

Well-Connectedness and Community Detection

Park et al. [1] is addressing the problem that standard community detection algorithms are returning *poorly-connected* clusters. They perform experiments on real-world and simulated networks for a suite of different methods. Here, I will focus on the Leiden algorithm (optimizing Constant-Potts (CPM), or Modularity (Mod)).

To improve the connectivity of output clusters, they introduce a new pipeline, Connectivity Modifier (CM). CM applies post-processing (according to user inputted parameters) that can work with a variety of base methods. While there are poorly-connected clusters, CM will uncluster them and recluster individually.

First, they explore some properties of the Leiden algorithm [0]. They found that Leiden+Mod behaved like Leiden+CPM with very small resolution ($<1e-5$), in terms of connectivity and sizes of the clusters. In general, refining the resolution (making it smaller) leads to more coverage (and larger size) of clusters, at the expense of their connectivity.

To formally define well-connectedness, they chose a threshold function $f(n) = \log_2(n)$, where n is the cluster size, to decide whether a cluster was well-connected or not. With this, indeed Leiden only produces mostly well-connected clusters with resolution $r \geq 0.1$, on the real-world datasets.

On real-world networks, applying CM decreases the node coverage to varying degrees based on the dataset and method. For coarse resolutions, CM does not significantly affect the results (indeed, there is a theoretical guarantee on connectedness). For fine resolutions, CM significantly drops the coverage, as the frequency of poorly-connected clusters increases. CM primarily *reduces* the cluster size, but occasionally it *degrades* the cluster into singletons, or *splits* into multiple smaller clusters.

The stages within the CM pipeline affect the output accuracy in different ways. First, the initial pre-filter of clusters with ≤ 1 edge connectivity decreases the accuracy. However, the resulting ability to improve poorly-connected clusters is neutral of beneficial to accuracy. Hence, for fine resolutions, we see improvements in accuracy (because higher proportion of clusters are poorly-connected), and minor drops in accuracy for coarse resolutions.

Overall, their results provide further evidence of *over-clustering* and the accuracy downfalls. Their CM method showcases the potential of post-processing as a way to mitigate it.

References

- [1] 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>.
- [0] V. A. Traag, L. Waltman, and N. J. van Eck. “From Louvain to Leiden: guaranteeing well-connected communities”. In: *Scientific Reports* 9.1 (Mar. 2019), p. 5233. ISSN: 2045-2322. DOI: 10.1038/s41598-019-41695-z. URL: <https://doi.org/10.1038/s41598-019-41695-z>.