Datathon - 4

[1]Arjun Verma : IMT2017008
CS732/DS732: Data Visualization
Course Instructors : Prof. Jaya Sreevalsan Nair
Technical Report - 4

Abstract. *This technical report contains a brief overview of the methodology involved in implementing two different matrix seriation methods and also gives an exposition of the implementation.*

1. Introduction

The provided paper titled "Matrix Reordering Methods for Table and Network Visualization" was the main point of reference for the provided implementations. An attempt was made to implement a number of algorithms and finally the **Spectral Methods** and **Dimension Reduction Approaches** were chosen as the final submissions.

2. Tools and Methodology

2.1. Tools

The exhaustive set of libraries used for generating the final inference visualizations involve :

- numpy
- pandas
- networkx
- matplotlib
- sklearn
- scipy

An additional point to be made here is that the algorithm implementations were done by trial and testing it on the *IRIS dataset* as it was less time consuming to work on an already completed/imputed dataset which required no pre-processing.

2.2. Methodology

In this section, we provide the distinct characteristics involved in each technique.

2.2.1. Methodology for Spectral Ordering

Brief overview of the algorithm :

- Input is a matrix X with n rows and m columns/attributes.
- Generate an adjacency matrix W and visualize the corresponding bipartite graph for this matrix X. The adjacency matrix is created by taking the pairwise euclidean distances amongst the n datapoints and then putting in a value of 1 if this distance is below a certain threshold else 0. This is in attempt to create a symmetric matrix that somewhat group closer datapoints together.
- Compute the Laplacian Matrix L by using the equation L = D - W, where D is the degree matrix and W is the adjacency matrix.
- Compute the Fiedler vector from the Laplacian matrix L by doing an eigen-decomposition and then picking up the vector corresponding to the smallest non-null eigenvalue.
- Re-order the n datapoints according to the sorted order of the computed Fiedler vector.

Key visualizations :

- The original ordering of matrix.
- The generated bipartite graph.
- The re-ordered matrix.

2.2.2. Methodology for Dimension Reduction Approach

Brief overview of the algorithm :

- Input is a matrix X with n rows and m columns/attributes.
- Compute the Covariance Matrix C for X.
- Compute the eigenvectors for the matrix C by doing an eigen-decomposition.
- Pick the first principal component,i.e the eigenvector corresponding to the largest eigenvalue.
- Re-order the n datapoints according to the sorted order of the first principal component.

Key visualizations :

- The original ordering of matrix.
- The re-ordered matrix.

2.3. Sample Visualizations

1. The **figure below** depicts the original ordering of the matrix X.

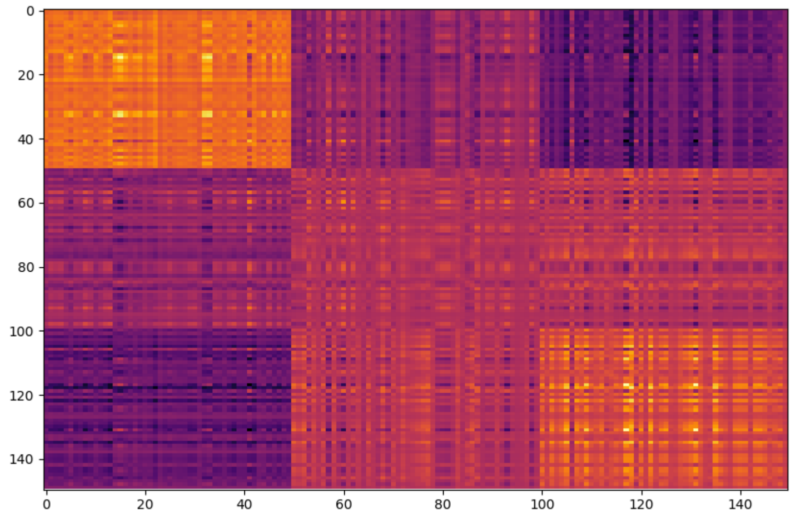

Figure 1. Original Ordering of the dataset

2. The **figure below** depicts the re-ordered matrix according to **Dimension Reduction**.

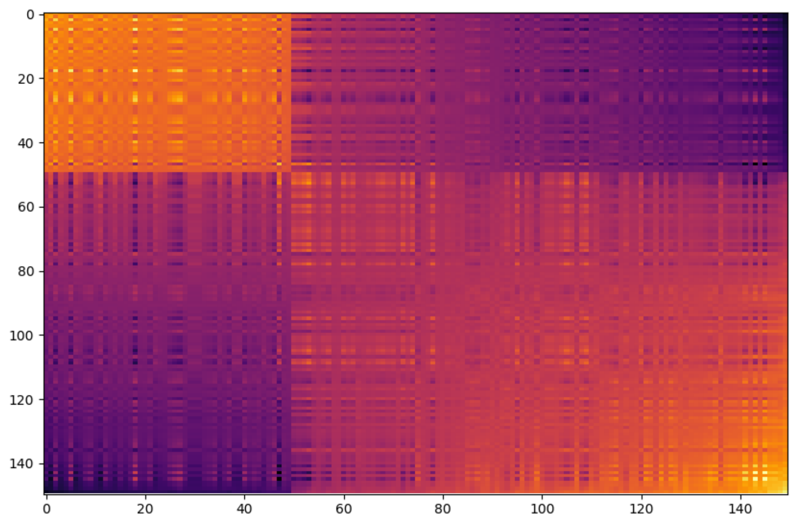

Figure 2. Dimension Reduction Method

3. The **figure below** depicts the computed bipartite graph of the matrix X for the **Spectral Ordering** method.

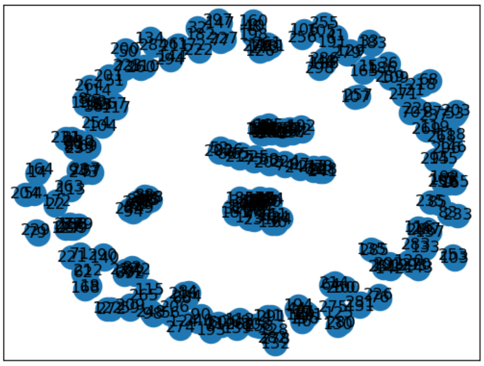

Figure 3. Bipartite Graph

4. The **figure below** depicts the re-ordered matrix X according to the **Spectral Ordering** method.

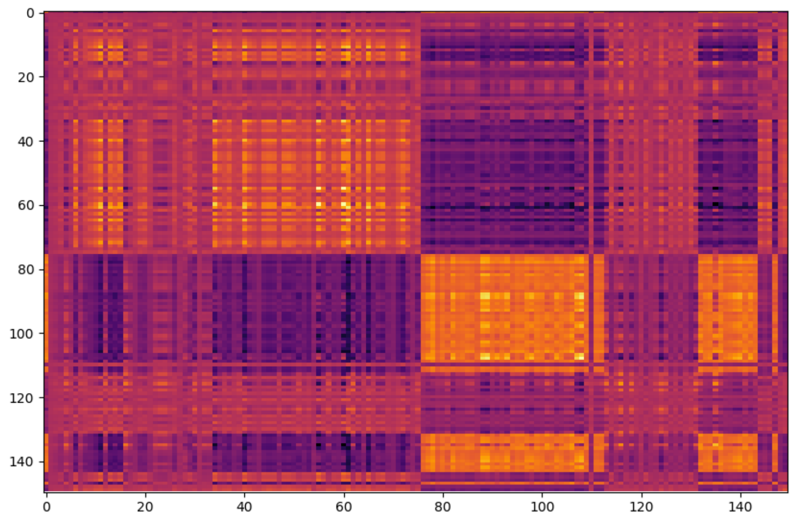

Figure 4. Spectral Ordering Method

3. References

1. Matrix Reordering Methods for Table and Network Visualization
2. Matrix algorithms for the seriation problem
3. History of Cluster HeatMaps
4. Corrgrams: Exploratory displays for correlation matrices
5. Spectral Graph Clustering