Testing the effectiveness of the fiedler vector binary clustering

Abstract

PURPOSE:Test the effectiveness of the fiedler vector as a method for binary clustering for small data sets

METHODS: The fiedler vector method was implemented and tested on the given data set. The number of edges between sets (transcluster edges) and edges within sets (intracluster edges) were recorded. The more effective method was the method that reduced the amount of transcluster edges.

RESULTS: The fiedler vector showed clear signs to much more effective with proportionally more intracluster edges when compared to the arbitrary clustered

CONCLUSIONS: The Fiedler Vector method resulted in a very effective binary clustering of data and should be implemented when binary data clusters are required.

Introduction

The objective was to determine whether or not the fiedler vector method of binary clustering is effective on small sets of vector data. Binary clustering is the action of grouping a set of data into specifically two distinct groups, minimizing the connection between groups. The fiedler vector method is the method of using the fiedler vector (the eigenvector that corresponds to the second smallest eigenvalue), and using the sign of the value to group the nodes into two distinct sets. The method will be tested to see if it can be used to create a binary clustering that has a significantly reduced amount of edges that span between the clusters and an increased amount of edges that span only one cluster when compared to an arbitrarily generated cluster.

Methods

The computations for the creation of the binary clustering were done in MatLab. For the fiedler vector method, the starting edge list was transformed into an adjacency matrix that represents the graph. Then a degree matrix was formed by putting the degrees of the nodes on the diagonal of a n x n zero matrix the same size as the adjacency matrix. Subtracting the adjacency matrix from the degree matrix yielded the laplacian matrix. The eigenvectors and eigenvalues of the matrix were found using the corresponding matlab functions. The second eigenvector corresponding to the second smallest eigenvalue was extracted. Nodes corresponding to the negative values were placed in one set, nodes corresponding to positive numbers were played in the other set.

The number of intracluster and transcluster for the fiedler clustered grouping was done by using MatLab to gather the total amount of edges. The transcluster edges were gathered visually from the output of the binary clustering. The number of edges for the arbitrary cluster was gathered in Python.

Results

Table 1: Shows the corresponding nodes for each set, the number of intracluster and transcluster edges for both models

Set	Feidler Clusters	Arbitrary Clusters
A	3,4,5,6,7,8,11,14,18,19	2,4,7,8,9,10,13,15,16,19
B	1,2,8,10,12,13,15,16,17,20	1,3,5,6,11,12,14,17,18,20
Intracluster Edges	78	24
Transcluster Edges	8	62

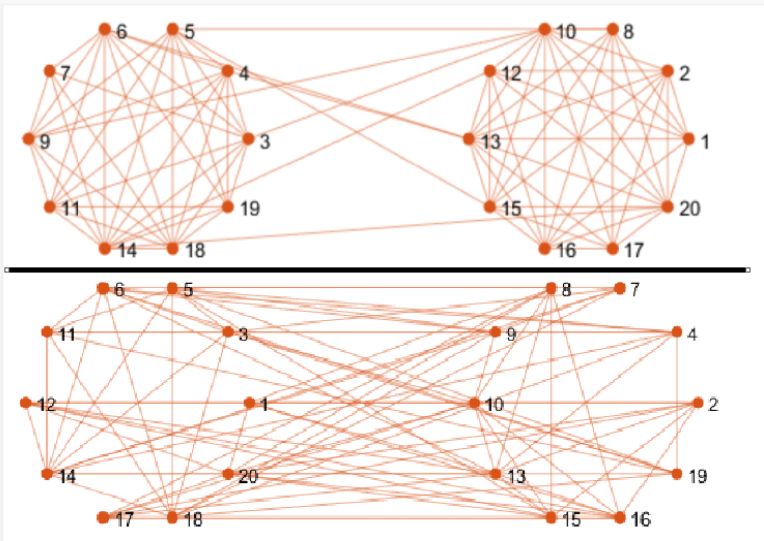

Figure 1: On top is a diagram of the fiedler clustered graph, On bottom is the arbitrarily clustered graph.

Discussion

The amount of edges that span clusters is very low compared to the amount of connections within clusters with a ratio of 39:4 of intracluster edges to transcluster edges. The method created a clustering that allowed for 90.70% of the edges to be within the clusters. This is compared to the 12:31 ratio of intracluster edges to transcluster edges. This results in the percent of intracluster edges being 27.91%. These factors point to the method being significantly better than arbitrary clustering, and decreasing transcluster edges from 72.09% down to 9.30%.

In conclusion, the Fiedler Vector method is a practical and successful way to create meaningful binary clusters for smaller data samples.