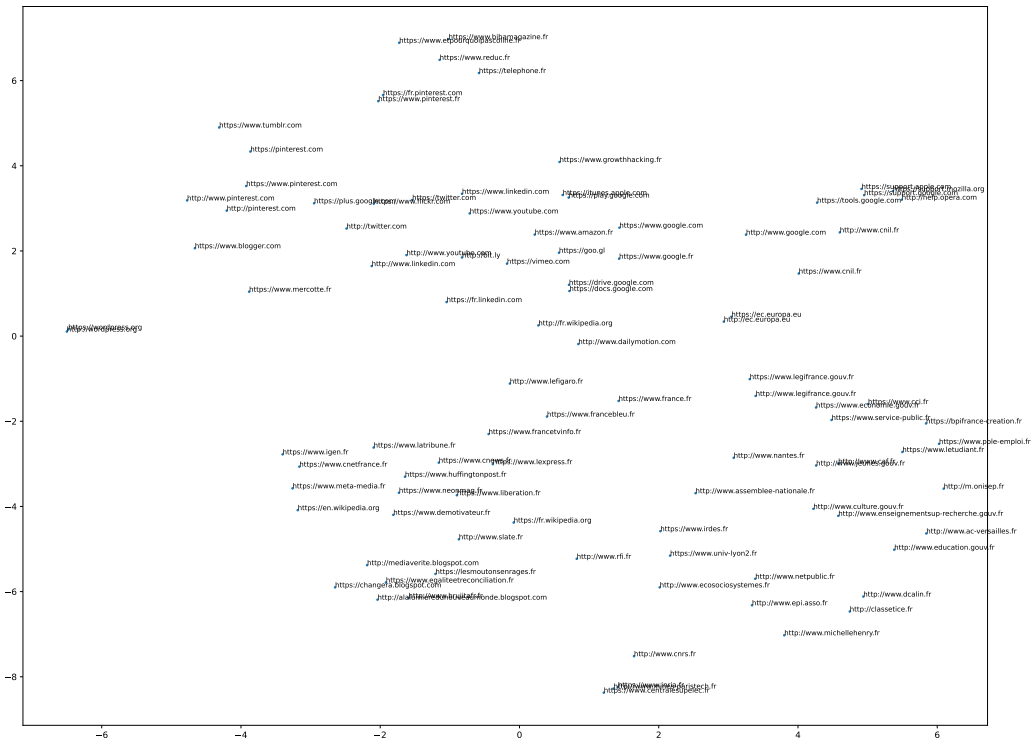
Martin JOLIF
martinjolif@gmail.com

Lab session # 5
ALTEGRAD 2024

11/15/24

Visualization of the French Web



1 Question 1

The DeepWalk architecture that we have seen just before for undirected graph can easily generalized to:

- For a directed graph, instead of choosing randomly a neighbor of the current vertex, we just have to choose randomly a vertex, which is the end of an edge that begins from the current edge (it's the same as before, we just consider the direction of the edges in the case of a directed graph).
- For a weighted graph, we can assign each weight to a probability ($p_{edgeij} = \frac{w_{edgeij}}{\sum_j w_{edgeij}}$). Instead of using a uniform distribution to choose the next vertex, we can use this new probability distribution that start from node i.

2 Question 2

We can observe that,

$$X_2 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = X_1$$

Therefore, the two embedding results are related by a reflection across the x-axis. This transformation preserves the relative distances between nodes while giving a different but equally valid embedding of the

graph structure. This is not surprising since DeepWalk embeddings are not unique. Any rotation, reflection, or translation of a valid embedding is also valid as long as it preserves the relative positions of the nodes that capture the graph's structure.

3 Question 3

After a message passing layer, the new feature vector of each node is the sum of the feature vectors of their neighbors. In task 10, our architecture is composed of 2 message passing layers. Therefore, the "receptive field" of each node in our GCN architecture is equal to 2.
In the same way, for a GCN architecture with k message passing layers, the "receptive field" of each node in our GCN architecture is equal to k. Therefore, the maximal number of edges separating a given node i from the nodes the features of which are taken into account in the prediction $\hat{Y}_i$ is k.

4 Question 4

- For the K_4 graph, as the graph is complete, we have:

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \text{ so, } \tilde{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \text{ and, } \tilde{D} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

  Therefore,

$$\hat{A} = \frac{1}{4}\tilde{A}$$

  and we assume that $X = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

  So, first,

$$\hat{A}XW^0 = \begin{pmatrix} -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \end{pmatrix} \text{ then, } Z^0 = f(\hat{A}XW^0) = \begin{pmatrix} 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \end{pmatrix}$$

  In the same way, we obtain,

$$\mathbf{Z^1} = \begin{pmatrix} 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \end{pmatrix}$$

- For the S_4 graph, we have (assuming the central node is the first one in the ordering):

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \text{ so, } \tilde{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \text{ and, } \tilde{D} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

  Therefore,

$$\hat{A} = \frac{1}{2}\begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & 1 \end{pmatrix}$$

  Then,

$$Z^0 = \begin{pmatrix} 0 & \frac{1+3\sqrt{2}}{8} \\ 0 & \frac{2+\sqrt{2}}{8} \\ 0 & \frac{2+\sqrt{2}}{8} \\ 0 & \frac{2+\sqrt{2}}{8} \end{pmatrix} \text{ and, } \mathbf{Z^1} = \begin{pmatrix} 0 & \frac{21+27\sqrt{2}}{160} & \frac{7+9\sqrt{2}}{64} \\ 0 & \frac{30+9\sqrt{2}}{160} & \frac{10+3\sqrt{2}}{64} \\ 0 & \frac{30+9\sqrt{2}}{160} & \frac{10+3\sqrt{2}}{64} \\ 0 & \frac{30+9\sqrt{2}}{160} & \frac{10+3\sqrt{2}}{64} \end{pmatrix} \approx \begin{pmatrix} 0 & 0.37 & 0.31 \\ 0 & 0.27 & 0.22 \\ 0 & 0.27 & 0.22 \\ 0 & 0.27 & 0.22 \end{pmatrix}$$

Finally, we observe that for both graphs, the first column of the matrix Z_1 is equal to zero (this is due to the ReLU activation function). Moreover, in the case of the K_4 graph, each row of the matrix Z_1 are the same. For the S_4 all the rows are the same except the first one, which corresponds to the central node.

If we had randomly sampled node features X from a random uniform distribution, then our rows of the final Z_1 matrix will be different with a very high probability.

Node classification on karate dataset

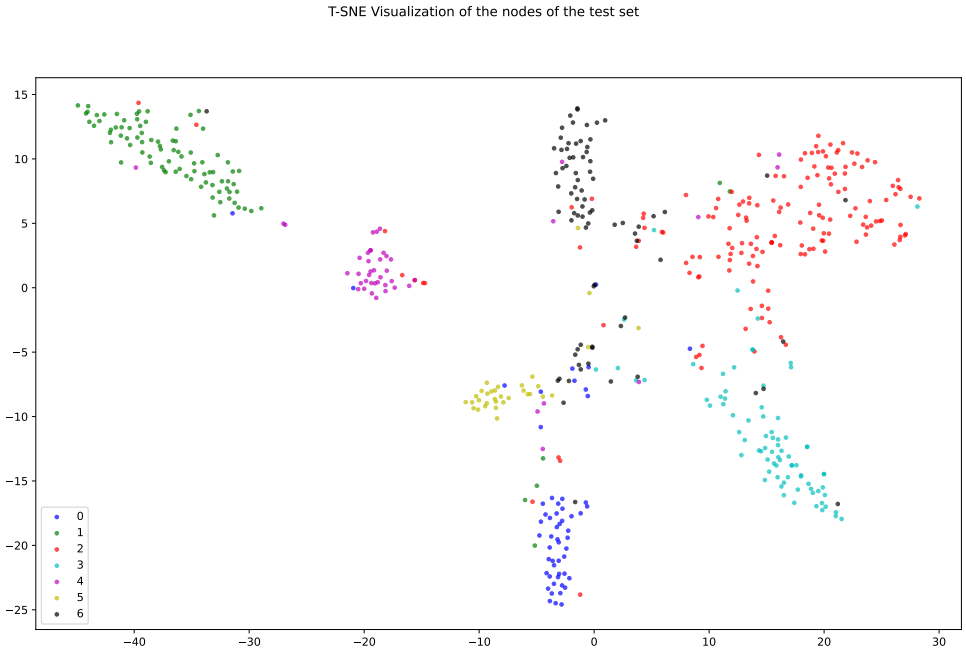
Logistic regression after a DeepWalk embedding gives an accuracy of 100%.

Logistic regression after a Spectral embedding gives an accuracy of 43%.

GNN with a identity matrix features gives an accuracy of 100%.

GNN with a ones vector features gives an accuracy of 28%.

Visualization of Node Representations for the Cora dataset



We obtain an accuracy of almost 86% which is quite good.