Martin BRICENO
jbriceno@telecom-paris.fr

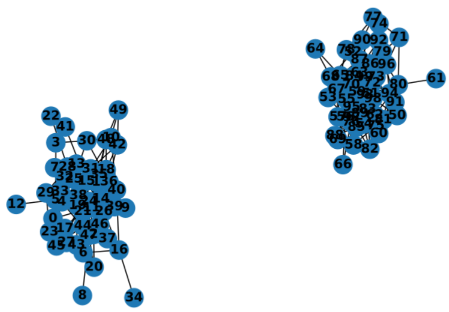
Lab session # 5
ALTEGRAD 2020

01/19/21

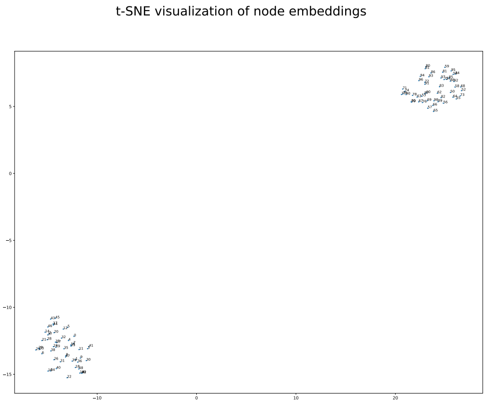
# 1 Question 1

To know how the embeddings nodes of the two connected components will be represented, we need to know that deepWalk captures the network structure in terms of neighborhood and community awareness [1]. This means that the distance between the latent dimensions must represent a metric to evaluate the social similarity between the corresponding network members. This way there will be a correspondence between the community structure in the input graphic and the final representation of the embeddings nodes.

Finally, since we have 2 separate components, the embeddings nodes will show **2 separate communities, each with 50 nodes, where the distances between the nodes of each community will be less than the distances between nodes of different communities** to represent the fact that nodes of different components have low similarity and those that are neighbors within a component have high similarity as we see in the figure below.



(a) Graph G with 2 connected components




(b) Embedding Space.

# 2 Question 2

Although the structure of both nodes is the same, deepwalk takes into account the information of the community to which it belongs (neighborhood) and we see that they do not share the same neighborhood for a random walk of short length. Moreover we can see that they belong to two different communities, **deepwalk will try to reflect this in the embedding space by placing them as far as possible not close to each other.**

On the other hand imagine two separate components of a network and two nodes in each component with the same structure. Even if the structure is the same it will not matter for DeepWalk since it will place them as far as possible by belonging to separate components to maximize the likelihood.

# 3 Question 3

Let's have the following matrix, where A is the adjacency matrix of G.

$$X = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^\top \; W^0 = \begin{bmatrix}0.8 & -0.5\end{bmatrix} \; W^1 = \begin{bmatrix}0.9 & -1.2 & 0.4 \\ 1.4 & -0.3 & -0.2\end{bmatrix} \; A = \begin{bmatrix}1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1\end{bmatrix}$$

- Now, we will normalize A:

$$\tilde{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2} \tag{1}$$

Where :

$$\tilde{A} = A + I = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{bmatrix} \ \tilde{D} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \tag{2}$$

then:

$$\tilde{A} = \begin{bmatrix} 0.5 & 0.25 & 0 & 0.25 \\ 0.25 & 0.5 & 0.25 & 0 \\ 0 & 0.25 & 0.5 & 0.25 \\ 0.25 & 0 & 0.25 & 0.5 \end{bmatrix} \tag{3}$$

- We will now calculate $Z^0$:

$$Z^0 = f(\tilde{A}XW^0),\ f = ReLu\ function \tag{4}$$

$$\tilde{A}XW^0 = \begin{bmatrix} 0.8 & -0.5 \\ 0.8 & -0.5 \\ 0.8 & -0.5 \\ 0.8 & -0.5 \end{bmatrix} \tag{5}$$

then:

$$Z^0 = \begin{bmatrix} 0.8 & 0 \\ 0.8 & 0 \\ 0.8 & 0 \\ 0.8 & 0 \end{bmatrix} \tag{6}$$

- We will now calculate $Z^1$:

$$Z^0 = f(\tilde{A}Z^0W^1),\ f = ReLu\ function \tag{7}$$

$$\tilde{A}Z^0W^1 = \begin{bmatrix} 0.72 & -0.96 & 0.32 \\ 0.72 & -0.96 & 0.32 \\ 0.72 & -0.96 & 0.32 \\ 0.72 & -0.96 & 0.32 \end{bmatrix} \tag{8}$$

then:

$$Z^1 = \begin{bmatrix} 0.72 & 0 & 0.32 \\ 0.72 & 0 & 0.32 \\ 0.72 & 0 & 0.32 \\ 0.72 & 0 & 0.32 \end{bmatrix} \tag{9}$$

**We see that the four rows are equal, this is normal given that being a cyclic graph, there is no difference between them therefore their embedding vectors must be equal**. Also because of the ReLu function a coordinate is null.

# 4 Question 4

**The accuracy obtained with the identity matrix was approximately 100%, while with the vector of ones it was approximately 28.57%**. Therefore we see that the result with the latter is much less efficient. **This is because in the first case the features of each node were different independent vectors, while in the second all are equal to the same numerical value.**

**This change removes flexibility from the model since it starts from the idea that each node is similar given its features.** In addition, this change in the initialization of the features affects the learning of the model, which prevents it from classifying correctly. Therefore the learning of the model depends on the initialization of the features of the nodes.

## References

[1] Rami Al-Rfou Bryan Perozzi and Steven Skiena. Deepwalk: Online learning of social representations. In *In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 701–710, 2014.