Graph Mining CSF426

Lab session 9 (evaluative)

Time: 5 pm – 7 pm

Date: 07-11-2024

Instructor IC – Vinti Agarwal

Instructions: All questions need to be answered. **You are required to write programs in jupyter notebook and submit .ipynb on canvas**. Please rename your solutions in format <ID-NAME-LABNO>. 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]**

Objective: To implement Laplacian Eigenmaps based encoder-decoder model to perform task of link prediction.

Dataset: The dataset provided to you is known as "karate club". The dataset has 34 datapoints, each belonging to one of 2 classes.

Steps of implementation:

1. Initialize the nodes with random features or embeddings of size 5.
2. Remove few random edges in graph (test edges). Objective will be to predict the edges between these nodes.
3. Implement Laplacian Eigenmaps decoder and compute the loss.

$$Loss = \left|\left|Z_u - Z_v\right|\right|_2^2 . A[u, v]$$

4. Compute gradient descent and update the embeddings accordingly.

$$emb_i^{t+1} = emb_i^t - \alpha * \left\{\frac{\delta L}{\delta emb_i}\right\}$$

$$where\ learning\ rate, \alpha = 0.005$$

5. Apply constraints of

$$emb^T . D . emb = 1$$

and,

$$emb^T . D . 1 = 0$$

to avoid arbitrary scaling and perform centralization to embeddings.
For more explanation, please refer the paper. *Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering.*

6. Repeat the steps 3 to 5 for 10 iterations.
7. Use the learned embeddings to predict the edges between test nodes using cosine similarity.
8. Compute the accuracy of link prediction.

9. Vary the number of iterations to 20 and 30, and analyse its effect on link prediction accuracy and plot the graph between iterations and accuracy.