Graph Mining CSF426
Lab session 6
Time: 2 pm - 4pm
Date: 24-09-2023

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.

[Total Marks =10]

Lab assignment for today is based on **Label Separation Algorithm (LSA)**, based on the research paper "Learning from local and global consistency" (Link)
You are provided with Iris Dataset which contains 150 data points and 3 class labels. Each data point has 4 features.

Steps to perform the experiments are as follows:

1. Extract the features from the dataset corresponding to each data point. You will have a matrix of size (150 * 4).
2. Using Gaussian similarity measure, construct a graph from the data with σ = 0.1

$$W_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) if\ i\ !=j, \qquad W_{ii} = 0\ otherwise$$

3. Extract the labels corresponding to each data point and convert into one-hot vector form. You will get a matrix of size (150 * 3)
4. Split your data into train/test into 70/30 ratio and change the labels for test nodes to 0. That is, a test node will have the corresponding label of (0 0 0) in one-hot encoded form.
5. Using the formula below, predict the labels for test nodes.

$$Y_{pred} = (I - \alpha S)^{-1} Y$$
$$S = D^{-1/2} . W . D^{-1/2}$$

6. Plot and compare the distribution of original and LSA predicted labels for data points.