位置上传图片关系、位置与位置相似关系.然后,提出基于稠密子图的软聚类算法来发现重叠社区,条件是要求稠密子图中超边权重能达到最大化.文献[18]提出基于概率模型的社区发现方法——rLinkTopic,综合考虑了 3 类属性:地理位置、上下文链接(社交关系)、主题消息.该方法侧重于从语义分析角度挖掘具有地理位置特征的重叠主题社区,得到单一结构类型的社区.另外,还有少数研究采用联合聚类方法来解决 LBSN 中多维异构关系的融合难题,文献[19,20]提出了以边为中心的联合聚类框架发现重叠结构的用户社区,既考虑了用户之间的社交关系,又考虑了用户与地点的签到连接,同时还加入了这两种实体的属性信息,虽然有效提高了社区发现质量,却忽视了由 LBSN 中多维关系数据稀疏造成的计算量较大的问题.为了克服当前社区发现中的数据稀疏问题,现有经典的矩阵分解方法,比如 SVD^[21]、NMF^[4,22]等是应对该问题的有效手段,其中,文献[4]提出了基于 NMF 的半监督社区发现方法,它在网络拓扑关系中融入了先验信息以提高社区检测性能,但该方法也仅能有效地应对单维数据的处理.由此可见,现有传统的基于矩阵分解的社区发现都是属于单向的聚类方法,仅针对单维数据处理有效.而在文献[23]中提出基于块值分解(block value decomposition,简称 BVD)的联合聚类方法则实现了两个维度关系数据的聚类,其给定的最小化目标函数 f 为

$$f(R, B, C) = \|Z - \mathbf{RBC}\|^2 \quad (1)$$

式(1)中, Z 代表原始矩阵, B 表示块值矩阵, R 和 C 分别为行系数和列系数矩阵.

受此启发,本文对该思路加以改进、扩展,使之适用于 LBSN 中融合多维关系的复杂社区挖掘,这种基于矩阵分解的联合聚类方法不仅能够有效地融合 LBSN 中用户-用户社交关系、地理-地理位置相关关系和用户-地理签到关系等多维度关系,还能借助矩阵分解技术较好地解决这些关系存在的数据稀疏问题,仿真实验结果表明,本文提出的针对 LBSN 的社区发现方法——Multi-BVD 能够有效地提高社区发现的质量和效率,所获得的具有地理特征的用户社区结构能够更好地体现出用户社区与地理位置标签间的兴趣关联性.

2 LBSN 网络建模及其用户社区发现方法

2.1 相关定义与形式化描述

LBSN 是一种由多维异构关系复合而成的新型网络,除了包含社交网络中的用户-用户的社交关系外,还有用户在地理空间上的签到行为形成用户-位置签到关系以及地理空间中的地理位置-位置相关关系.为了更好地发现 LBSN 中的复杂社区结构,本文首先给出 LBSN 的网络模型相关定义.

定义 1. 用户-位置关系网络(P).用户在地理位置上的签到行为形成的关系网络可表示成二部图结构,记为 $P=(U, L, C)$,其中, U 表示用户集, L 表示地理位置集, C 表示用户与位置间的签到关系集,即

$$C = \{(u_i, l_j) | u_i \in U, l_j \in L, U \cap L = \emptyset\}.$$

定义 1 表明,用户-位置关系网络实际上是一种仅包含签到关系的二模网络,体现了两类不同实体,即用户与地理位置间的联系,而对于同类实体内在的关系结构则分别被定义成如下两种网络图.

定义 2. 地理位置标签网络(G).地理位置标签网络可抽象成一个加权无向图 G ,记为 $G = (L, R, W)$,其中, L 为地理位置标签集合,即 $L = \{l_1, \dots, l_n\}$, R 表示地理位置间的关系边集,即 $R = \{e_{l_i l_j} | l_i, l_j \in L\}$,其中,任意边 $e_{l_i l_j} = (l_i, l_j)$ 表示无序对,亦即 $(l_i, l_j) \equiv (l_j, l_i)$,而 W 表示图中关系边的权重集合,即 $W = \{w(e_{l_i l_j}) | e_{l_i l_j} \in R\}$,其中,边权重 $w(e_{l_i l_j})$ 是指在边 $e_{l_i l_j}$ 的两端顶点 l_i 和 l_j 上共同签到的用户比重,如式(2)所示.

$$w_{l_i l_j} = \frac{|U_{(C_{l_i})} \cap U_{(C_{l_j})}|}{|U_{(C_{l_i})}| + |U_{(C_{l_j})}|} \quad (2)$$

式(2)中, $U_{(C_{l_i})}$ 表示所有在地理标签 l_i 上签到的用户集合,下标 C_{l_i} 类似于定义 1 中用户与某位置 l_i 的签到关系集合,从中提取出用户集即得 $U_{(C_{l_i})} = \bigcup_{(u_x, l_i) \in C_{l_i}} u_x$.

上述定义将地理位置之间的相关度表示成用户签到关系的交集,体现了相似用户的兴趣点,同时,地理位置间连接边的权值越大,表明地理位置标签间的关联性越强.

定义 3. 用户社交网络(S).用户社交网络可由无权有向图的形式表示,记为 $S=(U,O)$, 其中, U 表示用户集, O 表示用户社交关系的边集,即 $O=\{(u_i,u_j)|u_i,u_j\in U\}$, 其中,有序对 $\langle u_i,u_j \rangle$ 表示从 u_i 指向 u_j 的有向边.

在此基础上,本文进一步给出融合上述 3 种关系网络图的异构复合网络模型,即基于位置的社交网络的形式化定义.

定义 4. 基于位置的社交网络(L_N).基于位置的社交网络可表示成一种多模异构关系的复合网络图,记为 $L_N=S\times P\times G=(U,O)\times(U,L,C)\times(L,R,W)=(\langle U,L\rangle,O,C,R_w)$, 其中,包含了两类实体:用户集 U 和地理位置集 L ,3 种维度关系:用户与用户间的社交关系集 O 、用户与位置的签到关系集 C 、带权重的位置与位置关系集 R_w .

由定义 4 可知,LBSN 是由 3 种关系网络复合而成,其中,传统社交网络与用户-位置关系网络间的融合是以用户实体为主键进行关联的,而用户-位置关系网络与地理位置标签网络间的融合则是以地理位置为主键进行关联的.如图 1 所示 LBSN 的复合关系网络结构,图中分别以用户和地理位置两类对象为主体组成上下两层同质的关系结构,上层为传统社交网络,下层为地理位置标签网络,层级之间通过用户-位置签到行为建立起异构实体间的联系.

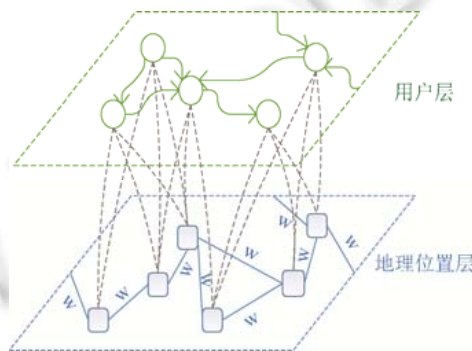


Fig.1 Structure of composite network relations in LBSN

图 1 LBSN 中的复合关系网络结构

基于此模型结构,本文进一步将定义 1~定义 3 的图结构分别转换为对应的邻接矩阵表示形式.首先,假设用户矩阵 $\mathbf{U}=\{u_{ij}\}_{M\times M}$ 表示社交网络关系的 M 阶方阵,用户 i 与用户 j 间的关系值 u_{ij} 如式(3)所示.

$$u_{ij} = \begin{cases} 1, & u_i \text{ follows } u_j \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

其次,地理标签矩阵 $\mathbf{L}=\{w_{ij}\}_{N\times N}$ 表示地理位置标签网络的 N 阶方阵,其中, w_{ij} 是地理位置标签 i 与 j 的关联强度,由式(2)得出.不难发现,该邻接矩阵是对称矩阵,即对任意的 i 和 j 均有 $w_{ij}=w_{ji}$.

然后,用户-位置签到关系矩阵可表示成 $\mathbf{A}=\{\alpha_{ij}\}_{M\times N}$ 的 $M\times N$ 阶矩阵,元素 α_{ij} 表示用户 i 在地理位置 j 上签到的概率,且满足条件 $\sum_j \alpha_{ij}=1$.如果将矩阵每列地理标签看作一类兴趣点,则矩阵 \mathbf{A} 中的每一行代表某用户在各兴趣点上的分布,该行元素组成的向量 $\mathbf{A}_i=[\alpha_{i1},\dots,\alpha_{ij}]$ 称为用户 i 的兴趣向量.由此,以上邻接矩阵 \mathbf{U} 、 \mathbf{L} 和 \mathbf{A} 分别表示了用户-用户社交关系、地理位置-位置关系和用户-位置签到关系等多维度关系.

2.2 融合多维异构关系的联合聚类社区发现方法

联合聚类的基本原理是通过行聚类和列聚类两个步骤进行迭代循环,直至收敛,本文在行聚类上实现用户社区划分,假设用户可被划分为 K_1 个社区,令矩阵 $\mathbf{R}=\{\rho_{ij}\}_{M\times K_1}$ 表示 M 行用户与 K_1 个社区的隶属关系,其

中, $\rho_{ij} \in [0,1]$, 若 $\rho_{ij} = 1$ 表示用户 i 只属于社区 j , 而 $\rho_{ij} = 0$ 表示用户 i 不属于社区 j , 对于任意行用户 i 都满足条件 $\sum_{j=1}^{K_1} \rho_{ij} = 1$. 在列聚类上, 假设地理位置标签可被聚类成 K_2 个簇, 令矩阵 $\mathbf{C} = \{\chi_{ij}\}_{K_2 \times N}$ 表示 N 列地理标签与 K_2 个簇的关系, 其中, $\chi_{ij} \in [0,1]$, 且 $\forall j \in N$, 都有条件 $\sum_{i=1}^{K_2} \chi_{ij} = 1$. 由此, 矩阵 \mathbf{R} 与 \mathbf{C} 分别在用户层和地理位置层构建了 $M \times K_1$ 与 $K_2 \times N$ 对应聚类实体的特征向量空间.

为了融合从行与列聚类的矩阵向量 \mathbf{R} 和 \mathbf{C} , 本文通过用户-位置签到关系矩阵 \mathbf{A} 建立两者间的关联, 同时再引入一个调和矩阵 \mathbf{B} 来解决拟合过程中 \mathbf{R}, \mathbf{C} 特征维度与 \mathbf{A} 不一致的问题. 设矩阵 $\mathbf{B} = \{\beta_{ij}\}_{K_1 \times K_2}$ 表示第 i 个用户社区和第 j 个地理标签簇之间的关系紧密程度, 其中, $\beta_{ij} \in [0,1]$, β_{ij} 越大, 表示二者之间的关系越紧密. 类似地, 本文还引入矩阵 $\mathbf{D} = \{\delta_{ij}\}_{K_1 \times K_1}$, 表示用户社区和用户社区之间的关联性, 以及矩阵 $\mathbf{E} = \{\varepsilon_{ij}\}_{K_2 \times K_2}$, 表示地理标签簇与地理标签簇之间的关联性. 在上述思路的引导下, 本文给出基于矩阵分解的联合聚类方法以发现非重叠的用户社区结构, 其融合了多维异构关系的社区划分目标函数 F , 如式(4)所示:

$$\min F = \min_{\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}} (\|\mathbf{A} - \mathbf{RBC}\|^2 + \|\mathbf{U} - \mathbf{RDR}^T\|^2 + \|\mathbf{L} - \mathbf{CEC}^T\|^2) \quad (4)$$

s.t. $\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E} \geq 0$ 且 $K_1 \ll M, K_2 \ll N$; 其中, $\|\cdot\|$ 表示Frobenius范数

式(4)中, 目标函数的最小代价由 3 种维度关系的矩阵分解确定, 多项式第 1 项体现了由用户-位置签到关系建立 K_1 个用户社区与 K_2 个地理标签簇间的关联性, 其中的 \mathbf{R} 与 \mathbf{C} 分别受后两维关系聚类结果的约束, 第 2 项表示矩阵向量 \mathbf{R} 是由用户社交关系形成的社区划分, 而第 3 项表示矩阵向量 \mathbf{C} 是对地理位置关系聚类而成的地理标签簇.

为求解该目标函数, 设 $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ 分别为 $\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ 的拉格朗日乘子, 且满足

$$\lambda_1 \in \mathfrak{R}^{M \times K_1}, \lambda_2 \in \mathfrak{R}^{K_1 \times K_2}, \lambda_3 \in \mathfrak{R}^{K_2 \times N}, \lambda_4 \in \mathfrak{R}^{K_1 \times K_1}, \lambda_5 \in \mathfrak{R}^{K_2 \times K_2}.$$

于是, F 对应的拉格朗日函数为

$$\begin{aligned} \text{La} = & F - \text{tr}(\lambda_1 \mathbf{R}) - \text{tr}(\lambda_2 \mathbf{B}) - \text{tr}(\lambda_3 \mathbf{C}) - \text{tr}(\lambda_4 \mathbf{D}) - \text{tr}(\lambda_5 \mathbf{E}) \\ = & \text{tr}(\mathbf{A}^T \mathbf{A} - 2\mathbf{A}^T \mathbf{RBC} + \mathbf{C}^T \mathbf{B}^T \mathbf{R}^T \mathbf{R}^T \mathbf{BC}) + \text{tr}(\mathbf{U}^2 - 2\mathbf{URDR}^T + (\mathbf{RDR}^T)^2) + \\ & \text{tr}(\mathbf{L}^2 - 2\mathbf{LC}^T \mathbf{EC} + (\mathbf{C}^T \mathbf{EC})^2) - \text{tr}(\lambda_1 \mathbf{R}) - \text{tr}(\lambda_2 \mathbf{B}) - \text{tr}(\lambda_3 \mathbf{C}) - \text{tr}(\lambda_4 \mathbf{D}) - \text{tr}(\lambda_5 \mathbf{E}) \end{aligned} \quad (5)$$

式(5)中, 符号 $\text{tr}(\cdot)$ 表示求矩阵的迹. 然后, 对式(5)分别用 $\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ 求偏导并使之等于 0, 可得:

$$\frac{\partial \text{La}}{\partial \mathbf{R}} = -2\mathbf{AC}^T \mathbf{B}^T + 2\mathbf{RBCC}^T \mathbf{B}^T - 2\mathbf{URD} + 2\mathbf{RDR}^T \mathbf{R} \mathbf{D} - \lambda_1 = 0 \quad (6)$$

$$\frac{\partial \text{La}}{\partial \mathbf{B}} = -2\mathbf{R}^T \mathbf{AC}^T + 2\mathbf{R}^T \mathbf{RBCC}^T - \lambda_2 = 0 \quad (7)$$

$$\frac{\partial \text{La}}{\partial \mathbf{C}} = -2\mathbf{B}^T \mathbf{R}^T \mathbf{A} + 2\mathbf{B}^T \mathbf{R}^T \mathbf{RBC} - 2\mathbf{E}^T \mathbf{CL}^T + 2\mathbf{ECC}^T \mathbf{E}^T \mathbf{C} - \lambda_3 = 0 \quad (8)$$

$$\frac{\partial \text{La}}{\partial \mathbf{D}} = -2\mathbf{R}^T \mathbf{UR} + 2\mathbf{R}^T \mathbf{RDR}^T \mathbf{R} - \lambda_4 = 0 \quad (9)$$

$$\frac{\partial \text{La}}{\partial \mathbf{E}} = -2\mathbf{LC}^T \mathbf{C} + 2\mathbf{CC}^T \mathbf{ECC}^T - \lambda_5 = 0 \quad (10)$$

再分别用 $\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ 对上述式(6)~式(10)求 Hadamard 积(\odot), 并代入由式(11)所示的 KKT 条件.

$$\left. \begin{aligned} l_1 \odot \mathbf{R} &= 0 \\ l_2 \odot \mathbf{B} &= 0 \\ l_3 \odot \mathbf{C} &= 0 \\ l_4 \odot \mathbf{D} &= 0 \\ l_5 \odot \mathbf{E} &= 0 \end{aligned} \right\} \quad (11)$$

由此, 可得如下等式(12)~等式(16):

$$(\mathbf{RBCC}^T \mathbf{B}^T + \mathbf{RDR}^T \mathbf{RD}) \odot \mathbf{R} = (\mathbf{AC}^T \mathbf{B}^T + \mathbf{URD}) \odot \mathbf{R} \quad (12)$$

$$\mathbf{R}^T \mathbf{AC}^T \odot \mathbf{B} = \mathbf{R}^T \mathbf{RBCC}^T \odot \mathbf{B} \quad (13)$$

$$(\mathbf{E}^T \mathbf{CL}^T + \mathbf{B}^T \mathbf{R}^T \mathbf{A}) \odot \mathbf{C} = (\mathbf{ECC}^T \mathbf{E}^T \mathbf{C} + \mathbf{B}^T \mathbf{R}^T \mathbf{RBC}) \odot \mathbf{C} \quad (14)$$

$$\mathbf{R}^T \mathbf{RDR}^T \mathbf{R} \odot \mathbf{D} = \mathbf{R}^T \mathbf{UR} \odot \mathbf{D} \quad (15)$$

$$\mathbf{LC}^T \mathbf{C} \odot \mathbf{E} = \mathbf{CC}^T \mathbf{ECC}^T \odot \mathbf{E} \quad (16)$$

在此基础上,可进一步求得 $\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ 的迭代更新公式.

$$\mathbf{R}_{new} = \mathbf{R} \odot \frac{\mathbf{AC}^T \mathbf{B}^T + \mathbf{URD}}{\mathbf{RBCC}^T \mathbf{B}^T + \mathbf{RDR}^T \mathbf{RD}} \quad (17)$$

$$\mathbf{B}_{new} = \mathbf{B} \odot \frac{\mathbf{R}^T \mathbf{AC}^T}{\mathbf{R}^T \mathbf{RBCC}^T} \quad (18)$$

$$\mathbf{C}_{new} = \mathbf{C} \odot \frac{\mathbf{E}^T \mathbf{CL}^T + \mathbf{B}^T \mathbf{R}^T \mathbf{A}}{\mathbf{ECC}^T \mathbf{E}^T \mathbf{C} + \mathbf{B}^T \mathbf{R}^T \mathbf{RBC}} \quad (19)$$

$$\mathbf{D}_{new} = \mathbf{D} \odot \frac{\mathbf{R}^T \mathbf{UR}}{\mathbf{R}^T \mathbf{RDR}^T \mathbf{R}} \quad (20)$$

$$\mathbf{E}_{new} = \mathbf{E} \odot \frac{\mathbf{LC}^T \mathbf{C}}{\mathbf{CC}^T \mathbf{ECC}^T} \quad (21)$$

上述每次迭代过程中 $\mathbf{A}, \mathbf{U}, \mathbf{L}$ 矩阵值是已知的.

由式(17)~式(21)可知, $\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ 矩阵值的计算与其他矩阵彼此间都是相互依赖的,本文采用交互优化求解策略迭代更新各个矩阵的值,直到 \mathbf{L}_a 收敛至设定的阈值 ξ ,由此可得到联合聚类的结果矩阵 \mathbf{R} . 然后,通过对 \mathbf{R} 中每个用户行向量求取 $\max_{j=1 \dots k_1} (\rho_{ij})$,即可确定用户社区归属.具体算法描述如下.

算法 1. Multi-BVD 联合聚类方法.

输入: 3 种维度关系矩阵 $\mathbf{A}, \mathbf{U}, \mathbf{L}$ 和用户社区数 K_1 , 地理标签簇数 K_2 ;

输出: 用户社区集合 $V[1] \dots V[k_1]$.

Step 1. 矩阵初始化: 初始化 \mathbf{R}, \mathbf{C} 矩阵服从均匀分布; 初始化 $\mathbf{B}, \mathbf{D}, \mathbf{E}$ 使其中每个元素的值为 1; $F = +\infty$

Step 2. Repeat:

$F' = F$;

根据式(17)~式(21),依次更新 $\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ 的值;

将更新后的 $\mathbf{R}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ 代入公式(4),计算 F ;

Until $|F' - F| \leq \xi$; //函数收敛至 ξ

Step 3. 分析聚类结果矩阵 \mathbf{R} , 确定用户所属社区:

初始化 K_1 个社区向量 $V[1] \dots V[k_1] = \{\}$;

FOR $i=1$ TO M //对于每个用户行向量

$t=1$;

FOR $j=2$ TO K_1

IF $\rho_{it} < \rho_{ij}$ THEN $t=j$;

ENDFOR

$add(i) \rightarrow V[t]$; //将用户 i 划分到隶属度最大的社区 t 中

ENDFOR

3 仿真实验及结果分析

3.1 实验环境与数据集

为了验证算法的有效性,实验数据集选取两种具有位置特征的社交网络:Yelp 和 Gowalla,基本参数见表 1. 对于 Yelp 数据集格式,主要包含了美国地区的用户、商户信息、用户朋友关系以及位置签到与评价打分数据,而 Gowalla 数据集中则包含了用户的朋友关系与位置签到信息,其中,签到位置上仅含有签到时间、经纬度与地理标号.

Table 1 Parameters of experimental datasets

表1 实验数据集的参数

数据集	用户数	社交关系数	商户数	签到记录数	标签数
Yelp	366 715	2 576 180	61 184	1 569 264	664
Gowalla	196 591	1 900 654	-	6 442 890	-

实验前,首先对数据集进行必要的清理与过滤,在 Yelp 中,我们除去了签到记录少于 20 条的不活跃用户,同时还简化了地理空间上的商户位置标记,将原始用户在商户上的签到记录转换为用户在商户对应的地理标签上的签到记录,最后经过预处理后的 Yelp 数据集中包含 11 341 位活跃用户,118 087 条关注关系和 616 804 条签到记录.而在 Gowalla 中,我们去除了少于 50 条签到记录的用户,然后使用 Hash 函数将地理标识映射成[1,200]区间内的兴趣标签号,最终该数据集中包含 9 562 位活跃用户,84 214 条朋友关系和 1 349 557 条位置签到记录.

我们将本文提出的社区发现算法(Multi-BVD)与文献[23]中的 BVD 算法进行实验效果对比,由于 BVD 算法仅考虑了用户-位置签到关系数据,因而其社区划分的目标函数只包含式(4)中的第 1 项,即 $\min_{R,B,C} \|A - RBC\|^2$. 为了便于分析和比较,Multi-BVD 与 BVD 这两种算法的联合聚类参数 K_1 和 K_2 都设置为 18,所有实验的运行环境为 Intel Core(TM)2 Duo CPU P9700@2.8GHz,8GB 内存,Windows 8.1 64 bit 操作系统,实验数据的预处理由 Python 实现,算法中的矩阵运算工具为 Matlab R2014b.

3.2 实验结果分析

图 2(a)~图 2(d)展现了 Multi-BVD 和 BVD 算法分别在 Yelp 和 Gowalla 数据集上获得的最大用户社区中包含的标签云结构,对于社区的标签云,越频繁出现的标签,呈现的标签云越大,表明用户兴趣的关注度也越高.统计结果发现,在 Yelp 数据集上,算法 Multi-BVD 的最大社区中共有 3 015 个用户和 210 个标签,而算法 BVD 的最大社区包含了 3 497 个用户和 213 个标签,虽然两者共同的标签数有 208 个,但各标签出现的频率有很大差别,Multi-BVD 的社区中出现最多的 3 大标签的兴趣主题为 Shopping、Bakeries 和 Steakhouses,而 BVD 的社区中最主要的 3 大兴趣主题是 Cafes、Seafood 和 Health&Medical.而在 Gowalla 数据集中,算法 Multi-BVD 的最大社区中有 1 767 位用户和 192 个标签,而算法 BVD 的最大社区包含了 2 143 个用户和 200 个标签,对比最流行的 top-6 标签,约有 1/3 不相同.综合比较后发现,Multi-BVD 的社区紧密融合了多维度关系,挖掘出了更清晰、准确的社区主题结构.

为了验证上述结论,图 3(a)~图 3(d)进一步给出了两种算法分别在 Yelp 和 Gowalla 数据集上划分的多维社区结构中用户社区和地理标签簇之间的关联性,由块值矩阵 B 得到,图中横轴为地理标签簇标号,纵轴为用户社区标号,网格中不同灰度的色块表示对应用户社区和地理标签簇之间的关联程度,色块颜色越深代表用户社区与对应的地理标签簇之间的关系越紧密.由图 3 中两种数据集上的对比结果可知,Multi-BVD 算法的联合聚类结果中用户社区与地理标签簇之间不仅存在一对一的联系,还有一对多的联系,而 BVD 算法的划分结果中用户社区与地理标签簇基本上只有一对一的联系.总体而言,Multi-BVD 的社区结构体现了更紧密的关联性和多兴趣主题标签特征.

图 4 显示了 Multi-BVD 算法在两种数据集上挖掘的用户社区之间的联系,由块值矩阵 D 来确定,图中横、纵坐标均为社区标号,色块灰度的深浅表示对应两个社区之间关系的强弱.但是,由于 BVD 算法仅考虑了用户-

签到关系数据而忽略了用户社交关系,因此无法体现用户社区之间的关联性.从图 4 中可以看出,不同深度的色块基本出现在网格的对角线位置上或呈对称分布,表明大多数用户社区内部具有较强的内聚性,并且社区之间也表现出较紧密的关联性.同时,结合图 3(a)与图 3(c)可知,这些有较强内聚性的用户社区都对应有清晰且紧密关联的地理标签簇,由此表明,Multi-BVD 算法发现的社区结构不仅能体现社区间的关联性,还具有更清晰、准确的兴趣标签簇.

图 5 和图 6 分别统计了不同社区规模下的用户数及其所包含的地理标签数.如图 5 所示,在 Yelp 和 Gowalla 数据集上,Multi-BVD 与 BVD 算法所挖掘的各社区用户数都大致相当,显然,这两种算法的社区结构具有长尾分布特征.图 6 进一步给出了图 5 所对应社区包含的标签数量,主要反映社区用户的兴趣范围,由图 6(a)可知,在 Yelp 数据集上两种算法的最大社区里都包含了 210 个左右的兴趣标签,约占总标签数的 1/3,而随着社区规模的减小,Multi-BVD 算法的社区标签数量在后半社区里少于 BVD 算法的社区标签数.而对于图 6(b)所示的 Gowalla 数据集,两种算法的最大社区里包含了 192 个标签,并且随着社区规模的减小而平缓下降,但 Multi-BVD 算法基本上都比 BVD 算法的社区标签数量要多.综合来看,Multi-BVD 算法挖掘的大多数社区的用户人均兴趣标签数都要高于 BVD 算法.



(a) Multi-BVD 算法的标签云:Yelp



(b) BVD 算法的最大标签云:Yelp



(c) Multi-BVD 算法的标签云:Gowalla



(d) BVD 算法的最大标签云:Gowalla

Fig.2 Comparison of cloud tags in maximum community between two algorithms

图 2 两种算法中最大社区的标签云对比

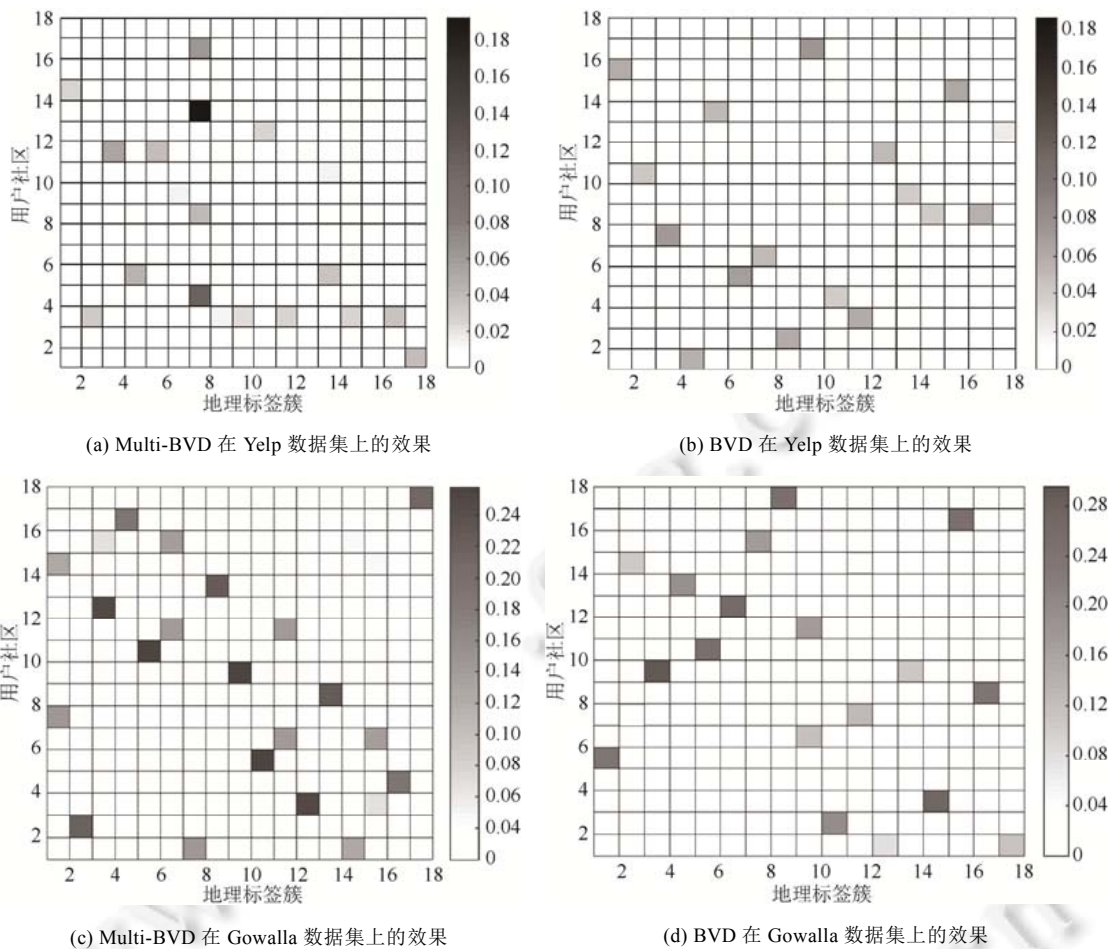


Fig.3 Association between user communities and clusters of geo-tags

图 3 用户社区和地理标签簇之间的关联

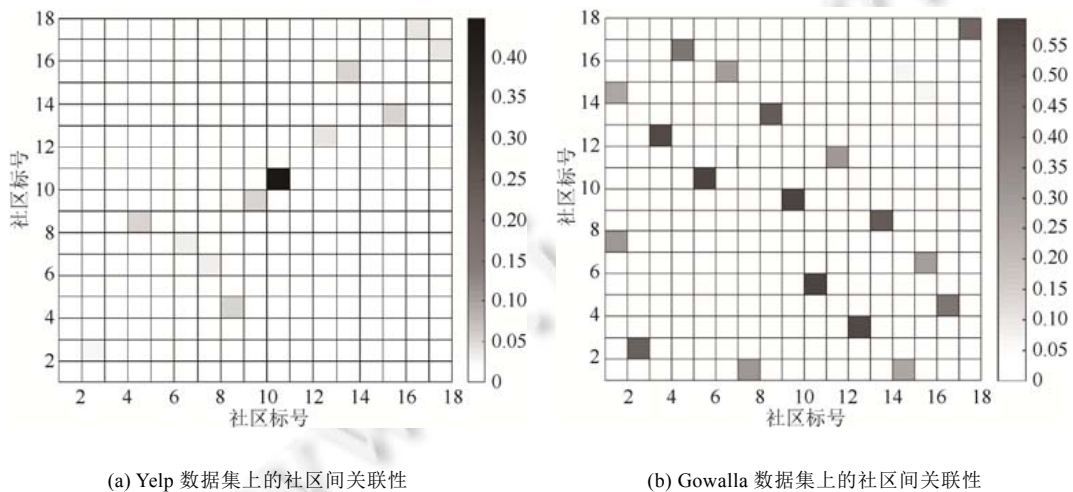


Fig.4 Relationships between user communities by algorithm Multi-BVD on two datasets

图 4 Multi-BVD 算法在两种数据集上挖掘的用户社区之间的联系

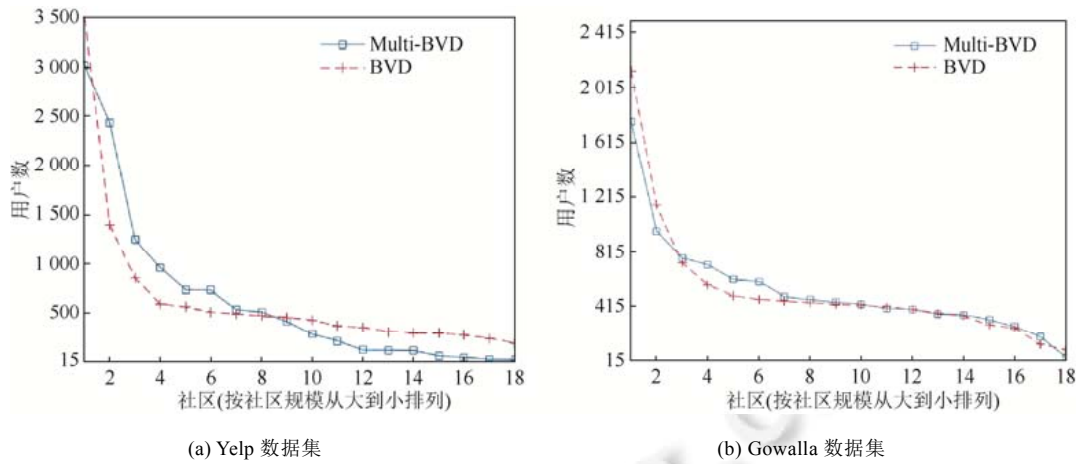


Fig.5 Statistic of the number of users in different community scales

图 5 不同社区规模下的用户数统计

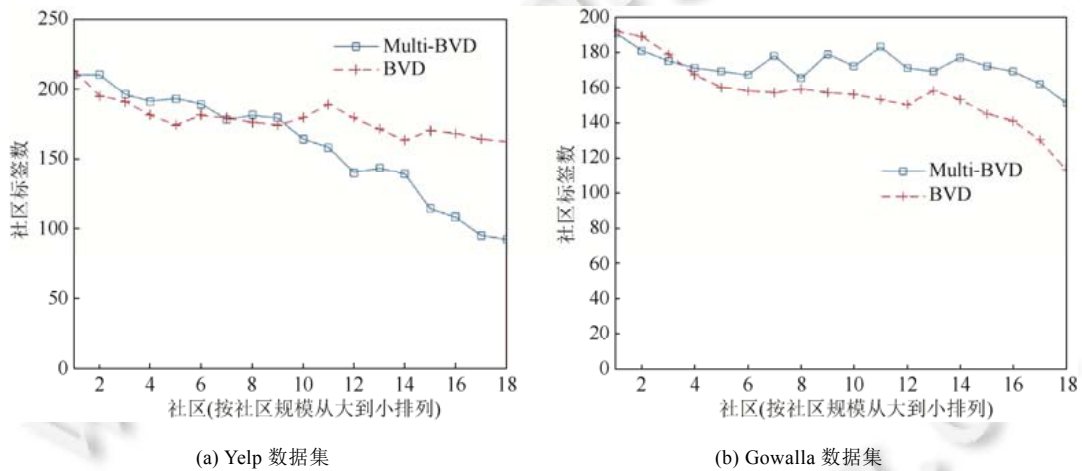


Fig.6 Number of geo-tags in different user communities

图 6 各用户社区所包含的地理标签数

3.2.1 社区结构特征比较

社区结构特征比较主要评价社区内部与社区间两方面的关系结构.首先,社区内部结构特征由用户关系平均度来体现社区内聚性,图 7 给出了 Multi-BVD 与 BVD 算法在两种数据集上的社区用户关系平均度的比较结果,不难发现,Multi-BVD 算法在 Yelp 和 Gowalla 数据集上的社区用户关系平均度分别为 8.8 和 7 左右,而 BVD 算法的各社区内用户关系平均度一般都在 2 以内,远低于前者.由此可见,Multi-BVD 比 BVD 的社区用户社交关系密度更高,具有更好的社区内聚性.

进一步地,对于社区间的特征结构,本文采用社区间亲密度指标来衡量一个社区与其他社区之间联系的紧密性,如定义 5 所示.

定义 5. 社区间亲密度:设某社区内部的用户关系边集合为 E_{in} ,外连边(与其他社区相连的用户关系边)集合为 E_{out} ,则该社区间亲密度可表示为

$$Closeness = \log \left(\frac{|E_{out}|}{|E_{in}|} \right) \tag{22}$$

式(22)中, $|E_{in}|$ 为内部边数量, $|E_{out}|$ 为外连边数量.

定义 5 中,当 $Closeness > 0$ 时,表示社区间的联系比社区内紧密;当 $Closeness < 0$ 时,表示社区内部的联系比社区间的联系更紧密.

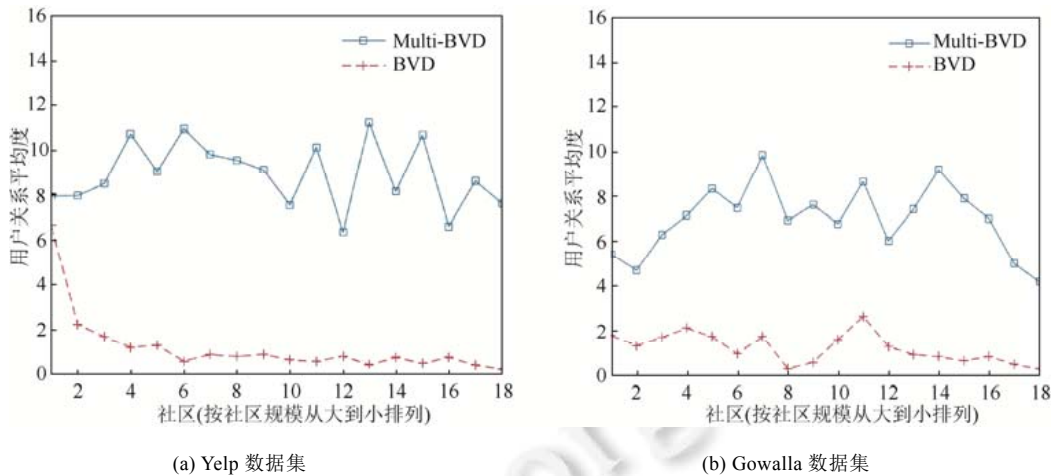


Fig.7 Average degrees of user relations in different inner communities

图 7 不同社区内部的用户关系平均度

图 8 给出了两种算法分别在 Yelp 和 Gowalla 数据集上的社区间亲密度比较结果,Multi-BVD 算法在两种数据集上所划分的各社区间亲密度均小于 0,且随着社区规模的减小而减少,最小值分别达到了-2.69 和-3.38,这验证了 Multi-BVD 算法的社区内部关系比社区间联系更紧密.而 BVD 算法在两种数据集上所得到的各社区间亲密度均大于 0,社区内在联系比较松散,且松散程度随社区规模的减小而增大,最大值分别为 2.25 和 3.2.由此表明,BVD 算法的各社区在用户关系维度上的内聚性较差.

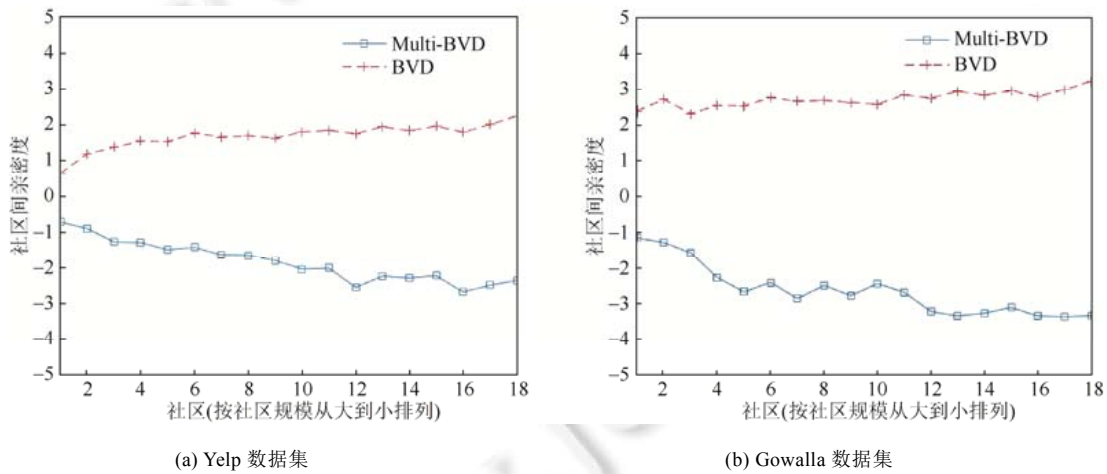


Fig.8 Closeness comparison of inter-communities with different scales

图 8 不同规模的社区间亲密度比较

从另一角度看,模块度^[2,3]是用来评价网络社区划分质量的重要指标,目前在有向网络中的模块度 Q 定义为

$$Q = \frac{1}{m} \sum_{i,j} \left(A_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right) \delta(C_i, C_j) \quad (23)$$

其中, m 表示网络边数, A_{ij} 是节点 i 与 j 的邻接矩阵, k_i^{out} 与 k_j^{in} 分别表示节点 i 的出度和入度.

根据上述公式,表 2 给出了 Multi-BVD 与 BVD 算法分别在两种数据集上划分成不同的用户社区数即 K_1 值变化时的模块度值,比较实验结果发现,Multi-BVD 算法在不同 K_1 值参数下均比 BVD 算法获得了更高的模块度值,社区划分质量较好,由此可见,Multi-BVD 算法能够更准确、合理地挖掘出不同的用户社区结构。

Table 2 Comparison of Q -modularity values

表 2 模块度值 Q 对比

	$K_1=8$		$K_1=16$		$K_1=20$		$K_1=24$	
	Multi-BVD	BVD	Multi-BVD	BVD	Multi-BVD	BVD	Multi-BVD	BVD
Yelp	0.325	0.204	0.341	0.166	0.393	0.208	0.294	0.187
Gowalla	0.286	0.182	0.373	0.151	0.402	0.213	0.328	0.195

3.2.2 社区兴趣相似度分析

更进一步地,本文采用标签 Jaccard 相似度来衡量社区内存在社交关系的用户间在签到关系上表现出共同兴趣标签的相似程度,它是描述社区内所有关联用户的共同兴趣指数,具体定义如下。

定义 6. 标签相似度:设某社区 C 内的用户社交关系边集为 E_c ,对于任意边 $\forall e_{ij} \in E_c$ 两端的用户 u_i 与 u_j ,其对应签到关系上的标签集合分别为 L_{u_i} 和 L_{u_j} ,则该社区的标签 Jaccard 相似度为

$$J = \frac{1}{|E_c|} \sum_{e_{ij} \in E_c} \frac{|L_{u_i} \cap L_{u_j}|}{|L_{u_i} \cup L_{u_j}|} \quad (24)$$

图 9 显示了 Multi-BVD 与 BVD 算法分别在两种数据集上取 $K_1=18$ 时的各用户社区所包含的兴趣标签相似度情况,不难看出,Multi-BVD 比 BVD 算法在不同社区规模下的兴趣标签相似度都高出 10% 以上,实验结果表明,Multi-BVD 算法的社区内用户间拥有更高的兴趣标签内聚性,保持了较好的兴趣标签结构。

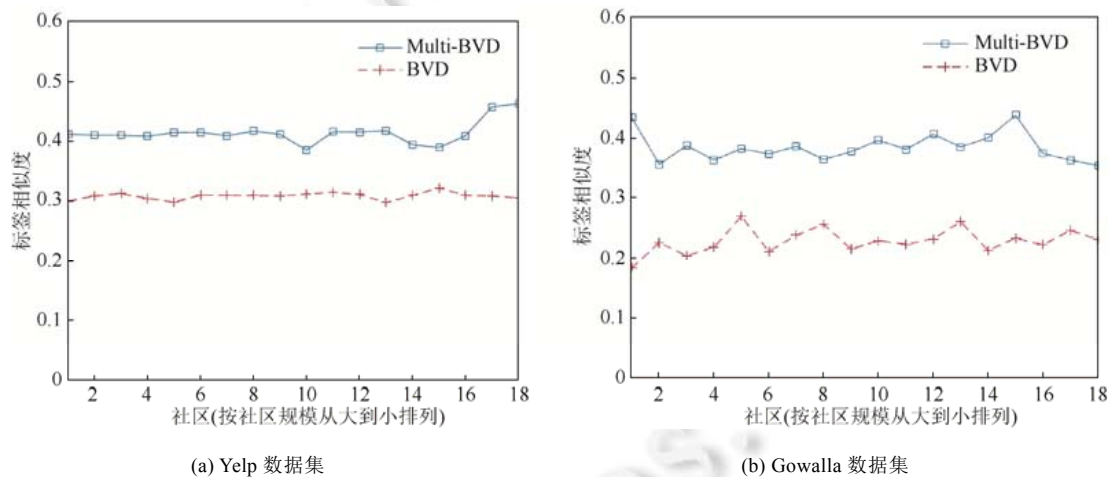


Fig.9 Similarity comparison of interest tags in different communities

图 9 各社区的兴趣标签相似度比较

4 结束语

本文开展针对 LBSN 这类具有多维异构关系网络中的复杂社区结构的研究,解决 LBSN 的社区划分中多维关系的关联性融合问题.文中首先给出了 LBSN 这种新型网络模型的形式化表示方法,并分析其所含 3 种关系网络的结构特征及关联性.然后,在此基础上,提出了一种基于联合聚类的非重叠用户社区发现方法——Multi-BVD,该方法首先设计出融合了用户-用户社交关系、地理位置-位置相关关系及用户-地理位置签到关系等多维关系的社区划分目标函数.为了求解该目标函数的最小代价值,我们采用拉格朗日乘子法得到矩阵变量的迭代更新规则,并运用块值矩阵分解技术获得函数收敛到最小值的用户隶属矩阵,以此确定最优的社区划分

结果.最后,实验结果表明,我们提出的联合聚类方法 Multi-BVD 能够有效地发现 LBSN 中具有地理特征的用户社区结构.在未来的工作中,我们将考虑对联合聚类的参数进行优化,以期能够针对不同数据集自动获得理想的社区划分簇数及社区结构.

References:

- [1] Girvan M, Newman MEJ. Community structure in social and biological networks. *Proc. of the National Academy of Sciences*, 2002, 12(99):7821–7826. [doi: 10.1073/pnas.122653799]
- [2] Newman MEJ. Fast algorithm for detecting community structure in networks. *Physical Review E*, 2004,6(69):1–5. [doi: 10.1103/PhysRevE.69.066133]
- [3] Liu Y, Kang XH, Gao H, Liu Q, W ZF, Qin ZG. A community detecting method based on the node intimacy and degree in social network. *Journal of Computer Research and Development*, 2015,52(10):2363–2372 (in Chinese with English abstract) [doi: 10.7544/issn1000-1239.2015.20150407]
- [4] Yang L, Cao X, Jin D, Wang X, Meng D. A unified semi-supervised community detection framework using latent space graph regularization. *IEEE Trans. on Cybernetics*, 2015,45(11):2585–2598. [doi: 10.1109/TCYB.2014.2377154]
- [5] Raghavan UN, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 2007,76(3):1–11. [doi: 10.1103/PhysRevE.76.036106]
- [6] Hu Y, Wang CJ, Wu J, Xie JY, Li H. Overlapping community discovery and global representation on microblog network. *Ruan Jian Xue Bao/Journal of Software*, 2014,25(12):2824–2836 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4721.htm> [doi: 10.13328/j.cnki.jos.004721]
- [7] Ahn YY, Bagrow JP, Lehmann S. Link communities reveal multi scale complexity in networks. *Nature*, 2010,466(7307):761–764. [doi: 10.1038/nature09182]
- [8] Zhou XP, Liang X, Zhang HY. User community detection on micro-blog using R-C model. *Ruan Jian Xue Bao/Journal of Software*, 2014,25(12):2808–2823 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4720.htm> [doi: 10.13328/j.cnki.jos.004720]
- [9] Sun YF, Li S. Similarity-Based community detection in social network of microblog. *Journal of Computer Research and Development*, 2014,51(12):2797–2807 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2014.20131209]
- [10] Brown C, Nicosia V, Scellato S, Noulas A, Mascolo C. The importance of being place friends: Discovering location focused online communities. In: *Proc. of the 2012 ACM Workshop on Online Social Networks (WOSN 2012)*. New York: ACM Press, 2012. 31–36. [doi: 10.1145/2342549.2342557]
- [11] Noulas A, Scellato S, Mascolo C, Pontil M. Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. In: *Proc. of the ICWSM*. AAAI Press, 2011. 32–35. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/3845/4388>
- [12] Ding X, Xu J, Chen GL. Exploring structural analysis of place networks using check-in signals. In: *Proc. of the IEEE Global Communications Conf. (GLOBECOM)*. Piscataway: IEEE Press, 2013. 3194–3199. [doi: 10.1109/GLOCOM.2013.6831563]
- [13] Srivastava S, Pande S, Ranu S. Geo-Social clustering of places from check-in data. In: *Proc. of the IEEE Int'l Conf. on Data Mining*. Washington: IEEE Computer Society, 2015. 985–990. [doi: 10.1109/ICDM.2015.16]
- [14] Nguyen T, Chen M, Szymanski BK. Analyzing the proximity and interactions of friends in communities in gowalla. In: *Proc. of the IEEE 13th Int'l Conf. on Data Mining Workshops*. Washington: IEEE Computer Society, 2013. 1036–1044. [doi: 10.1109/ICDMW.2013.60]
- [15] Liu Z, Huang Y. Community detection from location-tagged networks. In: *Proc. of the 22nd ACM SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems (SIGSPATIAL 2014)*. New York: ACM Press, 2014. 525–528. [doi: 10.1145/2666310.2666496]
- [16] Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 2005,435(7043):814–818. [doi: 10.1038/nature03607]

- [17] Zhao YL, Chen Q, Yan S, Chua TS, Zhang D. Detecting profilable and overlapping communities with user-generated multimedia contents in LBSNs. *ACM Trans. on Multimedia Computing, Communications, and Applications (TOMM)*, 2013,10(1):1–22. [doi: 10.1145/2502415]
- [18] Van Canh T, Gertz M. rLinkTopic: A probabilistic model for discovering regional LinkTopic communities. In: *Proc. of the IEEE Int'l Conf. on Advances in Social Networks Analysis and Mining (ASONAM)*. Piscataway: IEEE Press, 2014. 21–26. [doi: 10.1109/ASONAM.2014.6921555]
- [19] Wang Z, Zhou X, Zhang D, Yang D, Yu Z. Cross-Domain community detection in heterogeneous social networks. *Personal and Ubiquitous Computing*, 2014,18(2):369–383. [doi: 10.1007/s00779-013-0656-0]
- [20] Wang Z, Zhang D, Zhou X, Yang D. Discovering and profiling overlapping communities in location-based social networks. *IEEE Trans. on Systems Man & Cybernetics Systems*, 2014,44(4):499–509. [doi: 10.1109/TSMC.2013.2256890]
- [21] Drineas P, Frieze A, Kannan R, Vempala S, Vinay V. Clustering large graphs via the singular value decomposition. *Machine Learning*, 2004,56:9–33. [doi: 10.1023/B:MACH.0000033113.59016.96]
- [22] Xu W, Liu X, Gong Y. Document clustering based on non-negative matrix factorization. In: *Proc. of the 26th Annual Int'l ACM Conf. on Research and Development in Information Retrieval*. New York: ACM Press, 2003. 267–273. [doi: 10.1145/860435.860485]
- [23] Long B, Zhang Z, Yu PS. Co-Clustering by block value decomposition. In: *Proc. of the 11th ACM SIGKDD Int'l Conf. on Knowledge Discovery in Data Mining*. New York: ACM Press, 2005. 635–640. [doi: 10.1145/1081870.1081949]

附中文参考文献:

- [3] 刘瑶,康晓慧,高红,刘昕,吴祖峰,秦志光.基于节点亲密度和度的社会网络社团发现方法. *计算机研究与发展*,2015,52(10):2363–2372. [doi: 10.7544/issn1000-1239.2015.20150407]
- [6] 胡云,王崇骏,吴骏,谢俊元,李慧.微博网络上的重叠社群发现与全局表示. *软件学报*,2014,12(25):2824–2836. <http://www.jos.org.cn/1000-9825/4721.htm> [doi: 10.13328/j.cnki.jos.004721]
- [8] 周小平,梁循,张海燕.基于 R-C 模型的微博用户社区发现. *软件学报*,2014,12(25):2808–2823. <http://www.jos.org.cn/1000-9825/4720.htm> [doi: 10.13328/j.cnki.jos.004720]
- [9] 孙怡帆,李赛.基于相似度的微博社交网络的社区发现方法. *计算机研究与发展*,2014,51(12):2797–2807. [doi: 10.7544/issn1000-1239.2014.20131209]



龚卫华(1977—),男,湖北武汉人,博士,副教授,CCF 专业会员,主要研究领域为复杂网络,数据挖掘.



裴小兵(1971—),男,博士,副教授,CCF 专业会员,主要研究领域为机器学习,数据挖掘,电信网络管理.



陈彦强(1990—),男,硕士,主要研究领域为社交网络,数据挖掘.



杨良怀(1967—),男,博士,教授,CCF 专业会员,主要研究领域为数据挖掘,闪存数据库.