ABSTRACT

PURPOSE:

Use the Fiedler vector of a Laplacian matrix of a graph to cluster the vertices of a graph into two sets.

METHODS:

Input the array list (elist), use for loop to initialize the Laplacian matrix for the graph and output a vector that indicates the clustering of the vertices to two sets (-1 for Set1 and 1 for Set2), then cluster the vertices of the graph.

RESULTS:

Obtained two diagrams. The upper one is the graph in a rectilinear grid and the lower one is the graph when plotted according to a computed clustering of the vertices.

CONCLUSIONS:

Using the Fiedler vector of a Laplacian matrix of a graph can cluster the vertices of graphs in two sets clearly. However, it may not be the most effective method for the data which is more random, more chaotic and has larger differences in reality.

INTRODUCTION:

Cluster the vertices into two sets and make them into two graphs by the Fiedler vector and Laplacian matrix. Test if the Fiedler vector clustering vertices is an efficient method to organize data.

For any graph G(V,E) that has n vertices in V labeled v1,v2,...,vn and that has m edges labeled (v*i* vj), the adjacency matrix A(G) is the n × n matrix for which each entry is defined as

$$a_{ij} = \begin{cases} 1 & \text{if } (v_i\ v_j) \in \mathcal{E} \\ 0 & \text{if } (v_i\ v_j) \notin \mathcal{E} \end{cases}$$.(1)

Laplacian matrix of a graph is for any graph G(V,E) that has an adjacency matrix An×n(G) and a degree matrix Dn×n(G), the Laplacian matrix L(G) is the n × n matrix that is defined as L(G) def = D(G)−A(G). The graph that is used to define the Laplacian matrix is usually understood by context, so it is common to abbreviate Equation L(G) def = D(G)−A(G) as L =D−A.(2)

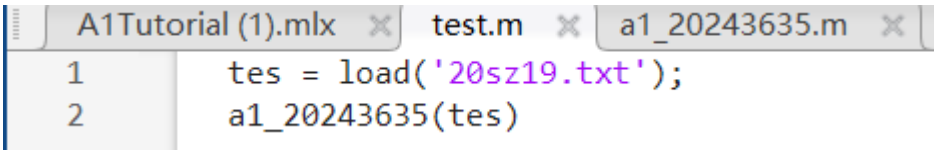
Scientific Question: If we have a larger dataset and get a more complex graph so that it's illegible to humans, is that still an effective method?

I think that depends on whether the diagram is used to display to people or let the machine to storage. To test this question, we could use a sample dataset that is known to be large in size. Then, after clustering the data, we can first see if it still is illegible to humans. If the clustered data becomes too complex for a human to interpret, then we could use a computer to try and make use of the data. If the clustered result is shown to be valid and still may be usable, then we may still consider clustering by Laplacian matrix and the Fiedler vector to be an effective method.

METHODS:

Write the matrix with the computation and initialize the matrix. Create a Laplacian matrix. Use the fiedler vector with the computation. Use for loop to iterate each coordinate and sort them into the correct set. Based on these two sets, plot the graph, cartesian and clustered. Create the graph.

Use the given txt file to test the code.



RESULTS:

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

Table 1: Clusters of vertices for 20sz19.txt. Each row contains the vertices of the cluster.

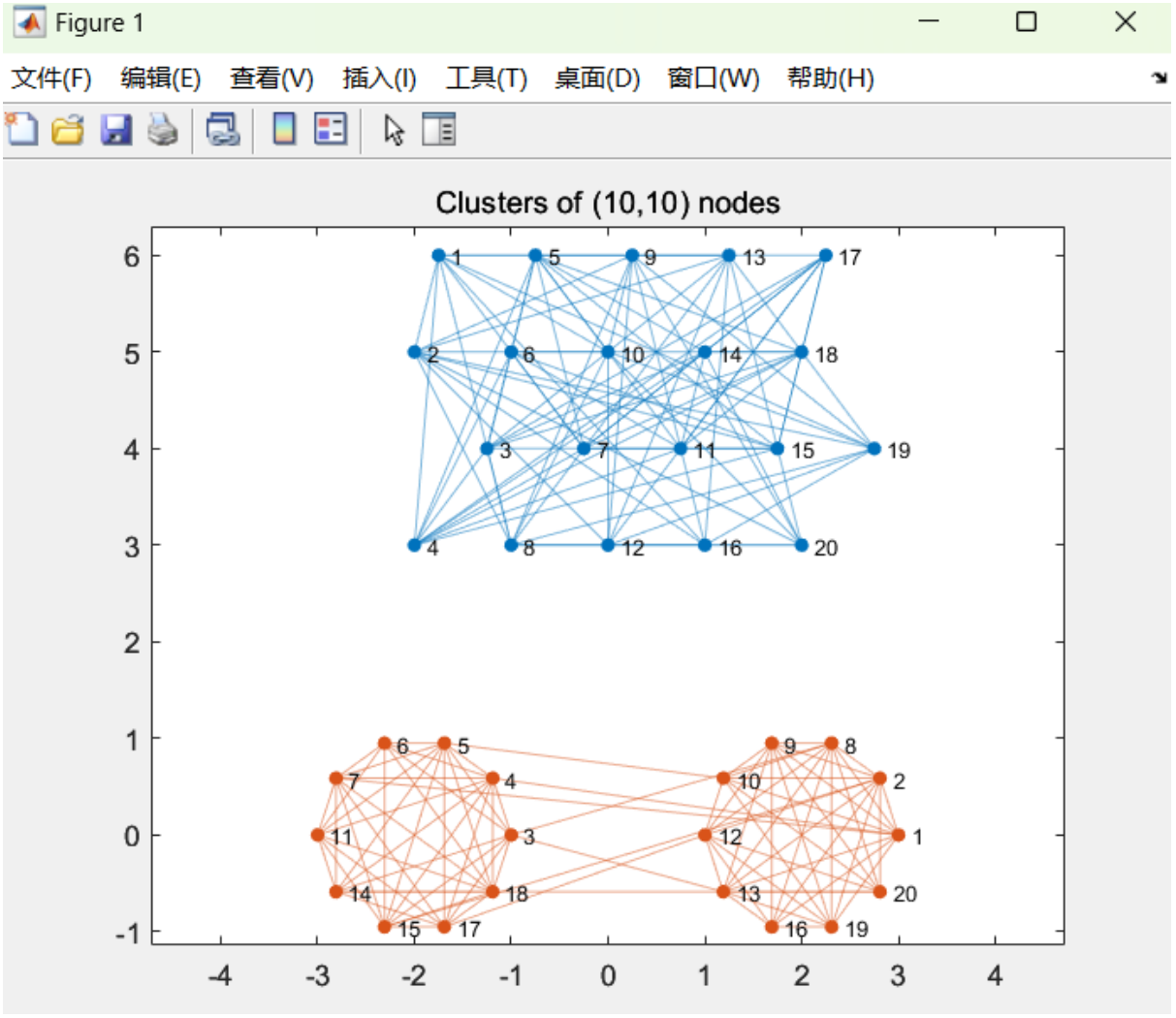


Figure 1: Diagrams of the test graph. The upper part shows the graph in a rectilinear grid. The lower part shows the graph when plotted according to a computed clustering of the vertices.

DISCUSSION:

🥲 It seems organized to sort into two diagrams, but the edges are a little jumbled. If the dataset is larger and more complex, this method may not be very effective.

🥹 It's effective for the given dataset because there's not too much data in, the graph is still legible.

🥹 The capacity of dataset can affect the amount of edges, so when the amount of data going to larger, the edges in graph will become more complex and illegible. Also because it needs to calculate the eigenvalues, the larger amount of data will slow down the code. It would not be effective at clustering real world datasets.

I think it can just cluster a small amount of data.

REFERENCES:

1. Ellis RE. Lecture notes from Week #1, Class 2. 2021.
2. Ellis RE. Lecture notes from Week #1, Class 3. 2021.