Bayan: A Branch-and-Cut Algorithm for Globally Maximizing Modularity

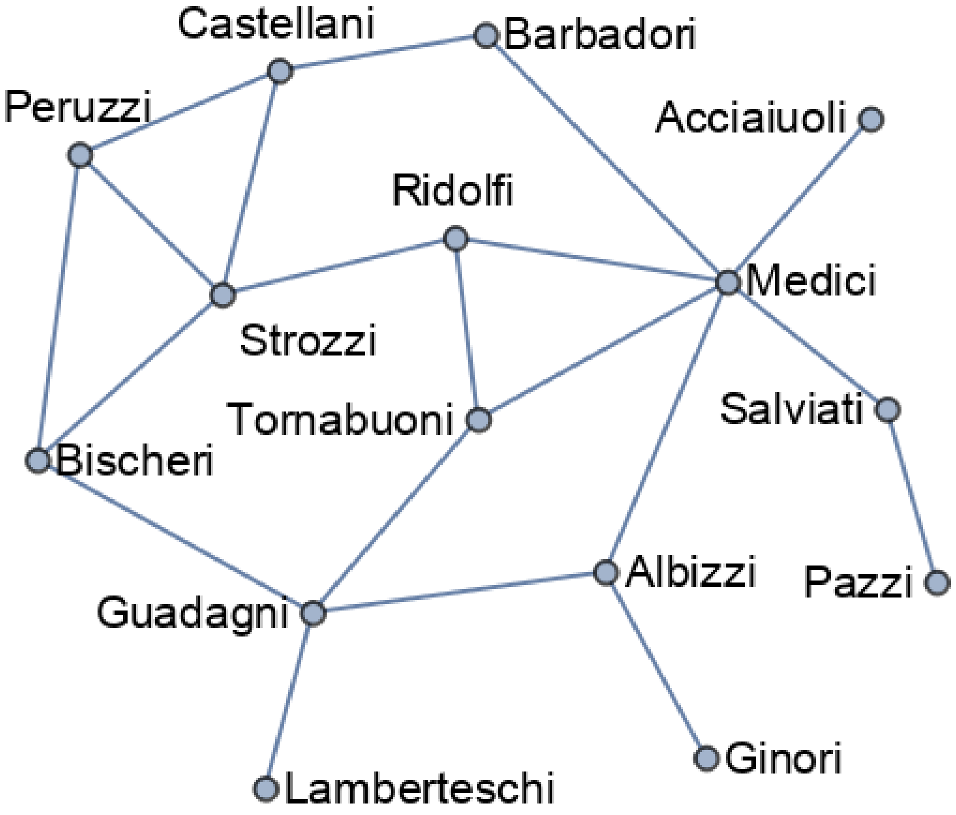
Samin Aref (University of Toronto)

Joint work with
Mahdi Mostajabdaveh (Huawei Technologies Canada)
and Hriday Chheda (University of Toronto)

CompleNet 2023
2023-04-26

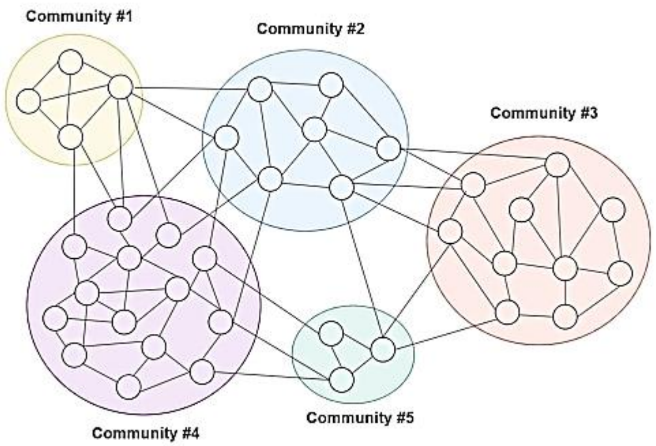
Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

What is the maximum modularity of the Florentine families network? (γ=1)



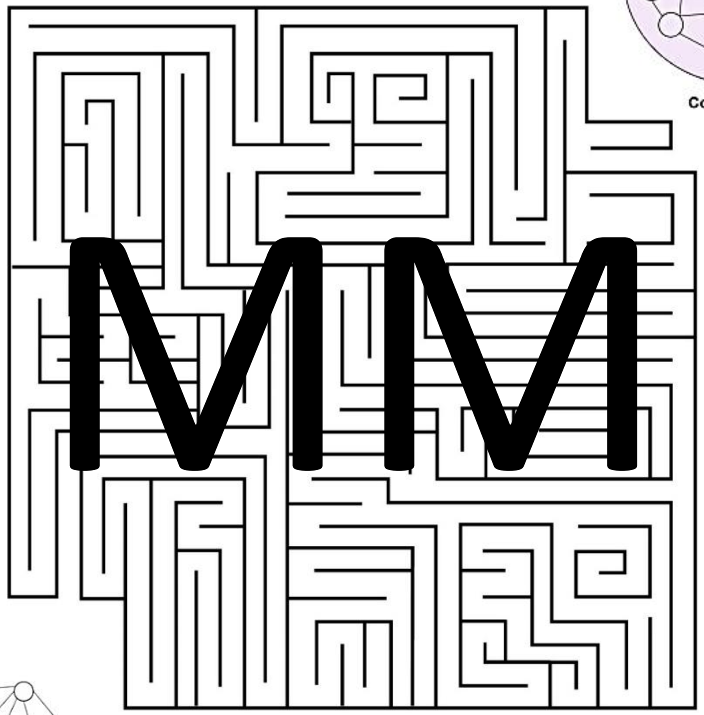
UNIVERSITY OF TORONTO

Output: communities

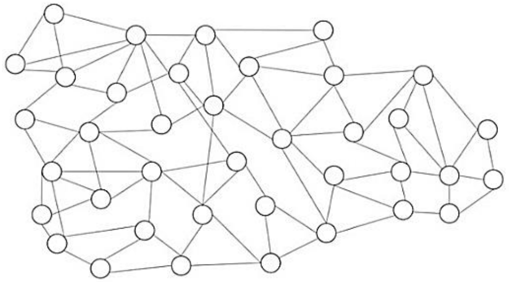


Detecting (assortative) communities via Modularity Maximization (MM)

Modularity b_{ij} : a function of degrees d_i, incidence a_{ij}, and the resolution parameter γ



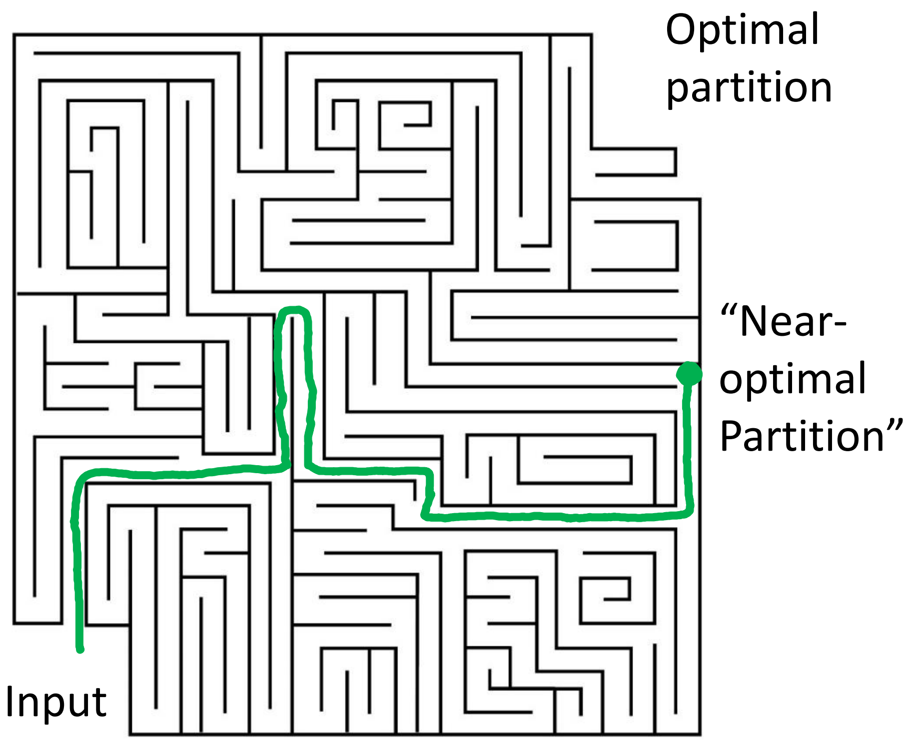
Input: a simple graph



Figures from timbr.ai

Approach 1: Modularity Maximization Heuristics

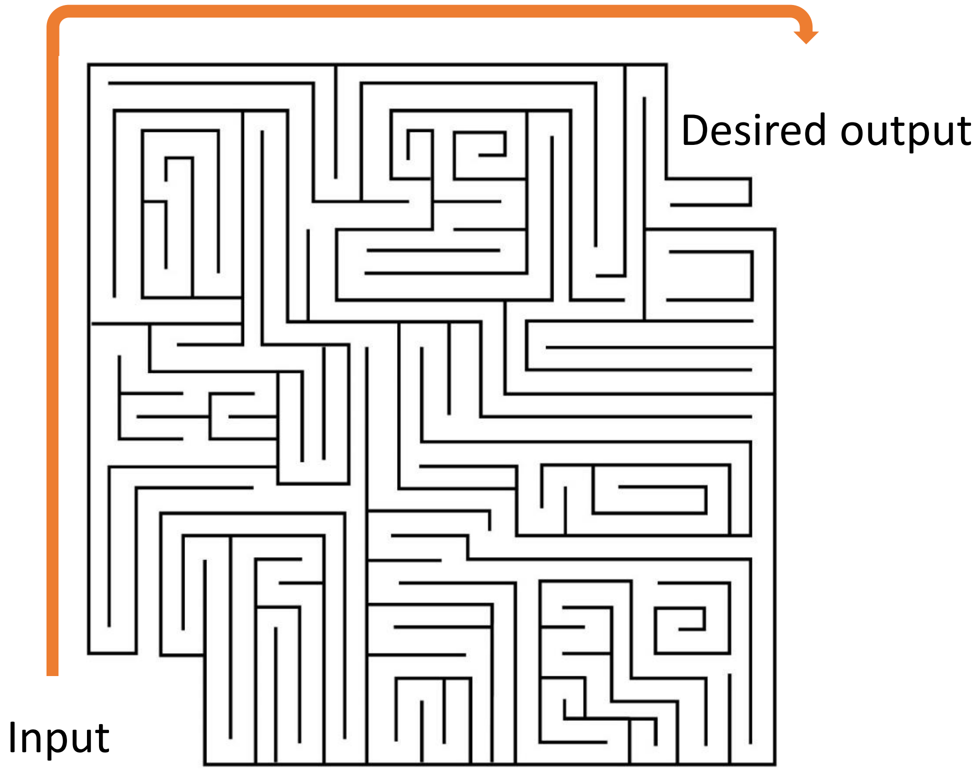
1. Edge Motif (EdMot) (Li et al. 2019)
2. Leiden (Traag et al. 2019)
3. Paris (Bonald et al. 2018)
4. Belief (Zhang & Moore 2014)
5. Combo (Sobolevsky et al. 2014)
6. Leicht-Newman (LN) (Leicht & Newman 2008)
7. Louvain (Blondel et al. 2008)
8. Greedy (CNM) (Clauset et al. 2004)



UNIVERSITY OF TORONTO

Approach 2: Not using modularity

1. Chinese whispers (Biemann et al. 2006)
2. RB (Reichardt & Bornholdt 2006)
3. Walktrap (Pons & Latapy 2006)
4. k-cut (Ruan & Zhang 2007)
5. Infomap (Rosvall & Bergstrom 2008)
6. Genetic Algorithm (Pizzuti 2008)
7. Label propagation (Cordasco & Gargano 2010)
8. CPM (Traag et al. 2011)
9. Significant scales (Traag et al. 2013)
10. Stochastic Block Model (SBM) (Peixoto 2014)
11. WCC (Prat-Pérez et al. 2014)
12. Surprise (Traag et al. 2015)
13. GemSec (Rozemberczki et al. 2019)



UNIVERSITY OF TORONTO

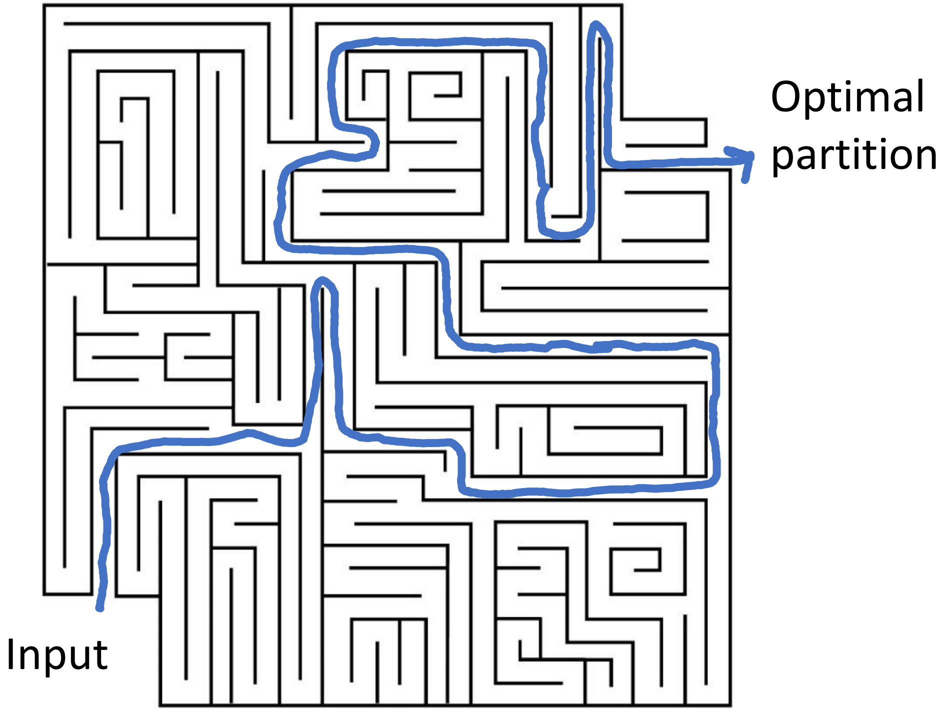
Approach 3: Exact/approximate modularity maximization

- Integer Programming - IP (Brandes et al. 2007)
- IP and LP rounding (Agarwal & Kempe 2008)
- Column generation (Aloise et al. 2010)
- Sparse IP and LP rounding (Dinh & Thai 2015)
- Approximation (Kawase et al. 2021)
- Our proposed method: Bayan (Aref et al. 2022)

$$\max_{x_{ij}} Q = \frac{1}{2m} \sum_{(i,j)\in V^2, i\leq j} b_{ij}(1 - x_{ij})$$

$$\text{s.t.} \quad x_{ik} + x_{jk} \geq x_{ij} \quad \forall k \in K(i,j) \subseteq V \setminus \{i,j\}$$

$$x_{ij} \in \{0,1\} \quad \forall (i,j) \in V^2, i \leq j$$

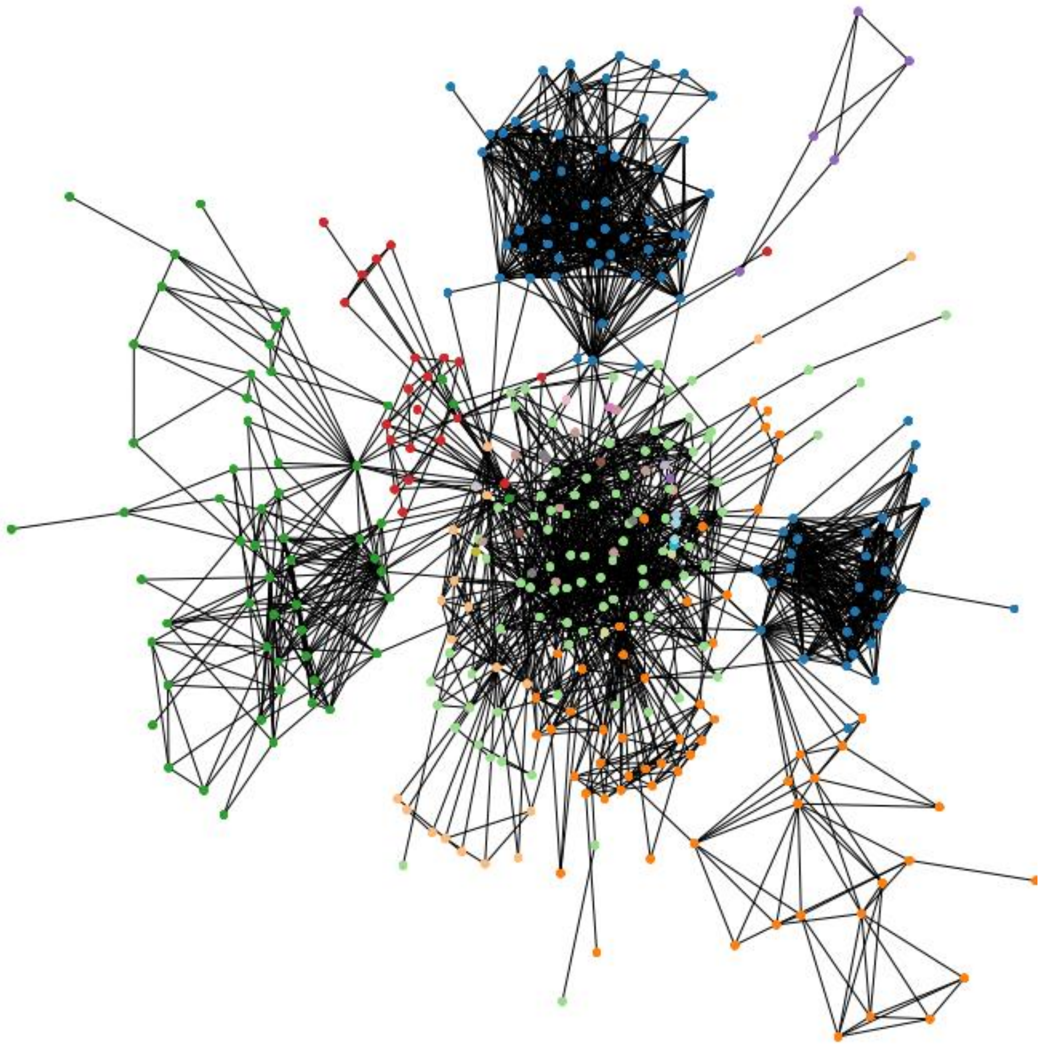


UNIVERSITY OF TORONTO

Tired of eye-balling the communities? Try Bayan.

Dataset:
facebook_friends*
m=1988

Q: modularity
k: number of communities
AMI: adjusted mutual information
(similarity to an optimal partition)



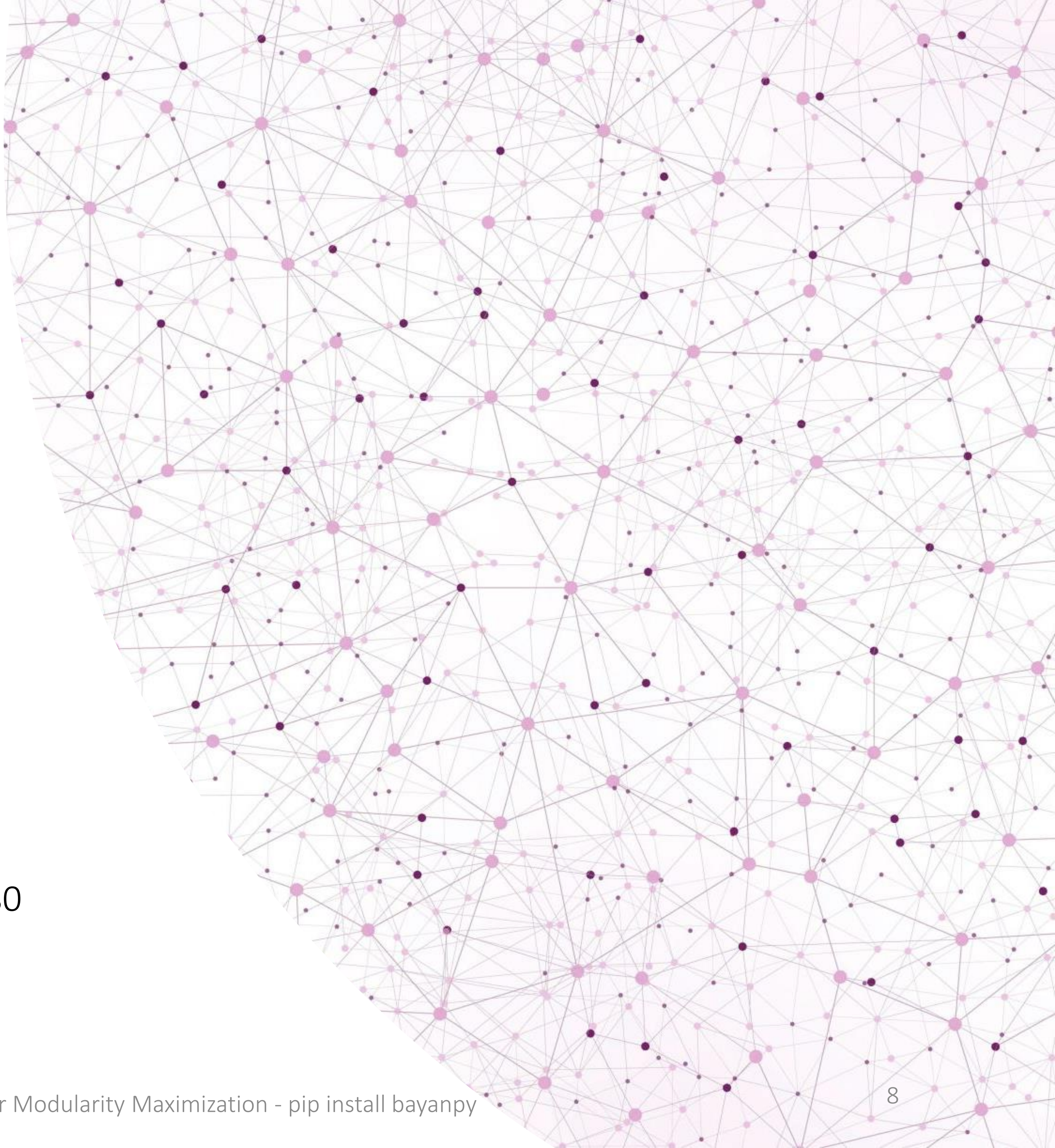
(a) Bayan, $Q^* = 0.7157714$, $k = 28$, AMI $= 1$

* Available in Netzschleuder

1

Heuristic modularity maximization algorithms rarely* maximize modularity.

*Only 16.9% of times according to our experiments on 80 networks

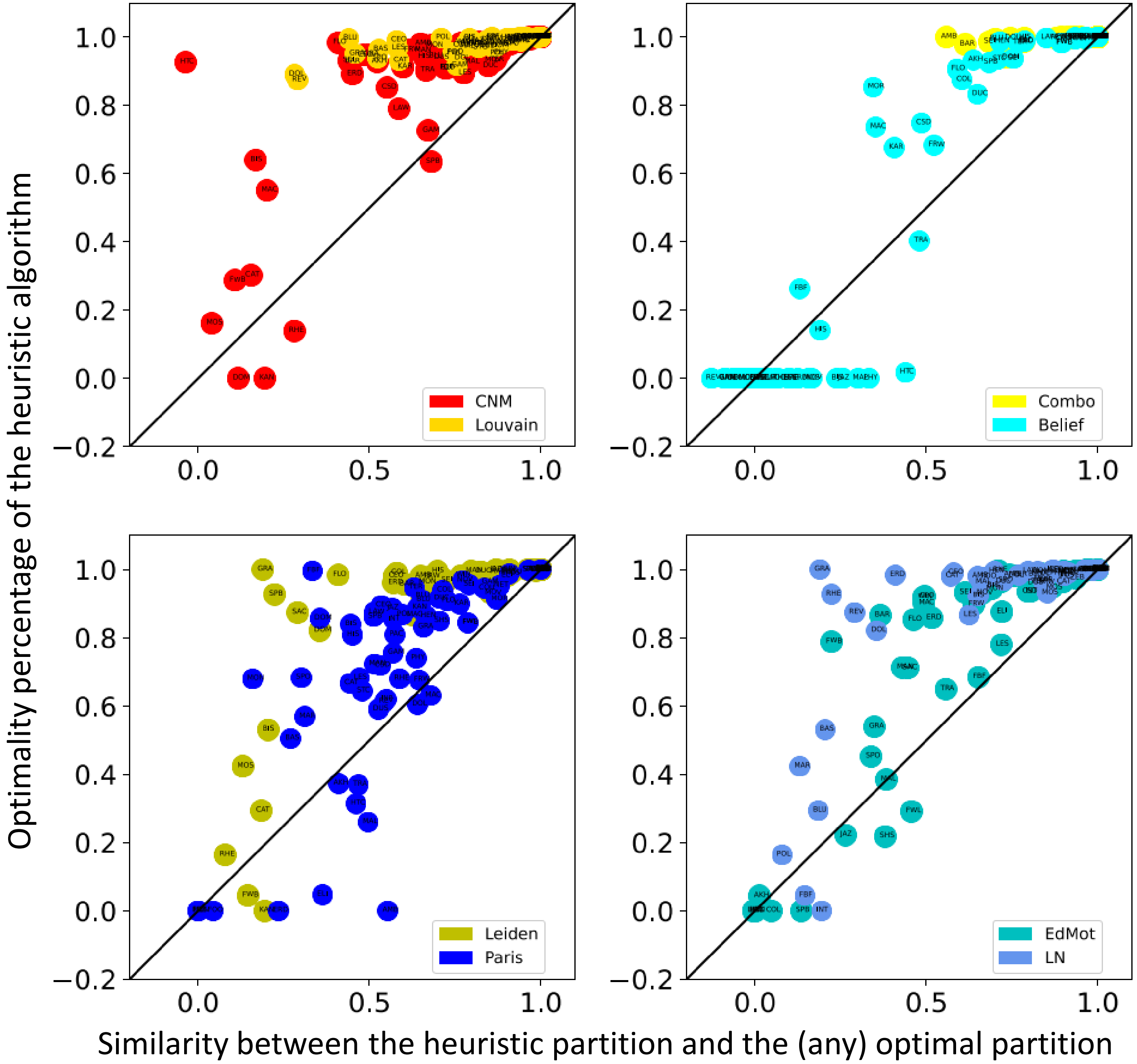


Test cases are 80 graphs with no more than 2812 edges:

- 60 real networks
- 10 Erdős–Rényi random graphs
- 10 Barabási–Albert random graphs

Each datapoint represents the performance of one algorithm on one test case.

1. Y-values: Many partitions are sub-optimal
2. X-values: Many partitions are dissimilar to any optimal partition
3. 45°-line: near-optimal partitions are not similar to any optimal partitions

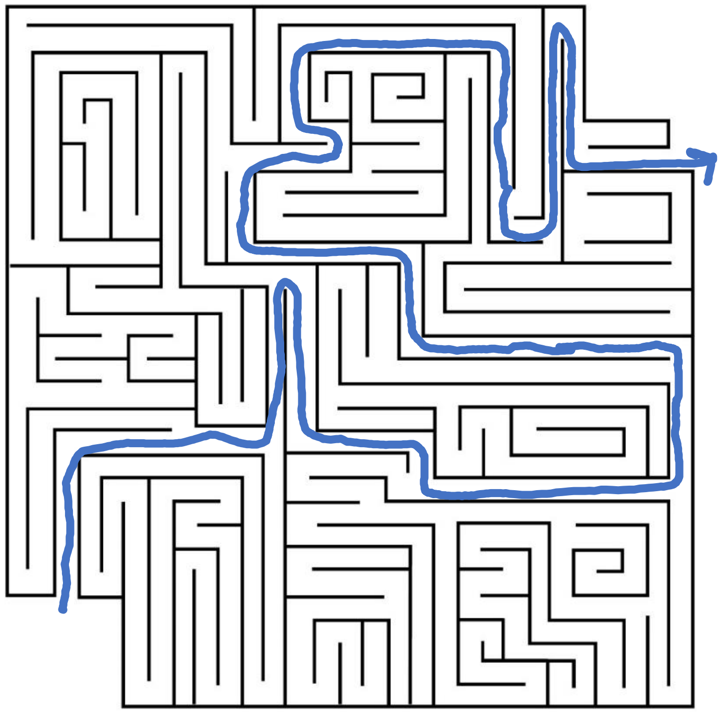
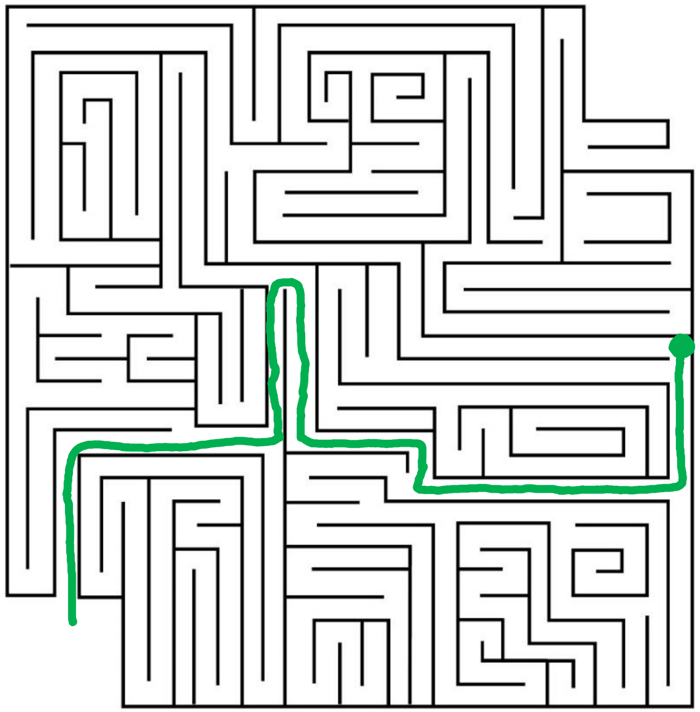
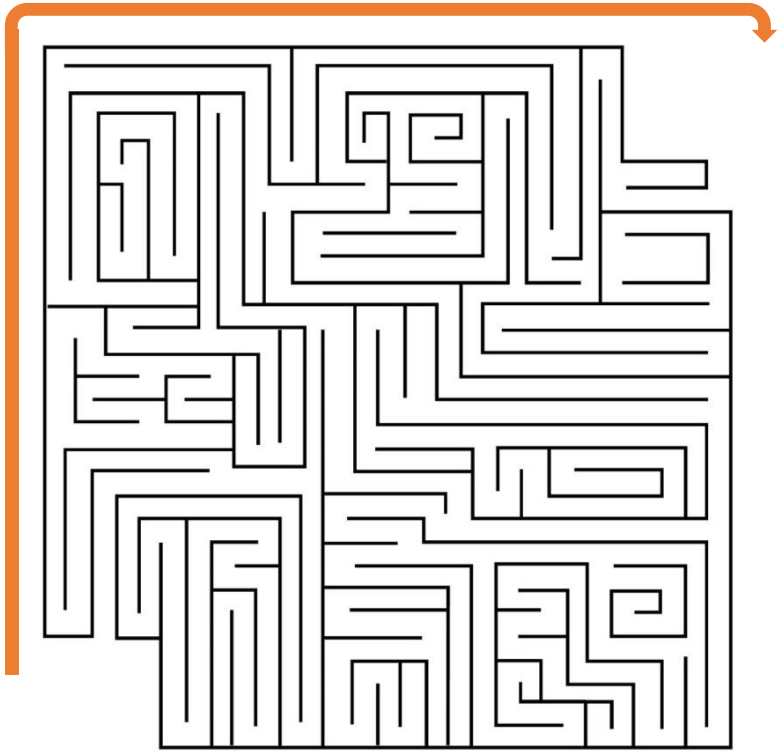


2

Suboptimal partitions of heuristic algorithms are disproportionately dissimilar to any optimal partition.

An x% suboptimality is often associated with a dissimilarity much larger than x% from any optimal partition.

Comparing Bayan with 21 other community detection algorithms

UNIVERSITY OF TORONTO

Comparing Bayan with 21 other community detection algorithms (on random LFR graphs)

Test cases:

- 500 LFR random graphs (generated with planted communities) with up to 300 nodes
- Fraction of inter-community edges μ

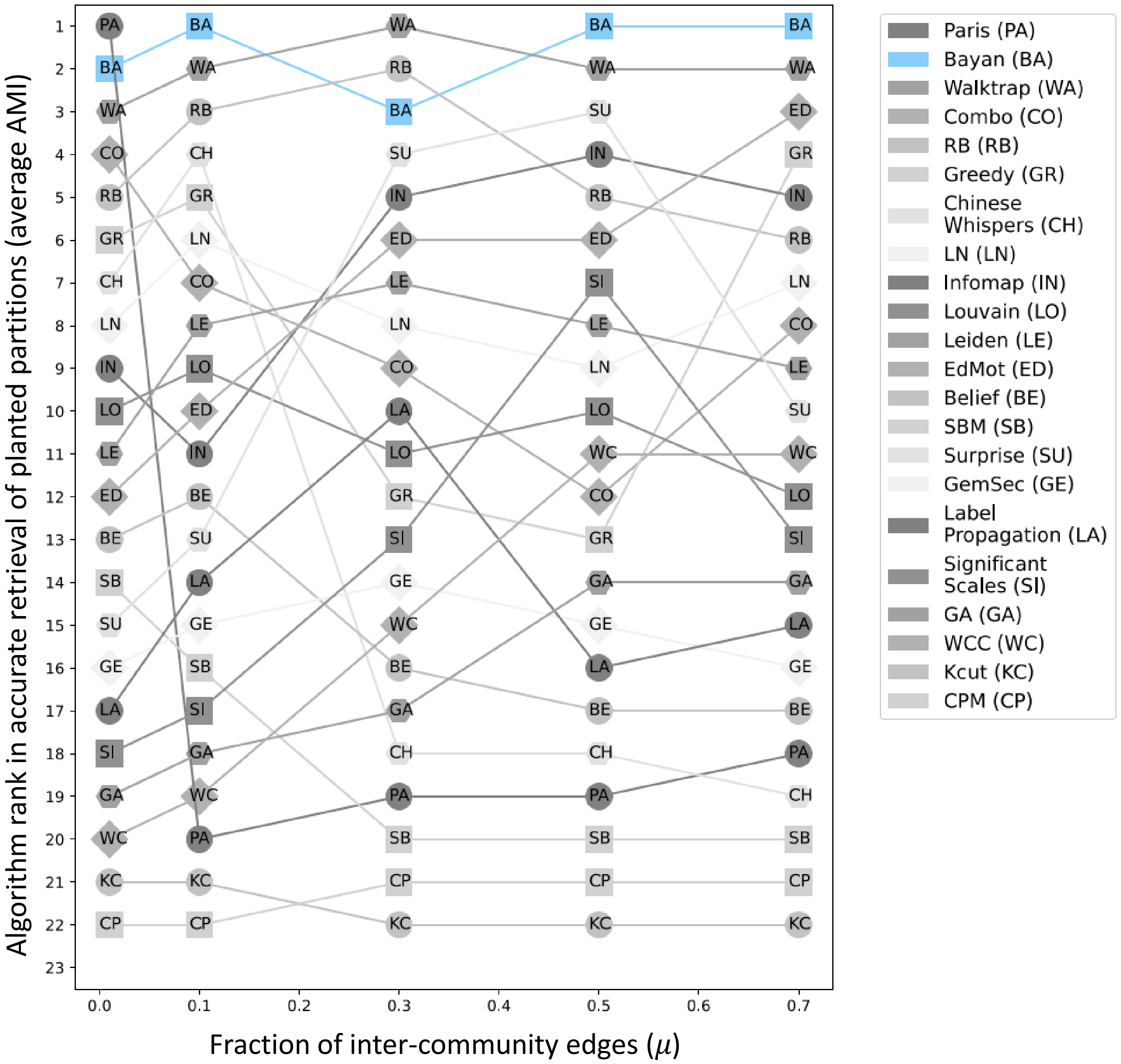
$$\mu \in \{1\%, 10\%, 30\%, 50\%, 70\%\}$$

Performance measure:

- Similarity with the ground-truth communities (adjusted mutual information averaged over 100 graphs)

UNIVERSITY OF TORONTO

Comparing Bayan with 21 other community detection algorithms (on random LFR graphs)



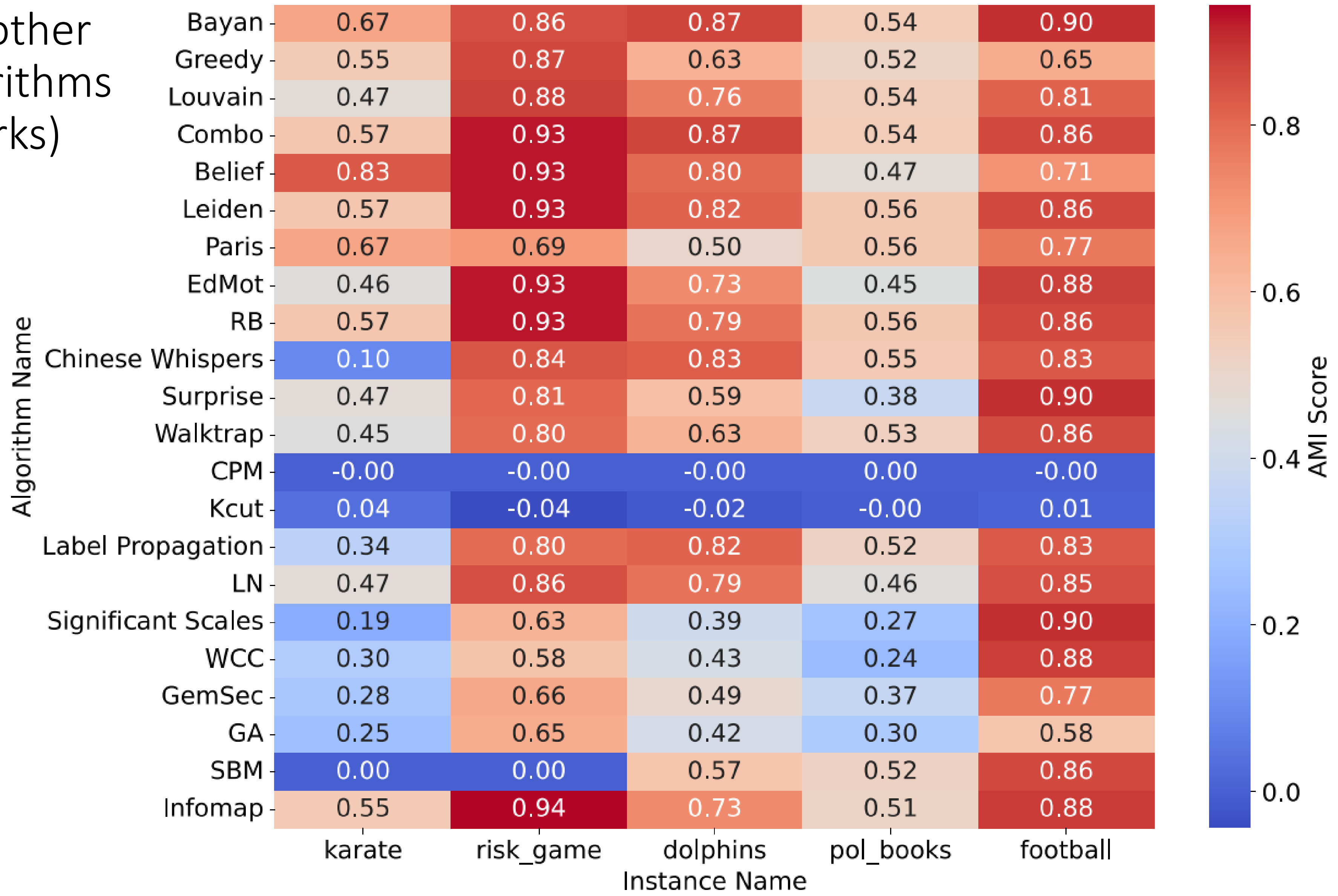
UNIVERSITY OF TORONTO

3

Bayan is more accurate than 21 other algorithms in retrieving planted communities of LFR graphs.

Bayan is several times faster than open-source and commercial solvers for maximizing modularity.

Comparing Bayan with 21 other community detection algorithms (on real network benchmarks)



UNIVERSITY OF TORONTO

4

Bayan maximizes modularity in mid-sized* networks and approximates maximum modularity in larger networks on ordinary computers.

*up to 3000 edges in their largest connected component

You can start using Bayan in Python today!

```
%pip install bayanpy
```

```
import networkx as nx
import bayanpy
```

```
G = nx.barbell_graph(5,2)
```

```
bayanpy.bayan(G)
```

Paper 1: arxiv.org/abs/2302.14698 (ICCS'23)
Paper 2: arxiv.org/pdf/2209.04562
GitHub Repo: github.com/saref/bayan
Project website: bayanproject.github.io

Try Bayan on your smartphone ->

Google Colab examples: tinyurl.com/bayancolab

UNIVERSITY OF TORONTO

Special thanks to my co-authors:

Mahdi Mostajabdaveh
Huawei Technologies Canada Co., Ltd.

Hriday Chheda
University of Toronto

UNIVERSITY OF TORONTO

References

Dinh, T.N., Thai, M.T.: Toward optimal community detection: From trees to general weighted networks. Internet Mathematics 11(3), 181–200 (2015)

Clauset, A., Newman, M.E., Moore, C.: Finding community structure in very large networks. Physical review E 70(6), 066111 (2004)

Biemann, C.: Chinese whispers: an efficient graph clustering algorithm and its application to natural language processing problems. In: Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing. TextGraphs-1, pp. 73–80. (2006)

Reichardt, J., Bornholdt, S.: Statistical mechanics of community detection. Physical Review E 74(1), 016110 (2006).

Pons, P., Latapy, M.: Computing communities in large networks using random walks. J. Graph Algorithms Appl. 10(2), 191–218 (2006)

Ruan, J., Zhang, W.: An efficient spectral algorithm for network community discovery and its applications to biological and social networks. In: Seventh IEEE International Conference on Data Mining (ICDM 2007), pp. 643–648. IEEE, (2007)

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment 2008(10), 10008 (2008).

Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences 105(4), 1118–1123 (2008).

Leicht, E.A., Newman, M.E.J.: Community structure in directed networks. Physical Review Letters 100(11), 118703 (2008).

Pizzuti, C.: GA-Net: A Genetic Algorithm for Community Detection in Social Networks. In: Proceedings of the 10th International Conference on Parallel Problem Solving from Nature — PPSN X - Volume 5199, pp. 1081–1090. Springer, Berlin, Heidelberg (2008)

Cordasco, G., Gargano, L.: Community detection via semi-synchronous label propagation algorithms. In: 2010 IEEE International Workshop On: Business Applications of Social Network Analysis (BASNA), pp. 1–8. IEEE, (2010)

Traag, V.A., Van Dooren, P., Nesterov, Y.: Narrow scope for resolution-limit-free community detection. Physical Review E 84(1), 016114 (2011)

Traag, V.A., Krings, G., Van Dooren, P.: Significant scales in community structure. Scientific reports 3(1), 1–10 (2013)

Peixoto, T.P.: Efficient monte carlo and greedy heuristic for the inference of stochastic block models. Physical Review E 89(1), 012804 (2014)

Prat-P´erez, A., Dominguez-Sal, D., Larriba-Pey, J.-L.: High quality, scalable and parallel community detection for large real graphs. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 225–236 (2014)

Sobolevsky, S., Campari, R., Belyi, A., Ratti, C.: General optimization technique for high-quality community detection in complex networks. Physical Review E 90(1), 012811 (2014)

Zhang, P., Moore, C.: Scalable detection of statistically significant communities and hierarchies, using message passing for modularity. Proceedings of the National Academy of Sciences 111(51), 18144–18149 (2014)

Traag, V.A., Aldecoa, R., Delvenne, J.-C.: Detecting communities using asymptotical surprise. Physical Review E 92(2), 022816 (2015)

Bonald, T., Charpentier, B., Galland, A., Hollocou, A.: Hierarchical graph clustering using node pair sampling. In: MLG 2018 - 14th International Workshop on Mining and Learning with Graphs, London, UK (2018)

Traag, V.A., Waltman, L., van Eck, N.J.: From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports 9(1) (2019).

Li, P.-Z., Huang, L., Wang, C.-D., Lai, J.-H.: EdMot: An edge enhancement approach for motif-aware community detection. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 479–487 (2019)

Rozemberczki, B., Davies, R., Sarkar, R., Sutton, C.: Gemsec: Graph embedding with self clustering. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 65–72 (2019)

1. Heuristic modularity maximization algorithms rarely maximize modularity.

2. Suboptimal partitions of heuristic algorithms are disproportionately dissimilar to any optimal partition.

3. Bayan is more accurate and more stable than 21 other algorithms in retrieving planted communities of LFR graphs.

4. Bayan maximizes modularity in mid-sized networks and approximates maximum modularity in larger networks on ordinary computers.

Try Bayan on your smartphone ->

Thank you!

Questions?

aref@mie.utoronto.ca
saref.github.io
@SaminAref

UNIVERSITY OF TORONTO