



















表7 真实社交网络数据集

名称	节点数	边数	真实社区数	网络描述
Karate	34	78	2	Zachary空手道俱乐部
Dolphins	62	159	2	海豚网络
Football	115	613	12	美国大学足球

Karate 为美国空手道俱乐部跆拳道俱乐部的真实划分。

Dolphin 数据集是 D. Lusseau 等人使用长达 7 年的时间观察新西兰 Doubtful Sound 海峡 62 只海豚群体的交流情况而得到的海豚社会关系网络。这个网络具有 62 个节点, 159 条边。节点表示海豚, 而边表示海豚间的频繁接触, 该图为无权图。

Football 网络, 根据美国大学生足球联赛而创建的一个复杂的社会网络。该网络包含 115 个节点和 616 条边, 其中网络中的结点代表足球队, 两个结点之间的边表示两只球队之间进行过一场比赛。参赛的 115 支大学生代表队被分为 12 个联盟。比赛的流程是联盟内部的球队先进行小组赛, 然后再是联盟之间球队的比赛。

此处选择了标准化互信息 (NMI) 评价指标来衡量算法得到的社区划分结果与实际社区的相似性分区结果比较。NMI 的计算公式如下:

$$NMI = \frac{-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} C_{ij} \log_2(C_{ij}N/C_i C_j)}{\sum_{i=1}^{C_A} C_i \log_2(C_i/N) + \sum_{j=1}^{C_B} C_j \log_2(C_j/N)} \quad (16)$$

其中,  $A$  和  $B$  代表社区网络的两个分区,  $C$  是混淆矩阵, 混淆矩阵  $C$  中的元素  $C_{ij}$  表示社区  $i$  除以  $A$  和社区  $j$  除以  $B$  的节点数。  $C_A(C_B)$  表示  $A(B)$  分区中的社区数,  $C_i(C_j)$  是混淆矩阵  $C$  中第  $i$  行 ( $j$  列) 元素的和,  $N$  是原始社区网络中的节点总数。当 NMI 值为 1 时, 表示  $A$  和  $B$  在社区网络中的划分相同。

由于 3 个真实社区数据集不含属性信息, 此处采用 Louvain 算法、DEMON 算法和 DOCET 算法与提出的 TPCDLP 算法进行比较。实验结果如表 8 所示: 在海豚数据集 (dolphins) 上, 本文提出的 MIFCD 算法的 NMI 值最高; 在空手道数据集 (karate) 上, 本文提出的 TPCDLP 算法的 NMI 值优于 DEMON 算法; 在足球数据集 (football) 上, TPCDLP 表现好于 DOCET 算法。可以看出, TPCDLP 能够基本实现真实社区划分。

表8 归一化互信息评价指标的实验结果

数据	Louvain	DEMON	DOCET	TPCDLP
Karate	0.687262	0.267277	0.611001	0.354738
Dolphins	0.475323	0.418992	0.406308	0.746055
Football	0.890316	0.672065	0.366826	0.519401

## 5 总结

本文提出了一种基于标签属性的拓扑势社区检测算法。该算法利用标签传播方法构造节点间的链接权重, 保证分割社区中的节点具有紧密的链接, 并保持区域内部属性特征高度一致。由于实际网络数据具有冗余关系、数据存储量大、数据分布离散等特点, 采用拓扑势最高的局部节点作为社区的核心节点进行社区划分的算法容易导致社区重叠度高、数量多, 因此, 在划分子社区之后, 利用子节点与属性特征之间的距离划分社区, 在保证社区节点之间的链接紧密性和属性相关性的同时, 能够解决细粒度独立社区问题。

## 参考文献

- Fazil M, Abulaish M. A hybrid approach for detecting automated spammers in twitter. *IEEE Transactions on Information Forensics and Security*, 2018, 13(11): 2707–2719. [doi: 10.1109/TIFS.2018.2825958]
- Sánchez-Oro J, Duarte A. Iterated Greedy algorithm for performing community detection in social networks. *Future Generation Computer Systems*, 2018, 88: 785–791. [doi: 10.1016/j.future.2018.06.010]
- Win HN, Lynn KT. Community detection in Facebook with outlier recognition. *Proceedings of the 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. Kanazawa, Japan. 2017.155–159.
- Chen YL, Chuang CH, Chiu YT. Community detection based on social interactions in a social network. *Journal of the Association for Information Science and Technology*, 2014, 65(3): 539–550. [doi: 10.1002/asi.22986]
- Sun XL, Lin HF. Topical community detection from mining user tagging behavior and interest. *Journal of the Association for Information Science and Technology*, 2013, 64(2): 321–333.
- Zhong C, Arisona SM, Huang XF, et al. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*,

- 2014, 28(11): 2178–2199. [doi: [10.1080/13658816.2014.914521](https://doi.org/10.1080/13658816.2014.914521)]
- 7 Girvan M, Newman MEJ. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(12): 7821–7826. [doi: [10.1073/pnas.122653799](https://doi.org/10.1073/pnas.122653799)]
- 8 贺超波, 汤庸, 刘海, 等. 一种集成链接和属性信息的社区挖掘方法. *计算机学报*, 2017, 40(3): 601–616.
- 9 李德仁, 王树良, 李德毅. *空间数据挖掘理论与应用*. 3版. 北京: 科学出版社, 2019.
- 10 淦文燕, 赫南, 李德毅, 等. 一种基于拓扑势的网络社区发现方法. *软件学报*, 2009, 20(8): 2241–2254.
- 11 Wang ZX, Li ZC, Yuan G, Y *et al*. Tracking the evolution of overlapping communities in dynamic social networks. *Knowledge-Based Systems*, 2018, 157: 81–97. [doi: [10.1016/j.knosys.2018.05.026](https://doi.org/10.1016/j.knosys.2018.05.026)]
- 12 刘世超, 朱福喜, 甘琳. 基于标签传播概率的重叠社区发现算法. *计算机学报*, 2016, 39(4): 717–729. [doi: [10.11897/SP.J.1016.2016.00717](https://doi.org/10.11897/SP.J.1016.2016.00717)]
- 13 Li Q, Zhou T, Lü LY, *et al*. Identifying influential spreaders by weighted LeaderRank. *Physica A: Statistical Mechanics and its Applications*, 2014, 404: 47–55. [doi: [10.1016/j.physa.2014.02.041](https://doi.org/10.1016/j.physa.2014.02.041)]
- 14 李东, 程鸣权, 徐杨, 等. 基于平均互信息的最优社区发现方法. *中国科学: 信息科学*, 2019, 49(5): 613–629.
- 15 Blondel VD, Guillaume JL, Lambiotte R, *et al*. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, 2008(10): P10008. [doi: [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008)]
- 16 Prat-Pérez A, Dominguez-Sal D, Larriba-Pey JL. High quality, scalable and parallel community detection for large real graphs. *Proceedings of the 23rd International Conference on World Wide Web*. Seoul, Republic of Korea. 2014. 225–236.
- 17 Coscia M, Rossetti G, Giannotti F, *et al*. Demon: A local-first discovery method for overlapping communities. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Beijing, China. 2012. 615–623.