Romy BEAUTÉ
romybeaute.univ@gmail.com

PART I : Node Embeddings

In this part of the lab, we implemented the DeepWalk algorithm (recently-proposed algorithm for generating node embeddings [1]), and use it to visualize some data and to perform node classification.

1 Question 1 :

Let G be a graph that consists of M connected components. Each connected component is a complete graph $K2$ (i.e., two nodes connected by an edge). Suppose you use the DeepWalk algorithm to embed the $2M$ nodes in some vector space. How do you expect the cosine similarities of the embeddings of the nodes within connected components and the embeddings of the nodes in different connected components to compare? A qualitative answer is sufficient here.

DeepWalk, is a novel approach for learning latent social representations of vertices [1]. Specifically, the distance between latent dimensions represents a metric for evaluating social similarity between the corresponding members of the network (homophily). Therefore, within the same component, the cosine similarity of the embedding of the nodes should be as close to 1 as possible, as the nodes will be considered as belonging to the same community. At the contrary, nodes belonging to different connected components will be treated as belonging to different communities, and the cosine components is therefore expected to be as small as possible.

2 Question 2 :

The i-th row of each matrix stores the embedding of node vi. How are the two embedding results related to each other? Is one of them more informative than the other?*Let G be the graph shown in Figure 1. Suppose you use the DeepWalk algorithm twice to embed the nodes of G in the 2-dimensional space, and you obtain the following two embedding matrices $X1$ and $X2$:*

$$X1 = \begin{bmatrix} -1.0 & 1.0 \\ -1.0 & 1.0 \\ -0.5 & 0.5 \\ 0.5 & -0.5 \\ 1.0 & -1.0 \\ 1.0 & -1.0 \end{bmatrix}$$

$$X2 = \begin{bmatrix} -1.0 & -1.0 \\ -1.0 & -1.0 \\ -0.5 & -0.5 \\ 0.5 & 0.5 \\ 1.0 & 1.0 \\ 1.0 & 1.0 \end{bmatrix}$$

The i-th row of each matrix stores the embedding of node vi. How are the two embedding results related to each other? Is one of them more informative than the other?

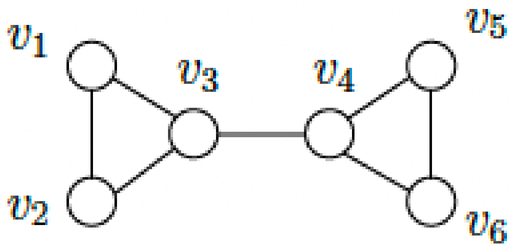

Figure 1: A synthetic graph G.

The two embeddings differ by they rotation : they have the same values, but are have different sign values. Consequently, both embeddings provide the same amount of information.

3 Question 3 :

The model you implemented consists of two message passing layers. Why do graph neural networks typically contain more than one message passing layer, and why don't they contain a very large number of such layers (e.g., larger than the diameter of the input graph)?

The reason why graph neural networks typically contain more than one message passing layer, is that it allows each node to have access to a broader neighborhood. This will result in a more meaningful final embedding layer and therefore a more qualitative representation of the graph's structure.
However, graph neural networks don't contain a very large number of such layers to limit introducing non-similar nodes in other node's neighborhood (too much information kills the information sort of). By introducing too many non-similar nodes in the neighborhood, the risk is to "blur" the information by creating very similar neighborhood, that won't be good information for classification anymore.

4 Question 4 :

$W^1 = \begin{bmatrix} 0.3 & 0.4 & 0.8 & 0.5 \\ 1.1 & 0.6 & 0.1 & 0.7 \end{bmatrix}$
Assume that there are no biases and that f is the $ReLU$ function. Compute (showing your calculations) matrix Z^1. The rows of this matrix correspond to the representations of the 4 nodes. What do you observe?*Let G be the P4 path, i.e., a tree with two nodes of degree 1 and the other two nodes of degree 2. Let also X be the 4 × 1 all-ones matrix, while W^0 and W^1 are defined as follows:*
$W^0 = \begin{bmatrix} 0.5 & -0.2 \end{bmatrix}$

$W^1 = \begin{bmatrix} 0.3 & 0.4 & 0.8 & 0.5 \\ 1.1 & 0.6 & 0.1 & 0.7 \end{bmatrix}$
Assume that there are no biases and that f is the $ReLU$ function. Compute (showing your calculations) matrix Z^1. The rows of this matrix correspond to the representations of the 4 nodes. What do you observe?

With assume G to be the $P4$ path, ie represented by a tree with 2 nodes of degree 1 and two other nodes of degree 2. We recall that the first-order proximity captures the direct neighboring relationships between vertices. The second-order proximity captures the 2-step relations between two vertices v and u (determined by the number of common neighbors shared by the two vertices). Let's first define the adjacency matrix A of G :

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

where the rows of the matrix Z contain the attributes of the nodes W^0 and W^1 as the trainable weight

matrices.
Given the adjacency matrix A of a graph, we can compute $A\Gamma as$: A = $D^(-1/2) \times (A+I) \times D^(-1/2)$ with D a diagonal matrix such that $D(ii)$

The output of the model will be given by : $Z = softmax(\hat{A}ReLU(AXW0)W1)$

References

[1] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 701–710, New York, NY, USA, 2014. Association for Computing Machinery.