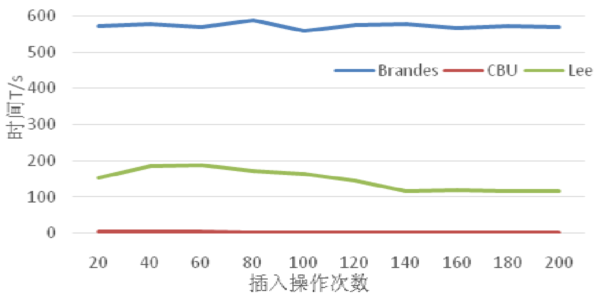

Fig.4 Comparison of the speed after adding edgeson the dataset ben_10000

图 4 在 ben_10000 数据集上插入边效率比较

接下来,我们探究社区过滤效果对于介数中心度更新计算的影响.表 3 展示了 CBU 算法的过滤能力.无论是合成数据集还是真实数据集,社区过滤效果均在 0.25 以下.在合成数据集上,删除操作的社区过滤效果略好,但是无法弥补删除操作后寻找最短路径过程所消耗的时间.在真实数据集 ego-Facebook 上,边添加操作后的社区过滤比高于边删除操作后的社区过滤比;在真实数据集 p2p-Gnutella25 上,边删除操作后的社区过滤比高于边添加操作后的社区过滤比.这与我们之前对于这两个数据集更新操作花费时间的分析是一致的.在真实数据集 gowalla 上,社区过滤效果比较突出,但是由于数据集节点数较大,受影响的点对仍然很多,更新速度的提高实际有限.整体上看,CBU 算法提高了介数中心度的更新速度,其采用的社区过滤策略是有效的.

Table 3 Filtration ratio of community in different datasets

表 3 不同数据集社区过滤比

数据集	ρ(CBU+)	ρ(CBU–)
ER_100000	0.21	0.14
WS_100000	0.13	0.13
BA_100000	0.16	0.06
ben_10000	0.10	0.05
ego-Facebook	0.10	0.05
p2p-Gnutella25	0.07	0.19
gowalla	0.000 3	0.000 7

5.4 参数实验

本节分析节点数量 n 以及社区发现的精度对实验结果的影响.

(1) 节点数量 n

针对 ER 合成图,设置节点的平均度数为 10,分析不同节点数量下,CBU 算法的加速效率.

从图 5 和图 6 可以看出:随着节点数的增加,3 种算法所消耗的时间也在不断增大.但是对于添加操作和删除操作,CBU 算法的时间开销均明显少于 Brandes 算法和 Lee 算法,效率至少提升了一个数量级.这表明,CBU 算法可以很好地提高节点介数中心度的更新效率.

(2) 社区精度的影响

针对表 1 中的合成图 ben_10000,采用不同的社区发现算法,计算出社区结果的精度.然后,在这些不同的社区结果上运行 CBU 算法,分析社区精度对于 CBU 算法效率的影响.

对于社区精度,我们采用的指标是 Modularity,这是 Newman[28]于 2004 年提出的一个定义在[–0.5,1)区间内的指标,由网络图中每个社区内连边数与期待值之差决定.Modularity 越大,社区内部实际连边越是高于随机期望,说明节点越有集中在某些社区内的趋势,即网络的模块化结构越明显,社区划分的精度越高.

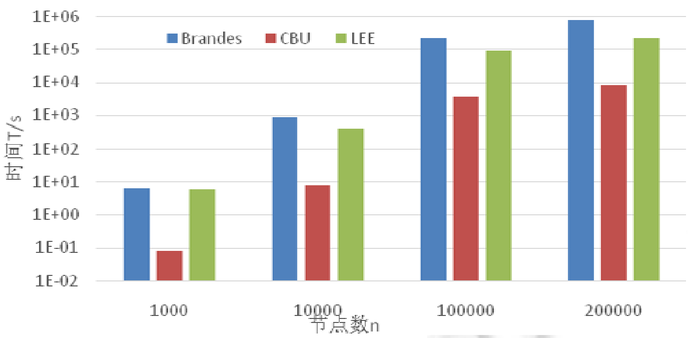

Fig.5 Comparison of the speed after adding an edge

图 5 插入边效率比较

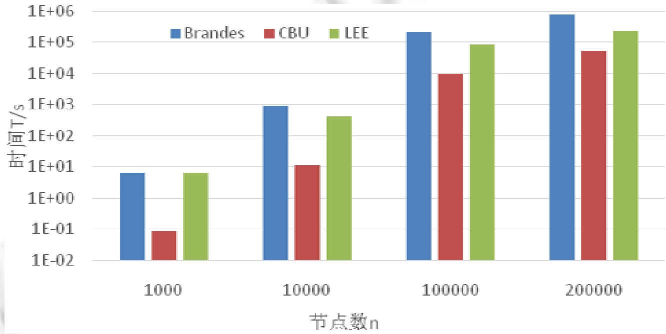

Fig.6 Comparison of the speed after deletingan edge

图 6 删除边效率比较

从表 4 可以看出:随着社区精度的提高,CBU 算法的效率越高,耗时越少.当 Modularity 较大时,提升更为明显,在 CNM 算法得到的社区上进行更新的耗时只有 HANP 的一半.总的来说,社区发现的精度越高,CBU 基于社区的过滤算法效果越好.

Table 4 Efficiency of updating with different communities

表 4 不同社区下的更新效率

社区发现算法	Modularity	时间 T/s	
		CBU(+)/s	CBU(−)/s
DSGE[17]	0.14	8.54	11.82
LPA[15]	0.25	7.54	9.30
HANP[16]	0.55	5.56	7.06
CNM[14]	0.61	3.40	3.13

6 结束语

为了解决在实际应用场景中节点介数中心度更新缓慢的问题,本文提出了基于社区的节点介数中心度更新算法 CBU,通过维护社区间最短距离集合、社区与节点的最短距离集合,快速过滤掉在网络更新过程中最短路径不变的点对,大大提高了节点介数中心度的更新效率.未来,我们将尝试利用社区的性质,对节点介数中心度进行近似计算.

References:

[1] Merrer EL, Tredan G. Centralities: Capturing the fuzzy notion of importance in social graphs. In: Proc. of the European Conf. on Computer Systems. 2009. 33–38. [doi: 10.1145/1578002.1578008]

[2] Bader DA, Kintali S, Madduri K, Mihail M. Approximating betweenness centrality. In: Proc. of the Workshop on Algorithms and Models for the Web Graph. 2007. 124−137. [doi: 10.1007/978-3-540-77004-6_10]

[3] Brandes U. A faster algorithm for betweenness centrality. Journal of Mathematical Sociology, 2001,25(2):163−177. [doi: 10.1080/0022250X.2001.9990249]

[4] Sariyüce AE, Saule E, Kaya K, Çatalyürek ÜV. Shattering and compressing networks for centrality analysis. In: Proc. of the Computer Science. 2012..

[5] Lee MJ, Lee J, Park JY, Choi RH, Chung CW. QUBE: A quick algorithm for updating betweenness centrality. In: Proc. of the Int'l Conf. on World Wide Web. ACM Press, 2012. 351−360. [doi: 10.1145/2187836.2187884]

[6] Lee MJ, Choi S, Chung CW. Efficient algorithms for updating betweenness centrality in fully dynamic graphs. In: Proc. of the Information Sciences. 2016. 278−296. [doi: 10.1016/j.ins.2015.07.053]

[7] Brandes U, Pich C. Centrality estimation in large networks. Int'l Journal of Bifurcation and Chaos, 2007,17(07):2303−2318. [doi: 10.1142/S0218127407018403]

[8] Geisberger R, Sanders P, Schultes D. Better approximation of betweenness centrality. In: Proc. of the Algorithm Engineering and Experimentation. 2008. 90−100. [doi: 10.1137/1.9781611972887.9]

[9] Girvan M, Newman ME. Community structure in social and biological networks. Proc. of the National Academy of Sciences of the United States of America, 2002,99(12):7821−7826. [doi: 10.1073/pnas.122653799]

[10] Huang FL, Zhang SC, Zhu XF. Discovering network community based on multi-objective optimization. Ruan Jian Xue Bao/Journal of Software, 2013,24(9):2062−2077 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4400.htm [doi: 10.3724/SP.J.1001.2013.04400]

[11] Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D. Defining and identifying communities in networks. Proc. of the National Academy of Sciences of the United States of America, 2004,101(9):2658. [doi: 10.1073/pnas.0400054101]

[12] Newman ME. Modularity and community structure in networks. Proc. of the National Academy of Sciences of the United States of America, 2006,103(23):8577−8382. [doi: 10.1073/pnas.0601602103]

[13] Newman ME. Fast algorithm for detecting community structure in networks. Physical Review E, 2004,69(6). [doi: 10.1103/PhysRevE.69.066133]

[14] Clauset A, Newman ME, Moore C. Finding community structure in very large networks. Physical Review E, 2004,70(6). [doi: 10.1103/PhysRevE.70.066111]

[15] Raghavan UN, Albert R, Kumara SR. Near linear time algorithm to detect community structures in large-scale networks. Physical Review E, 2007,76(3). [doi: 10.1103/PhysRevE.76.036106]

[16] Leung IX, Hui P, Lio P, Crowcroft J. Towards real-time community detection in large networks. Physical Review E, 2009,79(6). [doi: 10.1103/PhysRevE.79.066107]

[17] Chen J, Saad Y. Dense subgraph extraction with application to community detection. IEEE Trans. on Knowledge and Data Engineering, 2012,24(7):1216−1230. [doi: 10.1109/TKDE.2010.271]

[18] Huang J, Sun H, Song Q, Deng H, Han J. Revealing density-based clustering structure from the core-connected tree of a network. IEEE Trans. on Knowledge and Data Engineering, 2013,25(8):1876−1889. [doi: 10.1109/TKDE.2012.100]

[19] Perozzi B, Alrfou R, Skiena S. DeepWalk: Online learning of social representations. In: Proc. of the Knowledge Discovery and Data Mining. 2014. 701−710. [doi: 10.1145/2623330.2623732]

[20] Shang JW, Wang CK, Xin X, Ying X. Community detection algorithm based on deep sparse autoencoder. Ruan Jian Xue Bao/Journal of Software, 2017,28(3):648−662 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5165.htm [doi: 10.13328/j.cnki.jos.005165]

[21] Gong M, Li G, Wang Z, Ma L, Tian D. An efficient shortest path approach for social networks based on community structure. CAAI Trans. on Intelligence Technology, 2016,1(1):114−123. [doi: 10.1016/j.trit.2016.03.011]

[22] Cormen TH, Leiserson CE, Rivest RL, Stein C. Introduction to Algorithms. The MIT Press, 2001.

[23] Erdős P, Rényi A. On random graphs I. Publicationes Mathematicae, 1959,6:290−297.

[24] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature, 1998,393(6684):440−442. [doi: 10.1038/30918]

[25] Barabasi A, Albert R. Emergence of scaling in random networks. Science, 1999,286(5439):509−512. [doi: 10.1126/science.286.5439.509]

[26] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. Physical Review E, 2008, 78(4). [doi: 10.1103/PhysRevE.78.046110]

[27] SNAP datasets. http://snap.stanford.edu/data

[28] Newman ME, Girvan M. Finding and evaluating community structure in networks. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2004,69(2):26−113. [doi: 10.1103/PhysRevE.78.046110]

附中文参考文献:

[10] 黄发良,张师超,朱晓峰.基于多目标优化的网络社区发现方法.软件学报,2013,24(9):2062−2077. http://www.jos.org.cn/1000-9825/4400.htm [doi: 10.3724/SP.J.1001.2013.04400]

[20] 尚敬文,王朝坤,辛欣,应翔.基于深度稀疏自动编码器的社区发现算法.软件学报,2017,28(3):648−662. http://www.jos.org.cn/1000-9825/5165.htm [doi: 10.13328/j.cnki.jos.005165]

钱珺(1993－),男,江苏泰州人,硕士生,主要研究领域为社交网络,社区发现.

郭高扬(1994－),男,博士生,主要研究领域为社交网络,深度学习.

王朝坤(1976－),男,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为图和社交数据管理,音乐计算,大数据系统.