Tailoring Benchmark Graphs to Real-World Networks for Improved Prediction of Community Detection Performance

Catherine Schwartz[1,2], Cetin Savkli[1], Amanda Galante[1], and Wojciech Czaja[2]
[1] *Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723, USA and*
[2] *University of Maryland, Department of Mathematics, College Park, Maryland 20742, USA*

We introduce a new methodology for improving the understanding of how different community detection methods are expected to perform on a specific real-world network of interest. A common approach used to compare the performance of community detection methods is to measure their ability to detect ground truth communities in benchmark graphs. Studies that employ this approach are typically based on general benchmark model parameters that are selected to create benchmark graphs with realistic community structure. The authors of the studies then provide guidance on how to choose an appropriate method based on the performance results on the generated benchmark graphs. Unfortunately, researchers and practitioners may follow the guidance even though the real-world network they are working with may not have any resemblance to the benchmark graphs used in the study. For example, their real-world network may be highly connected, whereas the general benchmark model parameters may create graphs that are not. In this study, we demonstrate that, by running experiments on tailored benchmark graphs where model parameters are chosen to match a specific real-world network of interest as closely as possible, researchers can obtain a better understanding on how well community detection methods will work on that particular network. Since the popular LFR benchmark [1] (See FIG. 1a) cannot create certain community structures seen in real-world networks, we motivate researchers to additionally consider a recently proposed network growth model called the nPSO benchmark [2] (See FIG. 1b) when determining which benchmark model to use for their tailored benchmark graphs. For this study, our first real-world network of interest was the email-Eu-core network, a publicly available dataset from SNAP [3] which we reveal has a community structure similar to the community structure generated by the nPSO benchmark. We show that the performance of the community detection methods on the email-Eu-core network is highly correlated with the performance of the same methods on the corresponding tailored benchmark graphs (r = 0.93, FIG. 2a). This suggests that methods that performed well on the tailored benchmark graphs also performed well on the real-world network. Conversely, the performance of community detection methods on the email-Eu-core network was not correlated with the performance of the same methods on unrelated benchmark graphs (r = -0.25, FIG. 2b), meaning the methods that performed well on the unrelated benchmark graphs did not perform well on the email-Eu-core network. We further demonstrate how to create tailored benchmark graphs when a real-world network has no associated ground truth and introduce a tool that can be used to help ensure the appropriate community structure is reflected in the tailored benchmark graphs. We use another publicly available real-world network from SNAP [3], the DBLP collaboration network, that does not have non-overlapping ground truth communties and we use the mentioned tool to illustrate that the community structure of the DBLP network is similar to the community structure generated by the LFR benchmark. There are a number of similarities between how the community detection methods perform on the tailored benchmark graphs for the DBLP network and how the methods perform on the actual network itself. The results inform the type of errors the community detection methods are expected to make and identify which methods will tend to overpredict or underpredict the number of communities [4]. By utilizing tailored benchmark graphs, researchers and practitioners can select an appropriate community detection method in a more systematic way for a specific network they are studying and reach a better understanding of communities that are generated. This approach will increase trust and confidence in the resulting communities, which is particularly important if the communities are going to be used for downstream analyses.

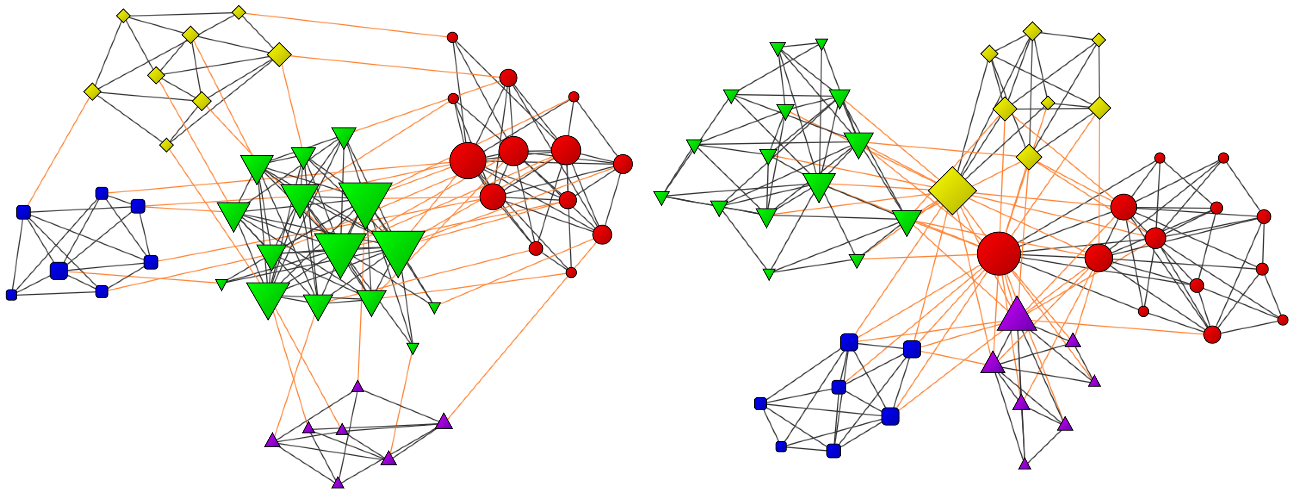
Keywords – community structure, community detection, benchmark graphs, network models

[1] A. Lancichinetti, S. Fortunato, and F. Radicchi, Benchmark graphs for testing community detection algorithms, Physical Review E **78** (2008).

[2] A. Muscoloni and C. V. Cannistraci, A nonuniform popularity-similarity optimization (npso) model to efficiently generate realistic complex networks with communities, New Journal of Physics **20** (2018).

[3] J. Leskovec and A. Krevl, SNAP Datasets: Stanford large network dataset collection, `http://snap.stanford.edu/data` (2014).

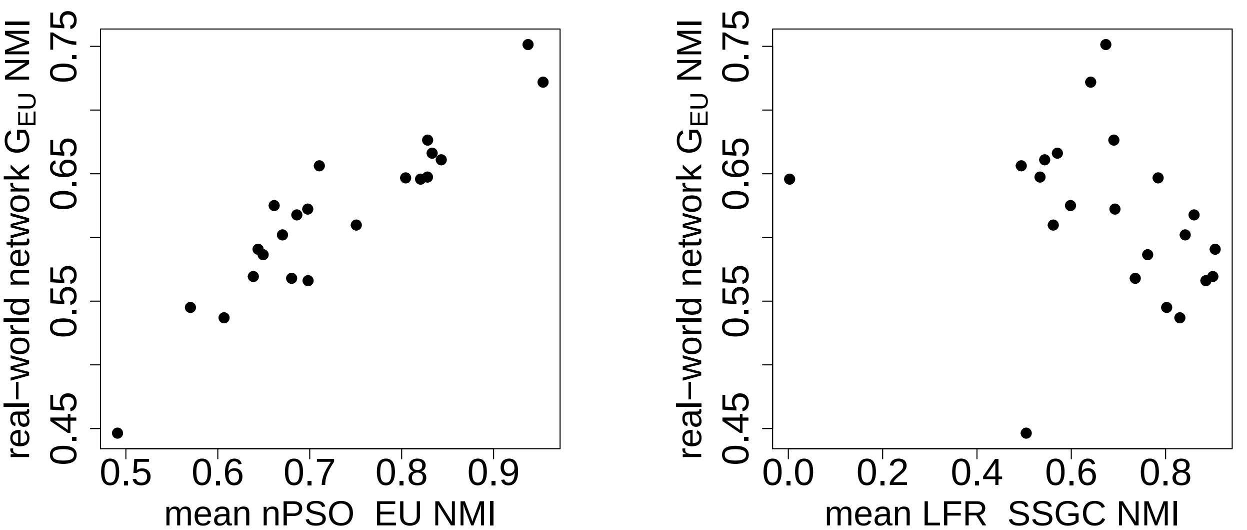
[4] C. Schwartz, *Analyzing Semi-Local Link Cohesion to Detect Communities and Anomalies in Complex Networks*, Ph.D. thesis (2021).



(a) Degree assortative community structure created with the LFR benchmark [1]

(b) Hierarchical community structure created with the nPSO benchmark [2]

FIG. 1: Example graphs with five communities. Color and shape indicates a vertex's community. Vertex size corresponds to its degree. Internal edges are black and external edges are orange.



(a) nPSO_EU are tailored nPSO benchmark graphs [2] for the G_{EU}, the email-Eu-core network

(b) LFR_SSGC are unrelated LFR benchmark graphs [1] for the G_{EU}, the email-Eu-core network

FIG. 2: Comparison of the mean normalized mutual information (NMI) from benchmark graphs to the NMI from G_{EU}, the email-Eu-core network [3], of different the community detection methods. Each point represents a different community detection method.