Linear Data Analysis

January 20, 2022

Identifying Binary Clusterings Of Vertices Using The Fielder Vector

Cain Susko

20244352

Abstract

This paper will show viability and process to group a graph into visual clusters given the set of edges of the graph. The algorithm used to derive the clusters was provided during cisc271 Lectures and was first documented by Miroslav Fielder in 1989, called the 'Algebraic Connectivity' of a graph (Ellis, Lecture Videos).

Introduction

The scientific objective was to demonstrate if it is possible to derive clusterings of vertices from a graph, given only a list of its edges.

This objective requires the use of numerous concepts found in Linear Algebra & Graph Theory. Particularly, the adjacency, degree, and Laplacian matrices as well as characteristic vectors and the Fielder vector. This objective is important as the clustering of vertices can be a subtle but important factor in the relations between nodes. an example of its use is in the tracking of precipitation in Australia (Tan et al., 2005).

The scientific question was to evaluate if using the Fielder vector and the algorithm for Algebraic Connectivity is a viable way to cluster a graph. This was done by creating a program in MatLab which uses said algorithm and then testing if the results of the said program correspond to a correct solution.

Methods

the algorithm used to find a Laplacian Matrix & binary clusters of a graph given its edges is:

1. Generate an Adjacency matrix from the given list of edges

2. Calculate the Degree matrix of the graph by first multiplying the Adjacency matrix by the ones vector to obtain the degree vector, and then diagonalizing it to become the Degree matrix

3. Using the Adjacency & Degree matrices calculate the Laplacian matrix, which is the difference of the Degree matrix minus the Adjacency matrix.

4. The Fielder vector is now found as the second eigenvector of the Laplacian matrix.

5. Finally, partition the Fielder vector into 2 groups of negative and positive values, the index of these values correspond to vertices while their value (+,-) determines which cluster on the graph the vertex is in.

the method with which the code was tested was to be given a control-data-set as well as the correct output for this data. Using these randomly generated files given by the course instructor (testedge.txt, testsets.txt, data and result respectively) (Ellis, a1.zip) it is possible to determine if the program is running as intended or not by comparing the output of the program to testsets.txt when using testedge.txt as the input. to initialize test data for input into the program one must enter 'load(testedge.txt)', there should appear a 'testedge' variable in the Work space window. Once this is done one must then run the code by executing 'a1_20244352(testedge)' in the command window and comparing the output vector 'ans' with 'testsets.txt'.

The results of this paper and its corresponding program are evaluated on the output of said program, using the data set exclusive to Cain Susko (20cjbs.txt) (Ellis, a1.zip). The evaluated outputs are a figure of the graph before and after clustering & a table of the vertices split in two groups after clustering.

Results

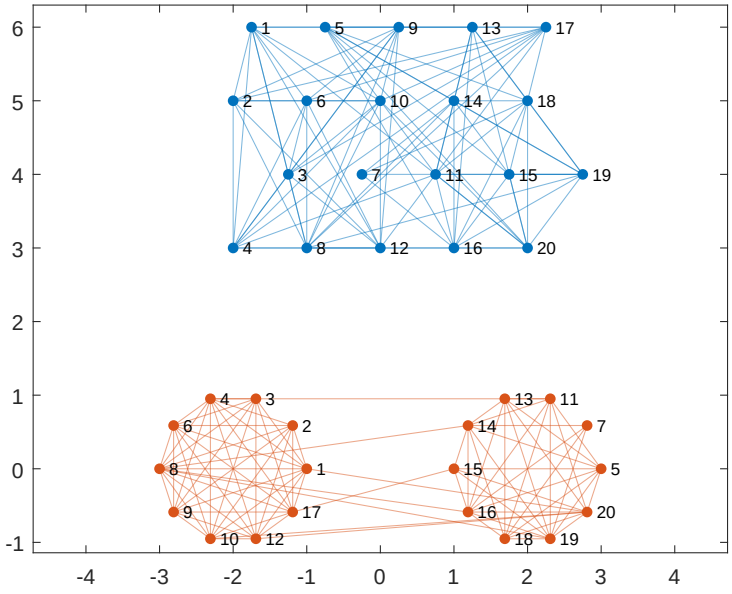

Figure 0.1: un-clustered and clustered graph derived from 20cjbs.txt

Below is a table that displays the vertices of each group which can bee seen as the bottom graph in Figure 0.1

Set 1	5	7	11	13	14	15	16	18	19	20
Set 2	1	2	3	4	6	8	9	10	12	17

Discussion

It can bee seen from the results that it is indeed possible to derive binary clusters of a graph solely from its edges.

From analyzing this data, it is possible to infer that these clusters can represent two independent but linked systems that can interact with each other through the pairwise nodes connecting them. Moreover this process could reveal unknown and important relations within large data sets that could aid in the research or management of the data.

The final result of this paper is that it is viable to use the Fielder Vector and the algorithm for Algebraic Connectivity to derive the binary clusterings of a graph.

Sources

Tan P-N, Steinbach M, Kumar V. Intro to Data Mining. 1st ed. New Jersey: P. Ed Australia; 2005.

Ellis R. Lecture Videos [Internet]. CISC271, Linear Data Analysis: Lectures. Queens University; 2022 [cited 2022Jan20]. Available from: https://research.cs.queensu.ca/home/cisc271/lectures.html

Ellis R. a1.zip [Internet]. CISC 271 Assignments. Queens University; 2022 [cited 2022Jan20]. Available from: https://onq.queensu.ca/content/enforced/639984-CISC271W22/myhtml/assignments.html