Evaluating Max-Min Modularity for Community Detection in Social Networks

Cassandra Spath

A community in a graph can be described as a set of nodes with more connections between them that have sparse connections with other nodes from other communities. These communities will be detected using max-min modularity [1]. This algorithm uses modularity which compares the difference between the connections in the network and the expected connections in a randomly distributed network [2]. It also uses the used-defined related node pairs in a community compared to the expected pairs. The pairs that are connect have positive community structure which traditionally pairs that aren't have negative community structure. However, by dividing the pairs that aren't connected into two groups, those that are definitely not related and those that might be related, only the unrelated pair will be penalized during scoring [1].

The node pairs can be used to modify the modularity algorithm resulting in the max-min modularity. After using max-min modularity, a hierarchical clustering algorithm can be used for optimization [1]. By merging pairs based on the pairs with the highest modularity, the modularities can be recalculated after each merge until there are no merges that will increase the modularity. This will result in the communities of the network. The performance of this algorithm can be tested against the ground truth communities for networks to determine the accuracy. Additionally, the traditionally modularity algorithm will be run to determine the effect of the pairs on the modularity metric.

Three datasets will be used for community detection. The karate club which is a mid-size network with 34 vertices and 78 edges. The communities are based on the divisions of a karate club when the instructor and president had disagreements that caused them to break into two different clubs [3]. The sawmill communication network which is based on the communication of sawmill workers. A change in management resulted in compensation changes and strikes. This network has ground truth communities based on the ages and languages spoken by the employees. This network includes 24 vertices and 38 edges [4]. The last network is the Mexican political elite network. This network has 35 vertices and 117 edges. The communities of this network are based on the occupations of the pollical elite (either military or civilian) [5]. Since these datasets are similar in size, these can evaluate the effectiveness of the clustering compared to the ground truth for mid-sized networks.

[1] Chen, Jiyang, Osmar R. Zaïane, and Randy Goebel. "Detecting communities in social networks using max-min modularity." *Proceedings of the 2009 SIAM international conference on data mining*. Society for Industrial and Applied Mathematics, 2009.

[2] Newman, Mark EJ, and Michelle Girvan. "Finding and evaluating community structure in networks." *Physical review E* 69.2 (2004): 026113.

[3] Zachary, Wayne W. "An information flow model for conflict and fission in small groups." *Journal of anthropological research* 33.4 (1977): 452-473.

[4] Michael, Judd H. "Labor dispute reconciliation in a forest products manufacturing facility." *Forest products journal* 47.11/12 (1997): 41.

[5] Gil-Mendieta, Jorge, and Samuel Schmidt. "The political network in Mexico." *Social Networks* 18.4 (1996): 355-381.