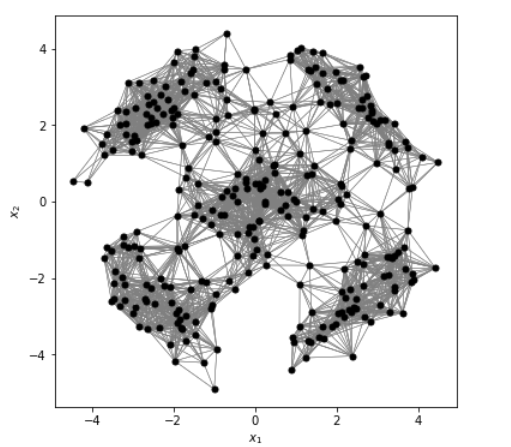
ENGR421 – HW8 REPORT

BERK BAHADIR BAHCETEPE

0071533

For this homework, I started by assigning the data from the .csv file. After I calculated the b matrix by using the formula given in the homework description. Briefly, my algorithm consists of 2 for loop for the rows and columns of the b matrix and assign 1 to the corresponding index if the Euclidian Distance of the each data point is less than the threshold (1.25) and assign 0 otherwise. After, I drew the connectivity graph using the b matrix and obtained a graph as such:



Later, I calculated the D, L and normalized L matrix using the formulas in the lecture notes and homework description. D matrix has the total number of connections for each data point at the diagonals and zeros for the rest of the matrix. L is the difference of D matrix and b matrix and the normalized L matrix which is L\_sym in my code is calculated with the help of the formula given in the description. After calculating L\_sym, I found the eigenvalues and vector of this matrix using the built-in function np.linalg.eig() of python. I sorted the indices of the eigenvalues from the the smallest to the greatest and took the indices of the smallest 2-6th eigenvalues in order to determine the 2-6th eigenvectors of the L\_sym matrix. Using these 5 matrices, I constructed the Z matrix and ran the k-means clustering algorithm on this matrix. In order to do that, I selected the 29, 143, 204, 271, 277th rows of the matrix as the initial centroids and plotted the clustering result graph as such:

Chart, scatter chart

Description automatically generated

In order to plot this, I updated the memberships with the initial centroids and updated centroids in z-plane however, in order to plot the centroids, I also required the updated centroids values in data\_set-plane since I was plotting the data point in data\_set-plane. Therefore, I also initialized a variable called centroids\_X and updated the centroids in both z-plane and data\_set-plane. While the former is used in the calculations of the memberships, the latter is used in plotting the clustering result.

In the end, I obtained two accurate graphs as wanted from me.