Spectral Histogram Project

Jia Zhang Supervisor: Professor David Bindel

The goal of this project is to generate spectrum of the Laplacian or adjacency matrices. Standard O(N^3) methods to compute the eigenvalues of a matrix are prohibitively expensive for graphs with more than a few
thousand nodes. But histograms or kernel density estimates of the eigenvalue density are useful visualizations, and these can be estimated without computing every eigenvalue[1]. This project is to implement a python estimate eigenvalue distributions with kernel polynomial method in an inexpensive way.

There are three major goals for this project:

1. Filtered moment computation code in Python (target: mid-October)
2. Histograms over wide variety of data sets (target: late October)
3. Classification (target: late November)
   (a) do unsupervised clustering of the graphs (e.g. via KNN)
   (b) do supervised classification based on labels in the repositories regarding graph types

During this semester, I have finished Majority of the goals. I have finished filtered moment computation code in Python, and added some extra utility function to help timing, reading and writing files, and generate eigenvalue for small datasets. This python function can perform reading smat files, smat.gz files, estimate Chebyshev moments, plot the integral of the density estimate based on first-kind Chebyshev polynomials, compare an eigenvalue histogram to an estimated histogram, jackson filtering, map adjacency to normalized adjacency, and converge original dataset files into smat.gz format files. I have collected many new datasets in a wide variety and formalized them into required form. The new data varies from animals, authorships, to citations and communications. I have also finished a k-means clustering function to cluster different datasets based on the calculated filtered moments. This function can take in all the datasets, generate sums of squared errors vs. the number of clusters and assign clusters to each input dataset. The only part that is not finished is the supervised classification function, due to the limit of time.

Resource:
https://github.com/jz685/MEngProj.git

Reference:
[1]David S. Bindel & David Gleich, Kernel polynomial estimates of graph spectra