Soël MEGDOUD
soel.megdoud@ens-paris-saclay.fr

Lab session # LAB5
ALTEGRAD 2024

11/25/24

Question 1

- DeepWalk could be generelized to directed graphs. We should user outgoing edges to select the next node.
- DeepWalk could be generelized to weighted graphs. We could choose the next node using a probability proportional to edge weights.

Question 2

The two embedding matrices are related through a linear transformation, which is an inversion of the 2nd dimension. Such variations arise due to the stochastic nature of DeepWalk, however the transformation is consistent through the nodes so the structural relationships in the graph, are preserved in both embeddings.

Question 3

- The GCN in Task 10 has 2 message-passing layers. Each message-passing layer aggregates information from the immediate neighbors of a node (1-hop neighborhood). After 2 message-passing layers, each node aggregates information from its 2-hop neighborhood (neighbors of neighbors). Thus, the maximal distance of nodes considered in the calculation of in the prediction for node i) is 2 edges.
- In a GCN architecture with k message passing layers,the maximal number of edges separating the node where we make the prediction and the nodes which are taken into account is k.

Question 4

Normalized Adjacency Matrix $\hat{A}$

- For the Complete Graph K_4:

  The adjacency matrix A for K_4 (with self-loops added) is:

$$\tilde{A} = A + I = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

  The degree matrix $\tilde{D}$ is:

$$\tilde{D} = \mathrm{diag}(4, 4, 4, 4).$$

  The normalized adjacency matrix $\hat{A}$ is:

$$\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2} = \left(\frac{1}{2}I\right)\tilde{A}\left(\frac{1}{2}I\right) = \frac{1}{4}\tilde{A} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}.$$

- For the Star Graph S_4:

  The adjacency matrix A (with self-loops added) is:

$$\tilde{A} = A + I = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

The degree matrix $\tilde{D}$ is:

$$\tilde{D} = \mathrm{diag}(4, 2, 2, 2).$$

$$\hat{A} = \begin{bmatrix} 0.25 & 0.3536 & 0.3536 & 0.3536 \\ 0.3536 & 0.5 & 0 & 0 \\ 0.3536 & 0 & 0.5 & 0 \\ 0.3536 & 0 & 0 & 0.5 \end{bmatrix}.$$

$Z_0 = \text{ReLU}(\hat{A}XW_0)$

$$XW_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -0.8 & 0.5 \end{bmatrix} = \begin{bmatrix} -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \end{bmatrix}.$$

- For K_4:

$$\hat{A}XW_0 = \hat{A} \begin{bmatrix} -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \end{bmatrix} = \begin{bmatrix} -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \end{bmatrix}.$$

$$Z_0 = \text{ReLU}(\hat{A}XW_0) = \begin{bmatrix} 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}.$$

- For S_4:

$$\hat{A}XW_0 = \begin{bmatrix} -1.0486 & 0.6554 \\ -0.6829 & 0.4268 \\ -0.6829 & 0.4268 \\ -0.6829 & 0.4268 \end{bmatrix}.$$

$$Z_0 = \text{ReLU}(\hat{A}XW_0) = \begin{bmatrix} 0 & 0.6554 \\ 0 & 0.4268 \\ 0 & 0.4268 \\ 0 & 0.4268 \end{bmatrix}.$$

$Z_1 = \text{ReLU}(\hat{A}Z_0W_1)$

- For K_4:

$$Z_0W_1 = \begin{bmatrix} 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.1 & 0.3 & -0.05 \\ -0.4 & 0.6 & 0.5 \end{bmatrix} = \begin{bmatrix} -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \end{bmatrix}.$$

Then,

$$\hat{A}Z_0W_1 = \begin{bmatrix} -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \end{bmatrix}.$$

Applying ReLU:

$$Z_1 = \text{ReLU}(\hat{A}Z_0W_1) = \begin{bmatrix} 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \end{bmatrix}.$$

- For S_4:

  $Z_0 W_1$:

$$\text{For node 0: } [0, 0.6554]W_1 = \begin{bmatrix} -0.2622 & 0.3932 & 0.3277 \end{bmatrix}.$$

$$\text{For nodes 1-3: } [0, 0.4268]W_1 = \begin{bmatrix} -0.1707 & 0.2561 & 0.2134 \end{bmatrix}.$$

  Next, $\hat{A} Z_0 W_1$:

$$Z_1 = \text{ReLU}\left(\begin{bmatrix} -0.2468 & 0.3700 & 0.3085 \\ -0.1781 & 0.2671 & 0.2226 \\ -0.1781 & 0.2671 & 0.2226 \\ -0.1781 & 0.2671 & 0.2226 \end{bmatrix}\right) = \begin{bmatrix} 0 & 0.3700 & 0.3085 \\ 0 & 0.2671 & 0.2226 \\ 0 & 0.2671 & 0.2226 \\ 0 & 0.2671 & 0.2226 \end{bmatrix}.$$

The representations Z_1 reflect the structural differences in the graphs:

- In K_4, all nodes are structurally identical, leading to identical representations.
- In S_4, the central node has a higher degree and different connectivity, resulting in a distinct representation compared to the outer nodes.

If we had randomly sampled node features X from a random uniform distribution, the initial feature variability would propagate through the network, resulting in more diverse node representations in Z_1, and the structural patterns observed would be less pronounced.

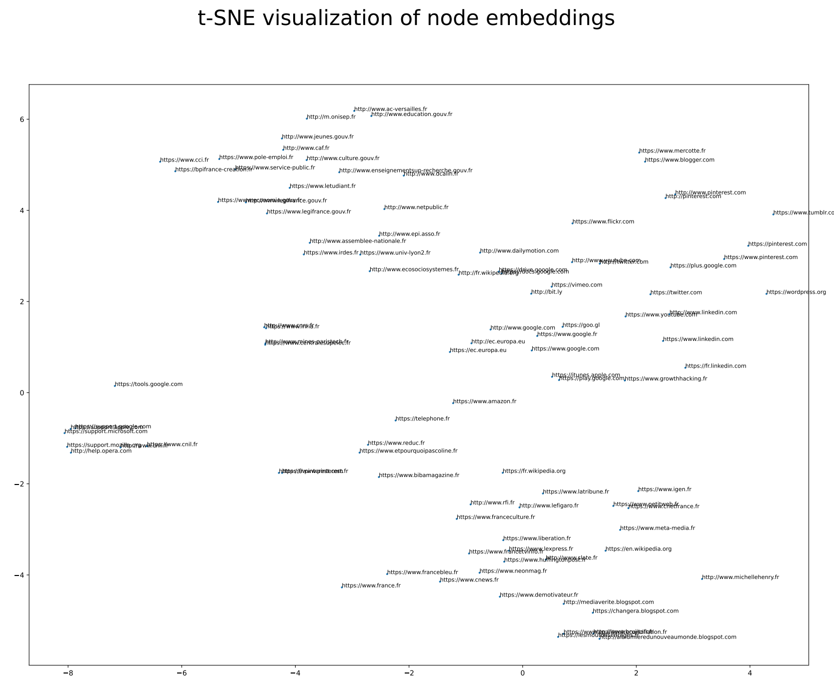

Figure 1: French Web

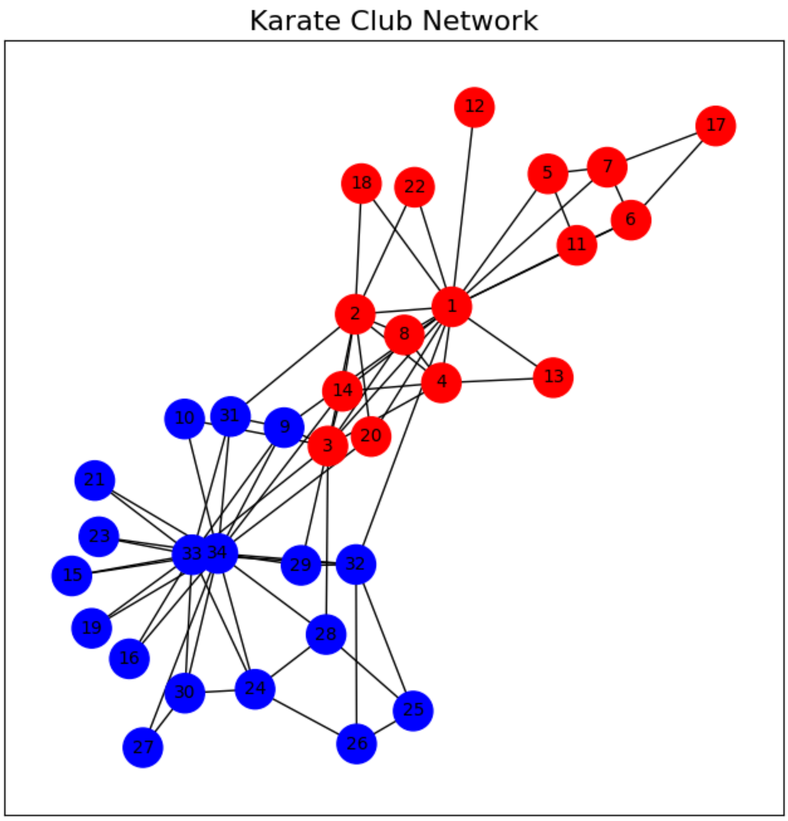

Figure 2: Karate Network Classification

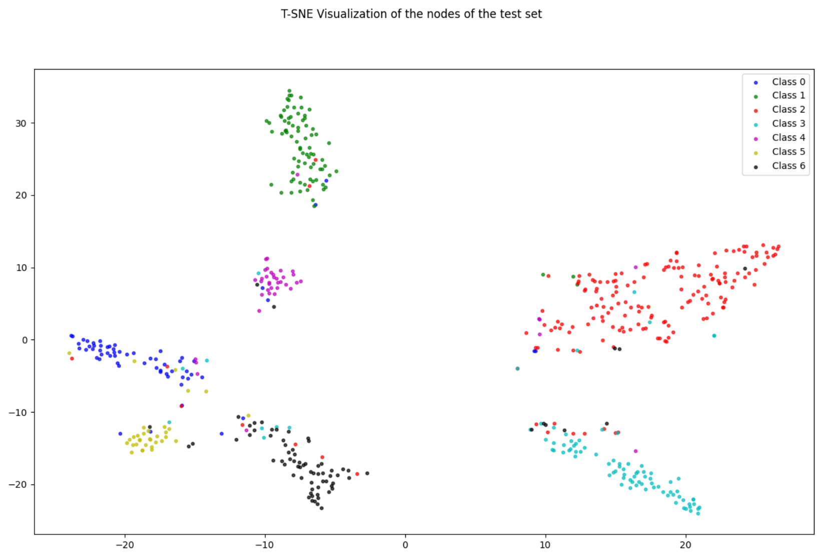

Figure 3: Cora Classification