Group Project, Development and Evaluation of Graph Clustering Algorithms

Report Due Date: 11:59pm, November 29 (Sunday), 2015

Purpose

- Practice of understanding existing graph clustering algorithms
- Practice of developing novel approaches having competitive performance

Description

In class, three categories of graph clustering algorithms were discussed: density-based algorithms, partition-based algorithms, and hierarchical algorithms. You implemented a density-based algorithm in Assignment-5. **Implement two more existing algorithms** in any category. **Develop your own algorithm** that is technically sound. When a protein-protein interaction network is used as input of the algorithms, the clusters as output represent potential protein complexes. **Justify that your own algorithm has competitive performance**. Use the f-measure to evaluate the clustering results by comparing to protein complex data provided. As the accuracy of the algorithms, compute an f-score for each output cluster by selecting the maximum f-score to a protein complex, and average the f-scores of all output clusters.

Protein-Protein Interaction Data Set

Yeast protein-protein interactions. Each row represents the interaction (edge) between two proteins (vertices).

Protein Complex Data Set

Yeast protein complexes identified by high-throughput and small-scale experimental methods. Each row represents a protein complex. The first column is an ID.

Submission

- A report in a research paper format

Note

- Submit your report by email attachment (requirement).
- Present your work on December 1 or December 3 (requirement). Check your presentation schedule [here](#).
- This project will be graded by the proposed algorithm (novelty and technical quality), performance evaluation, report, and presentation.