Improved Community Detection Using Stochastic Block Models

Park et al. [1] builds upon the previous work in the CM pipeline [2]. Indeed, they examine three treatments here:

- CC (Connected Components), each of the connected components represent an individual cluster

- WCC (Well-Connected Clusters), removing minimum cuts until well-connected

- CM (Connectivity-Modifier), WCC but with an additional re-clustering after removing a mincut

They first explore the properties of SBM clusters, and then evaluate their treatments on both real and synthetic networks. Their findings are different for small vs medium/large networks.

On small (≤ 1000 nodes) networks, SBM typically returns clusterings that are connected, and sometimes well-connected. Moreover, there are only small drops in node coverage due to the treatments (with CC affecting the least).

On the medium and large networks, SBM typically returns clusters that are disconnected. Here, the treatments do greatly drop the node coverage (indeed the magnitude is lowest for CC, then WCC, then CM). Note that the average returned cluster size after CM is larger than that for CC and WCC (cause it does re-clustering). Indeed, CC drops the average cluster size by an order of magnitude (or more).

To determine the effect of the treatments on accuracy, they further evaluate on synthetic (LFR) networks. Towards ARI, CC typically has neutral impact, WCC has neutral or beneficial, and CM's impact is unpredictable.

For theoretical explanation, they look at the description length formula. They sum the negative log-likelihood of the model (adjacency matrix given parameters) and of the priors. Specifically, for the edge-count matrix,

$$-\log p(e) = \log \binom{B(B+1)/2 + E - 1}{E}$$

where B and E are the number of blocks and edges. Hence, clusterings before the treatment are favored by the description length formula (despite the post-treatment having higher accuracy).

References

[1] Minhyuk Park et al. "Improved Community Detection Using Stochastic Block Models". In: *Complex Networks & Their Applications XIII*. Ed. by Hocine Cherifi et al. Cham: Springer Nature Switzerland, 2025, pp. 103–114. ISBN: 978-3-031-82435-7.

[2] Minhyuk Park et al. "Well-connectedness and community detection". In: *PLOS Complex Systems* 1.3 (Nov. 2024), pp. 1–25. DOI: 10.1371/journal.pcsy.0000009. URL: https://doi.org/10.1371/journal.pcsy.0000009.