Defining and evaluating network communities based on ground-truth

Yang and Lescovec [4] describe their metholodology in processing the ground truths in the SNAP database [3].

Recall that for community detection algorithms, we often use synthetic networks with ground truth partitions. For example, in the LFR software [2], the ground truth nodes have different connection probabilities (governed by the mixing parameter).

In empirical networks, we can use metadata to find a gold-standard labeling. For social networks (e.g. Youtube), users are nodes, friendships are edges, and user-created groups are communities. Despite this, we do not know if there are detectable topological differences between the gold-standard labeling (often by metadata).

The authors study scoring metrics for how "community-like" their clustering is. For individual clusters, they calculate the density, separability, and connectivity. They found that for their networks,

1. There is a big drop in quality after the top 5000 community (average rank across their metrics)

2. Some networks have communities that are highly separable, while others are highly dense (sometimes not both)

3. Modularity is inversely proportional to density, separability, and connectivity functions

4. Triangle Participation Rate (cohesiveness) and Conductance (separability) are the two metrics to use

   (a) They measure different qualities of communities

   (b) They are the least sensitive

Overall, this paper definitely feels a bit ad-hoc and the details feel like they're lacking. I'm not sure why they make a distinction between cluster scoring and cluster goodness. This top 5000 community threshold seems arbitrary and not something generalizable to different network sizes. Why do they include the "bad" scoring functions (e.g. modularity) into their top 5000 ranking, instead of just using TPR and Conductance. Is there a better way to weight, than just taking the average?

I read this paper to determine if and how can use these networks to measure accuracy for community detection. Reading this, I'm still not sure.

On another note, Justin Luke took a class with Lescovec! The software is publicly accessible, packaged and documented [1].

References

[1] Stanford Network Analysis Platform (SNAP). *snap*. https://github.com/snap-stanford/snap/. June 2025.

[2] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. "Benchmark graphs for testing community detection algorithms". In: (2008). DOI: 10.48550/ARXIV.0805.4770. URL: https://arxiv.org/abs/0805.4770.

[3] Jure Leskovec and Andrej Krevl. *SNAP Datasets: Stanford Large Network Dataset Collection*. http://snap.stanford.edu/data. June 2014.

[4]   Jaewon Yang and Jure Leskovec. "Defining and evaluating network communities based on ground-truth". In: *Knowledge and Information Systems* 42.1 (Oct. 2013), pp. 181–213. ISSN: 0219-3116. DOI: 10.1007/s10115-013-0693-z. URL: http://dx.doi.org/10.1007/s10115-013-0693-z.