

Ensemble clustering for graphs: comparisons and applications

Poulin and Théberge [2] further evaluate their previously published ensemble clustering method ECG [1].

They address some well known issues of modularity, including

1. The resolution limit — it tends to underfit when there are lots of clusters
2. Stability between iterations of Louvain

They do this by defining a weighting scheme: given multiple partitions, they form weights through co-association,

$$w_{\mathcal{P}}(u, v) = \begin{cases} w_* + (1 - w_*) \left(\frac{1}{k} \sum_{i=1}^k \mathbf{1}_{(u,v) \in P_i} \right) & \text{otherwise } (u, v) \in 2\text{-core of } G \\ w_* & \text{otherwise} \end{cases}$$

i.e. counting how many partitions place u and v together. On this weighted graph, they then run the Louvain algorithm to obtain a final clustering.

They show very positive results on all grounds, including on a (dubious) evaluation on real networks. ECG is better especially on low signal (e.g. LFR graphs with high mixing). Moreover, they show applications of their weighting beyond obtaining a partition. For example, the kurtosis of the weighting distribution is a good indicator for community structure within a network. In addition, this can be used for community search by using the weights to focus on specific seeds.

There are other issues of modularity, such as connectivity and homogeneity of clusters. Does ECG improve on these at all?

References

- [1] Valérie Poulin and François Théberge. “Ensemble Clustering for Graphs”. In: *Complex Networks and Their Applications VII*. Springer International Publishing, Dec. 2018, pp. 231–243. ISBN: 9783030054113. DOI: 10.1007/978-3-030-05411-3_19. URL: http://dx.doi.org/10.1007/978-3-030-05411-3_19.
- [2] Valérie Poulin and François Théberge. “Ensemble clustering for graphs: comparisons and applications”. In: *Applied Network Science* 4.1 (July 2019). ISSN: 2364-8228. DOI: 10.1007/s41109-019-0162-z. URL: <http://dx.doi.org/10.1007/s41109-019-0162-z>.