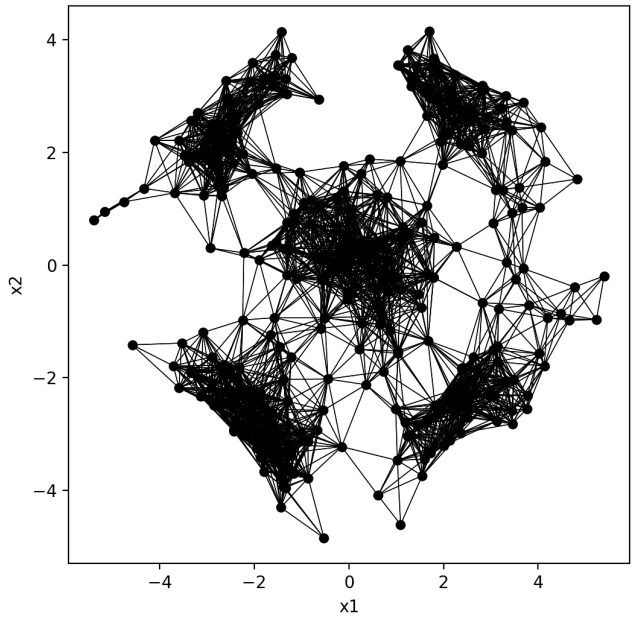
**ENGR 421 - HOMEWORK 6**
**Zeynep Öner**
**ID: 64912**

# Spectral Clustering

In this homework, we implemented a spectral clustering algorithm. We first need to calculate the distances between the data points. The distance used is the l2-norm Euclidean distance. We then form the $B$ matrix using these distances. If the distance is above the threshold $\delta = 1.25$, $B_{ik} = 0$. If not, $B_{ik} = 1$.

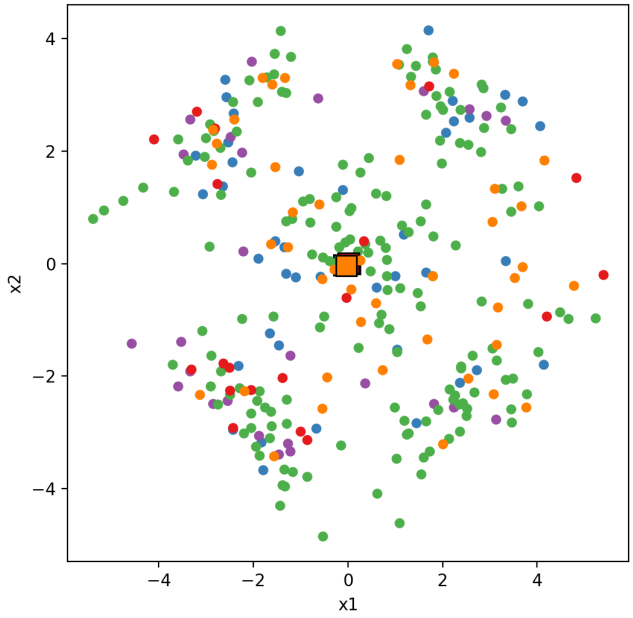
When plotting the connectivity matrix, we draw a line between the data points where $B_{ik} = 1$. The obtained plot is below:




It is a successful connectivity matrix. By using $B$, we then form the $D$ matrix. We count the number of 1 values in a particular row. The $D_{ii}$ value then becomes that count.

Afterwards, we calculate the symmetric Laplacian matrix. The symmetric Laplacian matrix is normalized and is calculated by $L = I - D^{-1/2}BD^{-1/2}$. It seems like my Laplacian matrix is correct, because I also tried using the `csgraph.laplacian` function and my calculation gave the same matrix. So my calculation using the formula above seems to be correct.

The code afterwards is a problem. When calculating the eigenvalues and eigenvectors, I use the `np.linalg.eig` function. It gives an output of very small numbers. Then I sort the eigenvalues, get the indices of the smallest eigenvalues and get the eigenvectors corresponding to those indices. By transposing that matrix, I get the Z matrix which is NxR, what we wrote in the lecture. However, when I plot the Z matrix and X matrix together, Z matrix gives very small values compared to X matrix. So when I do K-means clustering on Z matrix and plot the X matrix according to memberships generated, I get a bad picture.




My guess is that I am doing something wrong when generating the eigenvalues and eigenvectors. I could not figure out the reason behind this.