

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.

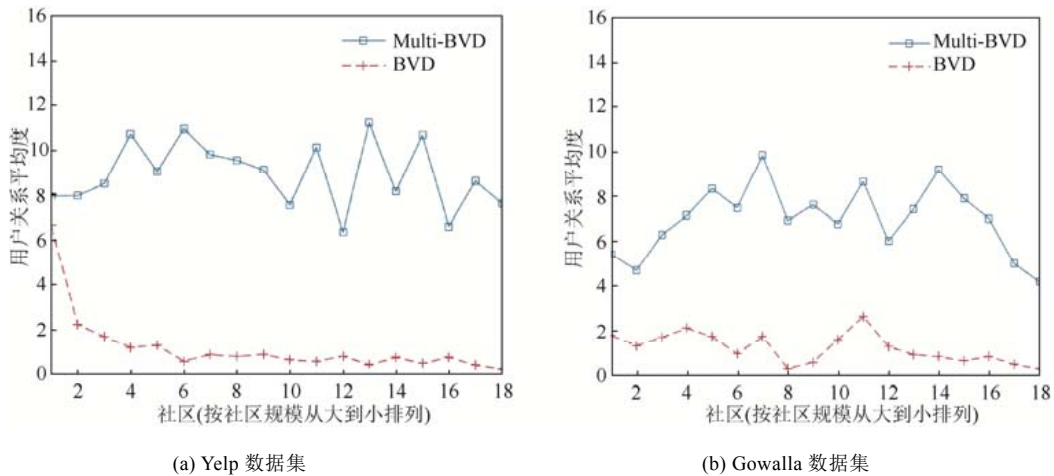


图 7 不同社区内部的用户关系平均度

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.

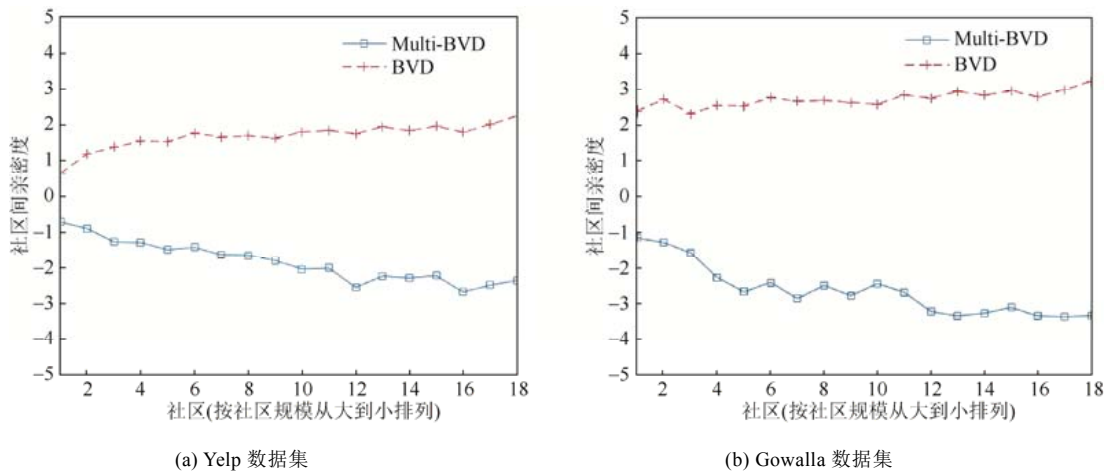


图 8 不同规模的社区间亲密度比较

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

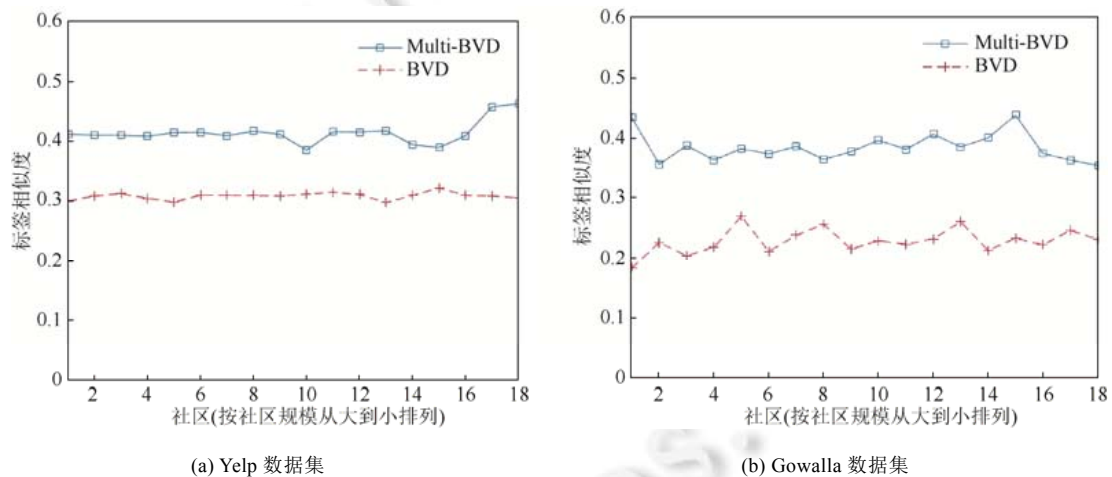
3.2.2 社区兴趣相似度分析

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。



(a) Yelp 数据集

(b) Gowalla 数据集

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.