

ID2222 – Data Mining

# Homework 4: Graph Spectra

Group 10

Xin Ren

Nikhil Dattatreya Nadig

2018-11-30

## Introduction

In this homework, we studied, implemented and tested the spectral graph clustering algorithm as described in the paper “On Spectral Clustering: Analysis and an algorithm” by Andrew Y. Ng, Michael I. Jordan, Yair Weiss. Using our implementation of the K-eigenvector algorithm, we analysed two sample graphs.

1. A real graph “example1.dat” -- This dataset was prepared by Ron Burt. He dug out the 1966 data collected by Coleman, Katz and Menzel on medical innovation. They had collected data from physicians in four towns in Illinois, Peoria, Bloomington, Quincy and Galesburg.
2. A synthetic graph “example2.dat”.

The implementation was done with Matlab.

## SpectralClustering.m

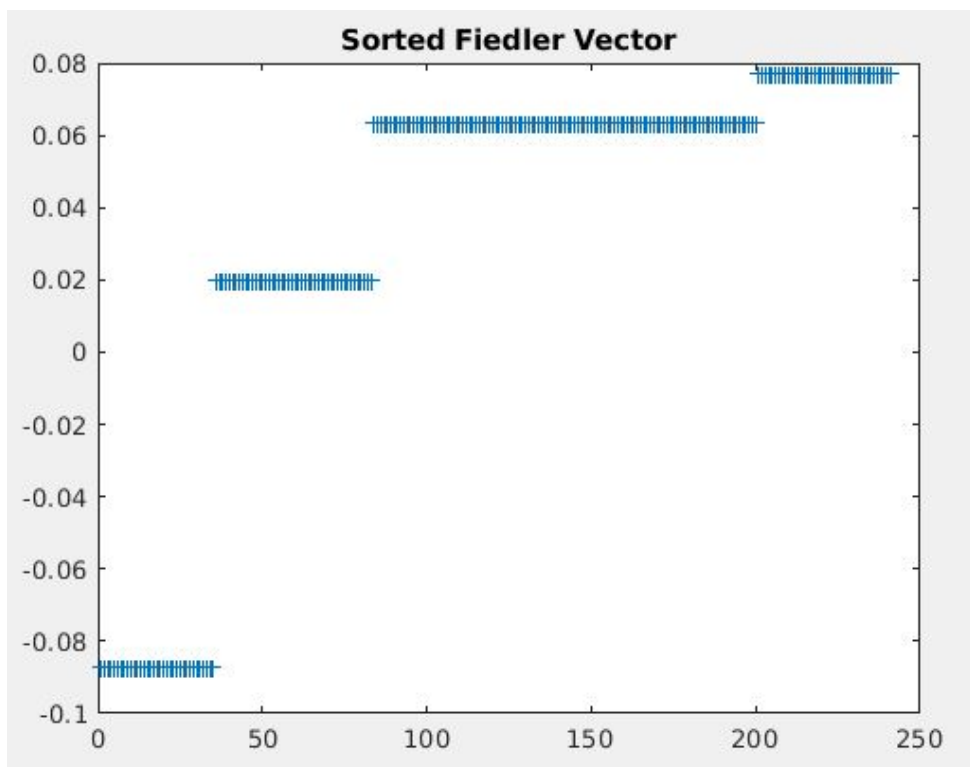
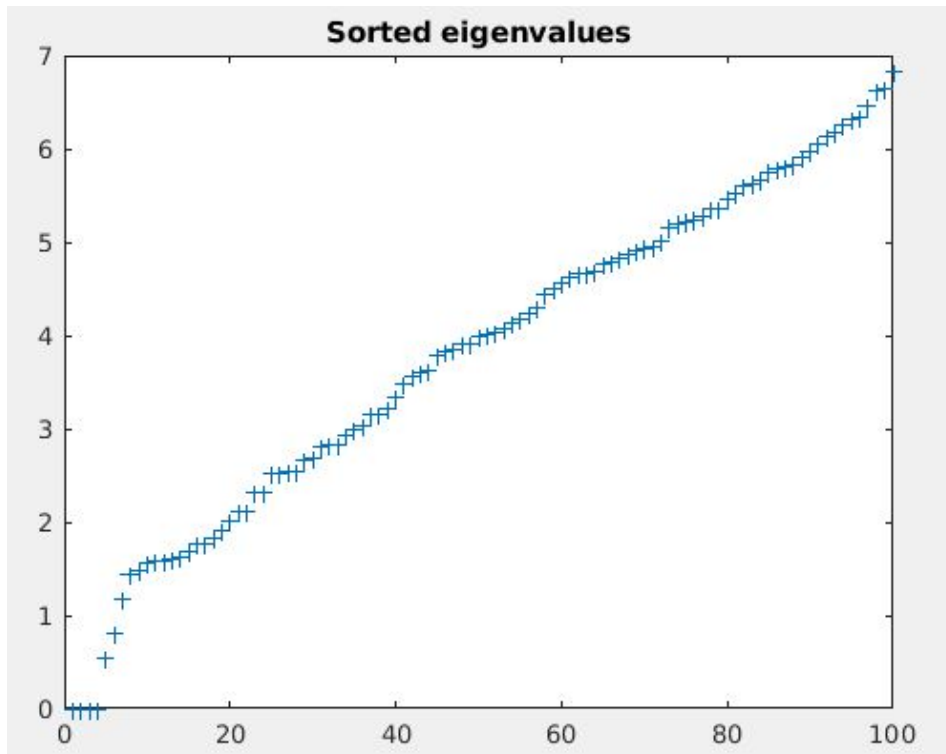
This implements the algorithm in the paper. It generates the following figures:

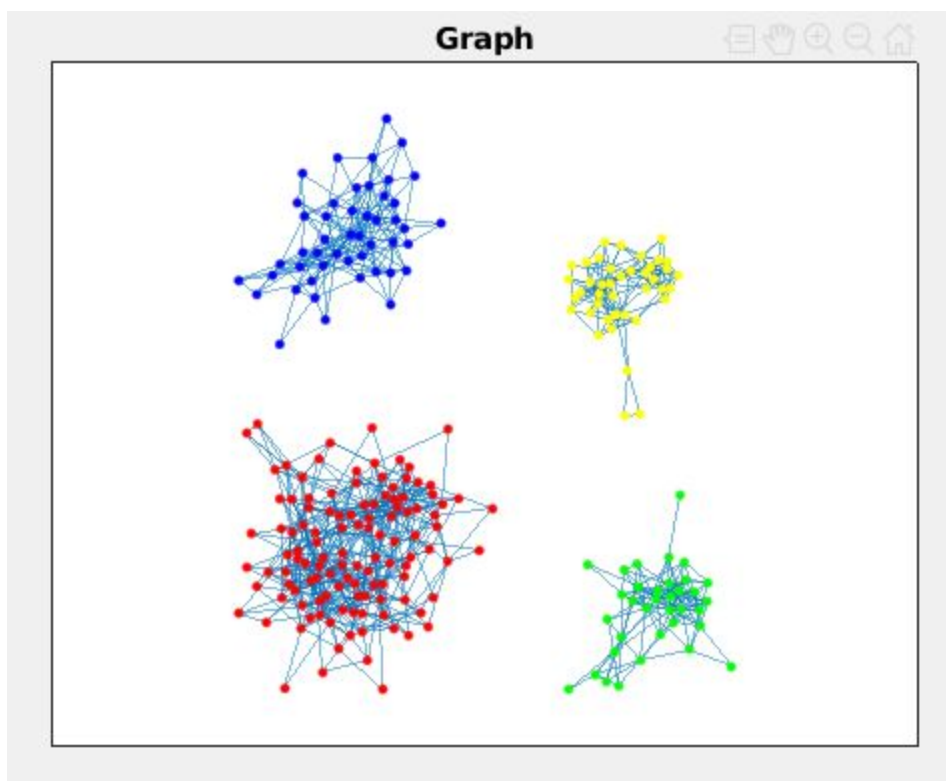
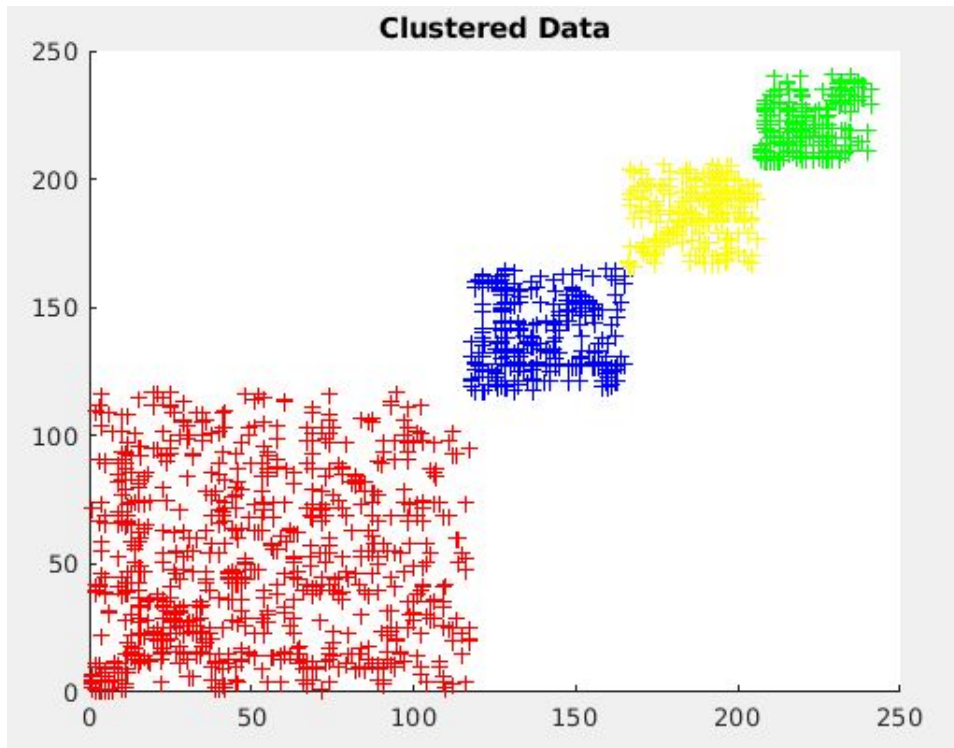
1. Graph figure with nodes highlighted according to clustering result from the algorithm
2. Adjacency matrix figure with edges highlighted (Clustered Data) according to clustering result from the algorithm
3. Figure of sorted eigenvalues of Laplacian matrix
4. Figure of Sorted Fiedler Vector of Laplacian matrix

## Result

1. example1.dat

From the Sorted Fiedler Vector graph, we can see that there are 4 communities so we picked K as 4.





2. example2.dat

From the Sorted Fiedler Vector graph, we can see that there are 2 communities so we picked K as 2. These 2 communities are highly connected.

