

## Graph Mining CSF426

Lab session 5

Time: 2 pm - 4pm

Date: 16-09-2023

Instructions: All questions need to be answered. **You are required to write programs in jupyter notebook and submit .ipynb.** For theoretical questions, you can type answers in the jupyter notebook itself. There is no need to create a separate text file.

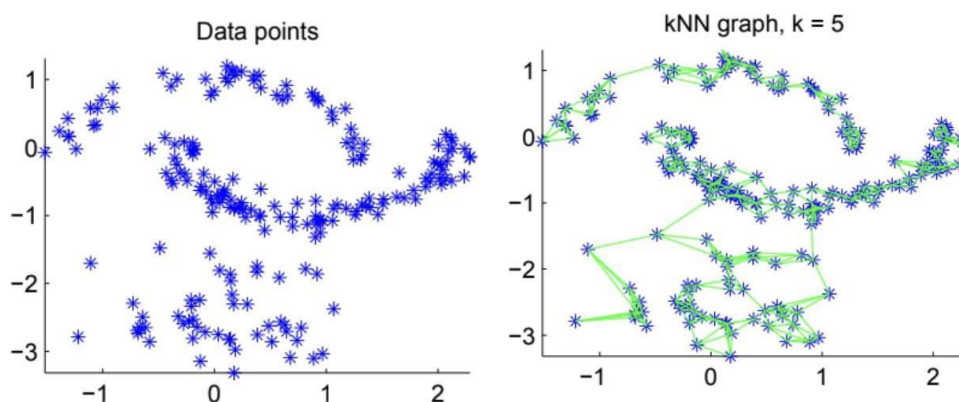
*Note: You are allowed to use your previous code from previous lab sessions for data and part c.*

**[Total Marks =10]**

- a) Create a similarity graph based on *k-nearest neighbor* approach (a reference paper is uploaded on canvas, see page no. 03) and display its variation on different values of  $k$  between 2 to 5. (An example data and corresponding 5-NN graph are shown below on some other dataset). You are free to choose your own similarity/distance metric (preferably depending on the domain the data comes from), and no general advice will be given.

**[3 marks]**

*Note: (For drawing graphs in 2-d plane depicting edges between points, students are allowed to seek help from TAs).*



- b) Construct the weighted adjacency matrices for the k-NN similarity graph on varying values of  $k$  i.e. 2, 3, 4 and 5 and plot the sparsity level of graph (in %) vs  $k$ . **[3 marks]**
- c) Design a spectral clustering algorithm for the 2 moons data using the similarity graph for different values of  $k$  i.e. 2, 3, 4, and 5, created in the last lab session. Display the clustering results by selecting different colors for the data points belonging to different clusters. **[4 marks]**