Comparison of Clustering and Community Detection Algorithms

Erika Griechisch

The datamining became an important field in the last few years. Usually every single data (record) connects to a person, and people have relationships between each other. After determining these relationships, we can use graph representation to understand the connections in the background. In graph based datamining, we can determine the important points and the significant groups in networks.

In our study we examined and tested some of the clustering algorithms and community detection methods on a few small world graphs. The clustering algorithms which we investigated are the algorithm based on edge betweenness, the leading eigenvector method, the label propagation method, the greedy modularity maximalization method, and the spinglass method based on statistics. Moreover we present the results of two community detection algorithms: the clique percolation method and the algorithm based on N^{++} sets.

The modularity is an important measure of a particular division of network. We gave an extension of the modularity based on the definition of fuzzy partition matrix. Our aim was to maximize the extended modularity, because optimum determines a fuzzy partition matrix, where every row of the matrix represents the membership values of a node.

The optimum of the extended modularity on the three test graphs usually gave a strict partition, thus we got a binary fuzzy partition matrix. In some cases we got overlaps.