Fast label propagation algorithm for community detection

Vincent A. Traag
Leiden University, Centre for Science and Technology Studies, Leiden, The Netherlands
v.a.traag@cwts.leidenuniv.nl

Lovro Šubelj
University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia
lovro.subelj@fri.uni-lj.si

There exist various techniques that try to improve our understanding of networks. One such technique is to cluster the nodes of a network into communities, such that nodes within a community are relatively densely connected while they are relatively sparsely connected between communities. There exists a wide variety of approaches to detect communities in networks, each offering different interpretations and associated algorithms. For large networks, there is the additional requirement of speed.

A technique that takes a heuristic approach is the label propagation algorithm [1] (LPA), which runs in near-linear time. Simply put, LPA works by iteratively updating the community label of each node to a label that is most common among its neighbors. We propose a fast variant of LPA [3] (FLPA), which is based on processing a queue of nodes whose neighborhood recently changed. In partitions found by either LPA or FLPA, we prove that each node is guaranteed to have most links to its assigned community.

We first analyse LPA and FLPA on theoretical graphs such as a star, a cycle and a complete graph. We prove that both algorithms find the same partitions, but that FLPA has lower asymptotic complexity Θ. Next, we thoroughly test FLPA on benchmark graphs and empirical networks (Figure 1). We find that the partitions found by LPA and FLPA are largely comparable, while FLPA can run up to 700 times faster than LPA (Table 1).

Our results show that FLPA is generally preferable to LPA. When using label propagation, we believe our fast variant will bring benefits at no additional costs. We consider

Table 1: Speedup for large empirical networks.

Network	Alg.	Time (s)	Speedup
`com-dblp`	LPA	185.0 ± 105.0	189×
	FLPA	1.0 ± 0.4	
`roadnet-ca`	LPA	940.4 ± 341.8	162×
	FLPA	5.8 ± 1.9	
`us-patents`	LPA	26704.4 ± 12100.4	705×
	FLPA	37.9 ± 12.3	
`foursquare`	LPA	977.3 ± 357.3	64×
	FLPA	15.3 ± 4.7	
`livejournal`	LPA	2248.1 ± 1259.9	30×
	FLPA	74.4 ± 26.3	
`twitter-sample`	LPA	1343.5 ± 544.6	93×
	FLPA	14.5 ± 4.9	
`bitcoin`	LPA	2937.7 ± 1077.3	80×
	FLPA	36.6 ± 13.0	

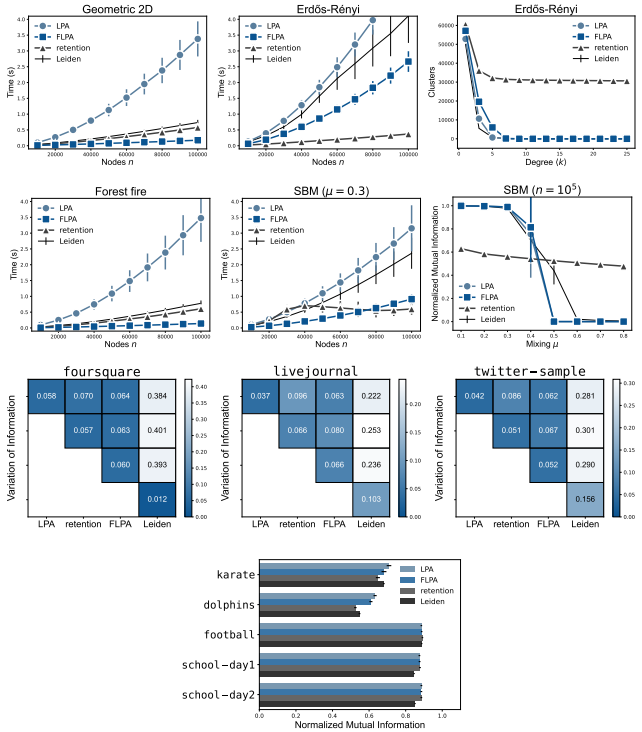

Figure 1: *(top)* **Benchmark graphs.** *(middle)* **Large empirical networks.** *(bottom)* **Small networks with known sociological partitioning.**

FLPA to be useful as a quick initial look at a network, although other slower methods are arguably more robust and preferable [2]. The suggested speedup might also be relevant in the context of other applications of label propagation.

[1] U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E*, 76(3):036106, 2007.

[2] L. Šubelj. Label propagation for clustering. In P. Doreian, V. Batagelj, and A. Ferligoj, editors, *Advances in Network Clustering and Blockmodeling*, Computational and Quantitative Social Science, pages 121–150. Wiley, New York, 1st edition, 2020.

[3] V. A. Traag and L. Šubelj. Large network community detection by fast label propagation. *Sci. Rep.*, pages 1–11, 2023. `https://arxiv.org/abs/2209.13338`.

This work has been supported in part by Slovenian Research Agency ARRS under the program P5-0168.