CS631 Cyber Security of Critical Infrastructure
Homework 6 – Report

Sanjeev Lal (19111077)

Avesh Kumar Agrawal (19111020)

Manish Patel (19111051)

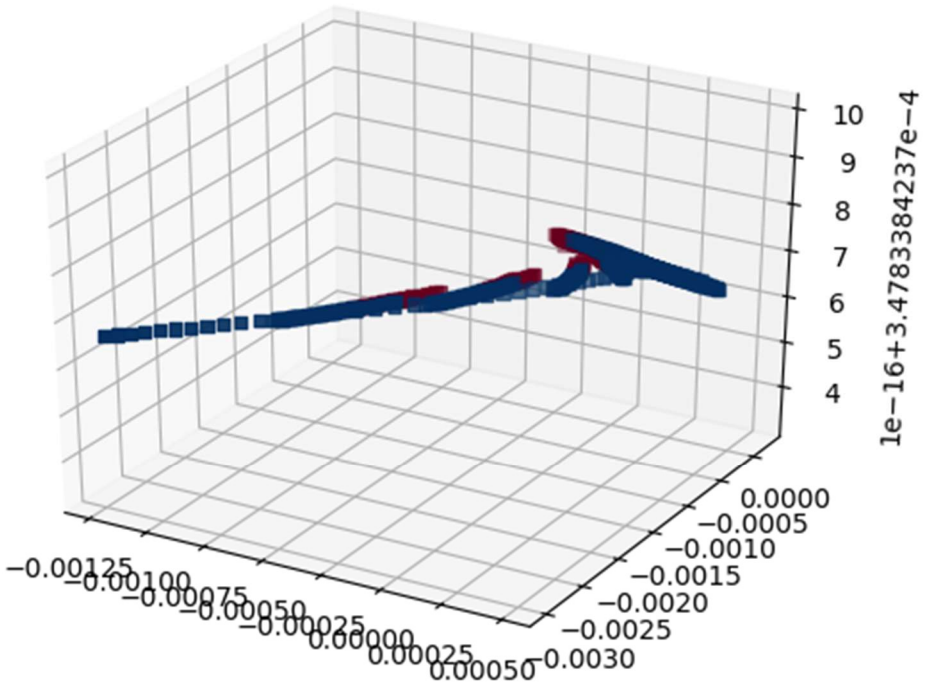
Diffusion Map

1) Diffusion map is a dimensionality reduction or feature extraction algorithm which computes a family of embeddings of a data set into Euclidean space (often low-dimensional) whose coordinates can be computed from the eigenvectors and eigenvalues of a diffusion operator on the data.
2) We have made first the kernel matrix (K) through Gaussian kernel with parameters as "gamma=10/132".
3) D is the diagonal matrix consisting of the row sums of K.
4) Created diffusion matrix (P) by normalizing the rows of kernel matrix.
5) Calculated the eigenvalues and eigenvectors of diffusion matrix (P) with the help of P'.

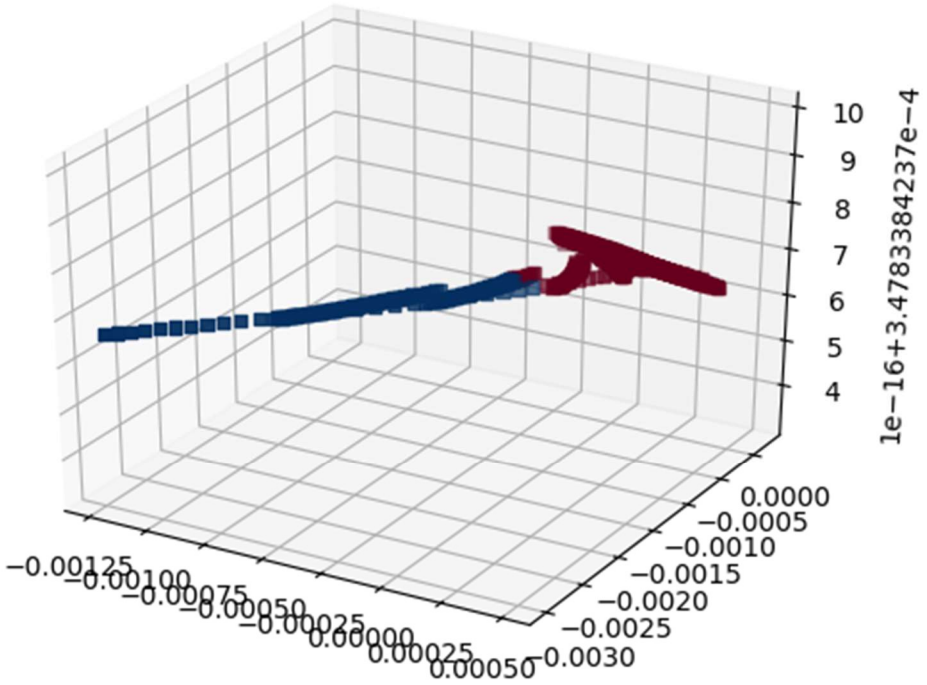
$$where \quad P' = D^{-1/2} K D^{-1/2}$$

6) Took the Path length as t=1.
7) Mapped the data in diffusion space.
8) Performed clustering on all the data points with all dimensions and also performed clustering on all data points with 3-dimensions.
9) Plotted the graph for 3-dimension clustering.

Original cluster of 3-D data



Plot after clustering of 3-D data



Conclusion

- ➔ We should not apply diffusion map over this data as the accuracy is coming very low around **60%**.
- ➔ Data is not well separable in the diffusion space.
- ➔ Inter-cluster distance is very low for this data.