CS 292F Final Project Proposal

Mert Kosan

May 16, 2021

Project Description.

The project will implement Spectral Clustering algorithm. I conduct experiments on different graph datasets and with their different representations. I will work alone in this project which will be implemented on Python 3.

Experiments.

The project will consist of the quality and computational cost comparison of spectral algorithm on different representation of graphs. I use L, N, and A (Laplacian form, Normalized Laplacian Form, and Adjacency Matrix Form respectively) as different representations for each dataset.

L and N are defined as the following, where D is a diagonal degree matrix.

$$L = D - A$$

$$N = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$$

Datasets.

The project will be able to use any matrix structure dataset (e.g. graphs). Tabular datasets can be also represented as similarity graphs that can be fed into the algorithm. I want to make the algorithm is flexible for getting any datasets the user provides. However, I will use some graph datasets to explain in my final report. I am also open for suggestions of any dataset that needs to be explored in the scope of this project and the final report. I'm thinking to extract datasets from the following websites:

- https://sites.stat.washington.edu/spectral/datasets.html
- http://cs.joensuu.fi/sipu/datasets/
- Community Detection Datasets: http://snap.stanford.edu/data/index.htmlcommunities

Deliverables.

I share GitHub repository (https://github.com/mertkosan/spectral-clustering) with all deliverables which is necessary to explain the project. I will share the below items on GitHub repository as well as Gradescope if needed.

- Source Code (i.e. Python Implementation)
- Final Report
- Final Datasets (or their links)