9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

A Branch-and-Cut Method for Globally Maximizing Modularity and Discovering Network Communities

Keywords: Community detection, network science, modularity maximization, integer programming, graph optimization

Extended Abstract

Community detection is a classic problem in network science with extensive applications in various fields (Fortunato and Newman). The most commonly used methods are the algorithms designed to maximize modularity across different ways that a network can be partitioned into communities. Most heuristic algorithms for modularity maximization tend to scale well for large networks, but this come at a cost: their partitions have no guarantee of proximity to an optimal solution (Good et al.) and these sub-optimal solutions tend to have partitions disproportionally dissimilar to an optimal partition. Given these results, we propose a community detection algorithm based on exact or approximate maximization of modularity acknowledging that there will be a limit to the size of the largest network that such an algorithm can handle.

We propose the Bayan algorithm which, unlike the existing methods, returns network partitions with a guarantee of either optimality or proximity to an optimal solution. At the core of the Bayan algorithm is a branch-and-cut scheme that solves a sparse integer programming formulation (Dinh and Thai) of the modularity maximization problem to optimality or approximate it within a factor. It takes advantage of structural properties of input networks and leverages the state-of-the-art optimization approaches (e.g., valid inequalities, variable fixing, pre-solve graph reduction, and lower bounding heuristics) to push the limits for solving a fundamental NP-complete problem exactly and efficiently. Through extensive experiments, we demonstrate Bayan's distinctive capabilities not only in maximizing modularity, but more importantly in accurate retrieval of ground-truth communities.

We compare Bayan to 22 community detection algorithms including the modularity-based heuristics, optimization-based heuristics, as well as other community detection methods which do not rely on modularity or optimization. We use Lancichinetti-Fortunato-Radicchi (LFR) benchmark graphs as test instances with ground-truth partitions. We measure the the performance of each method based on the normalized Adjusted Mutual Information (AMI) (Vinh et al.) of the its partition with the planted ground-truth partition from the LFR graph generation process. The 22 algorithms we compare Bayan with are known as Clauset-Newman-Moore (CNM), Chinese whispers, Reichardt-Bornholdt (RB), walktrap, k-cut, Louvain, Infomap, Leicht-Newman (LN), Genetic Algorithm (GA), label propagation, CPM (Traag et al.), significant scales (Traag et al.), Stochastic Block Model (SBM), Weighted Community Clustering (WCC), combo, belief, surprise (Traag et al.), head-tail (Jiang and Ma), Paris, Leiden, EdMot (Li et al.), and Gemsec (Rozemberczki et al.). For this evaluation, we randomly generate 100 LFR benchmark graphs for each value of parameter μ chosen from the set $\{0.01, 0.1, 0.3, 0.5, 0.7\}$. The parameter μ is a control parameter for LFR which determines the fraction of inter-community edges incident to each node.

Figure 1 illustrates the relative ranking of the algorithms, including Bayan, according to AMI averaged over 100 LFR graphs generated based on each value of μ. Bayan is ranked first in two out of five experiment settings, and it comes second in the other experiments. We can observe from this analysis that Bayan has the most stable performance in detecting the planted partition compared to the other 22 community detection algorithms considered. In Figure 1, the three algorithms walktrap, Paris, and Chinese whispers seem to comparatively perform well for small values of μ, but their comparative performance weakens substantially when the association between structure and communities is weakened by μ set to values larger than 0.1. Contrary to this pattern, there are algorithms like Edmot and GA whose comparative performance improves when μ increases. Averaging the AMI values for all 500 LFR graphs, Bayan has the highest performance followed by walktrap and Leiden respectively.

Not only Bayan's comparative level of performance remains stable for LFR graphs over a wide range of μ, it also returns the highest AMI values for the common benchmarks of real networks with ground-truth clusters[1]. The performance of Bayan as an exact global modularity maximization algorithm also reveals the theoretical capability limits of maximum-modularity partitions in accurate retrieval of communities. Overall our analysis points to Bayan as a suitable choice for a methodologically grounded detection of communities through exact (approximate) maximization of modularity in networks with up to 10^4 edges (and larger networks). Prospective advances in graph optimization and integer programming can push these limits further.

References

Dinh, Thang N. and My T. Thai. "Toward optimal community detection: From trees to general weighted networks". *Internet Mathematics*, vol. 11, no. 3, 2015, Publisher: Taylor & Francis, pp. 181–200.

Fortunato, Santo and Mark EJ Newman. "20 years of network community detection". *Nature Physics*, vol. 18, 2022, pp. 848–850.

Good, Benjamin H., et al. "Performance of modularity maximization in practical contexts". *Physical Review E*, vol. 81, no. 4, 2010, Publisher: APS, p. 046106.

Jiang, Bin and Ding Ma. "Defining least community as a homogeneous group in complex networks". *Physica A: Statistical Mechanics and its Applications*, vol. 428, 2015, Publisher: Elsevier, pp. 154–160.

Li, Pei-Zhen, et al. "Edmot: An edge enhancement approach for motif-aware community detection". *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 479–487.

Rozemberczki, Benedek, et al. "Gemsec: Graph embedding with self clustering". *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*, 2019, pp. 65–72.

Traag, Vincent A., et al. "Detecting communities using asymptotical surprise". *Physical Review E*, vol. 92, no. 2, 2015, Publisher: APS, p. 022816.

Traag, Vincent A., et al. "Narrow scope for resolution-limit-free community detection". *Physical Review E*, vol. 84, no. 1, 2011, Publisher: APS, p. 016114.

Traag, Vincent A., et al. "Significant scales in community structure". *Scientific reports*, vol. 3, no. 1, 2013, Publisher: Nature Publishing Group, pp. 1–10.

Vinh, Nguyen Xuan, et al. "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance". *Journal of Machine Learning Research*, vol. 11, no. 95, 2010, pp. 2837–2854. jmlr.org/papers/v11/vinh10a.html.

[1]These benchmark networks are accessible from the network repository Netzschleuder with the following names: dolphins, football, karate, polbook.

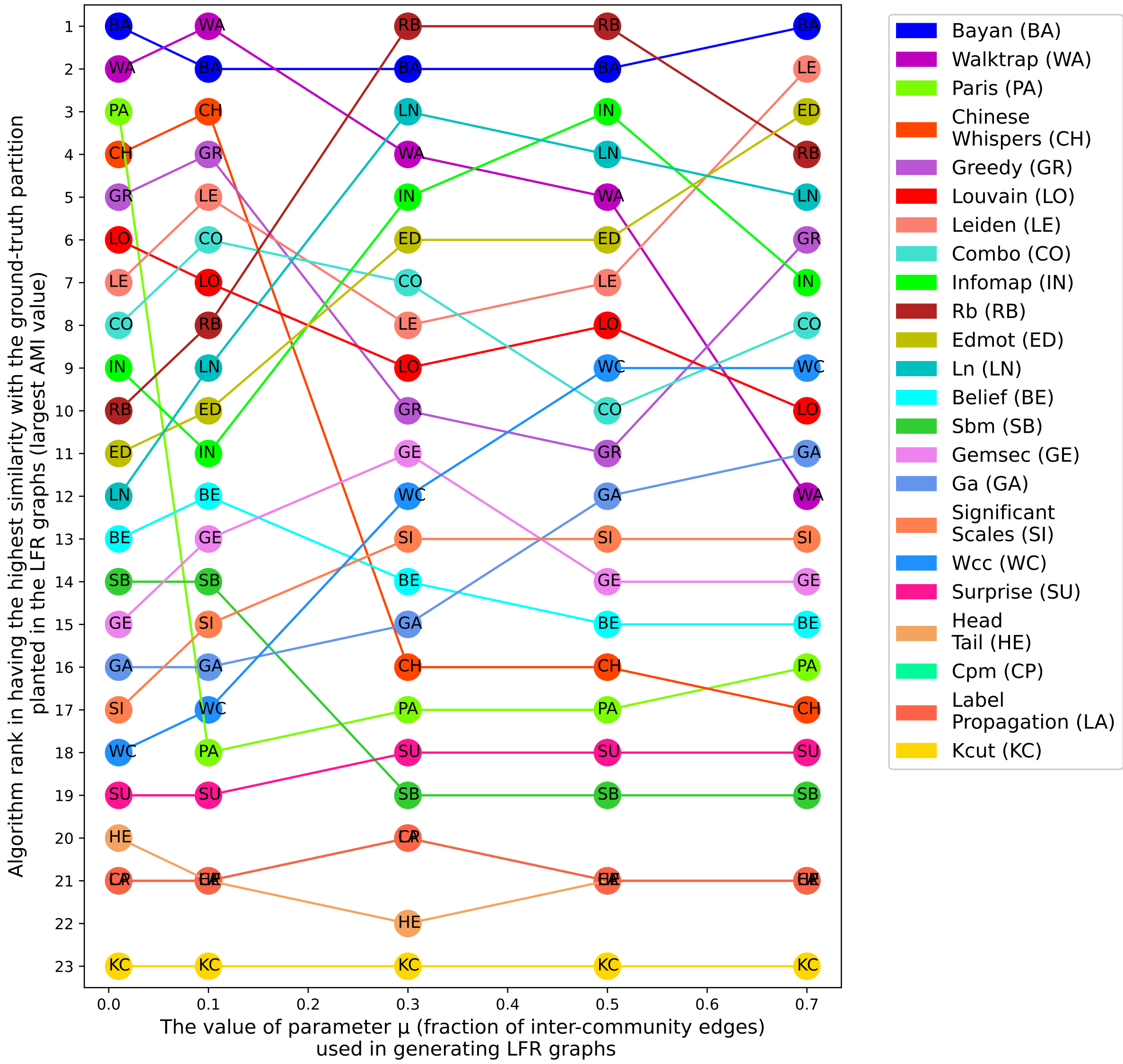

Figure 1: Performance ranking of 23 algorithms on retrieving ground-truth planted communities measured by normalized Adjusted Mutual Information (AMI) averaged over 100 LFR graphs for each value of μ (Magnify this high-resolution figure on screen for the details.)