Virgile Rennard
virgile@rennard.org

Lab session # 5
ALTEGRAD 2019

12/19/19

1 Question 1

It suffices to change the uniformity of the randomly chosen following node to show that we do not want to pick a previous node. Instead of choosing a vertex at random in the set of connected vertexes, we do it on the set of connected vertexes that weren't the one chosen on the previous iteration.

2 Question 2

In order to do Graph classification instead of the node classification, we take the idea from [1]. In it is described an end to end Deep Learning architecture for graph classification.
It works in three steps;

- Graph convolution layers that extract vertices local substructure features and define a consistent vertex ordering
- A sort pooling layer, that sorts the features of vertex under the predefined order and unifies input sizes
- Finally, an ensemble of convolutional and dense layers read the sorted graph representations in order to make predictions.

If we have a graph G and its node information matrix $A \in \mathbb{R}^{n \times c}$, the convolution layer is :

$$Z = F(\widetilde{D}^{-1}\widetilde{A}XW)$$

This convolution aggregates node information in the local neighborhoods so that it can extract local substrucutres information. To do so, is stacked different graph convolution layers :

$$Z^{t+1} = F(\widetilde{D}^{-1}\widetilde{A}X^tW^t)$$

with $Z^0 = X$, and the subsequents Z are the output of the t^{th} graph convolution layer.
The following Sortooling layer has an input as a $n \times \sum_{t=1}^{h} c_t$ tensor where each row is a vertex's feature descriptor and is columns are feature channels. Its ouput is a tensor with the vertex features sorted in a consistent order. It also unifies the sizes of the output tensors.
The sorted tensor is then passed through a sequence of CNN layers and dense layers capped by a softmax layer to give the classification.

3 Question 3

The GNN ouperformed the simple DeepWalk + logistic regression in he node classification task. The GNN is a more complex architecture that can better grasp the needed concepts for the classification.

4 Question 4

In the case where we change the initialization as a full matrix of 1 instead of the identity, the output is worse. As the weight in the first layer will be equally distributed for each entry, compared to the specific weighting of the Identity matrix.

References

[1] Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. An end-to-end deep learning architecture for graph classification. In *AAAI*, 2018.

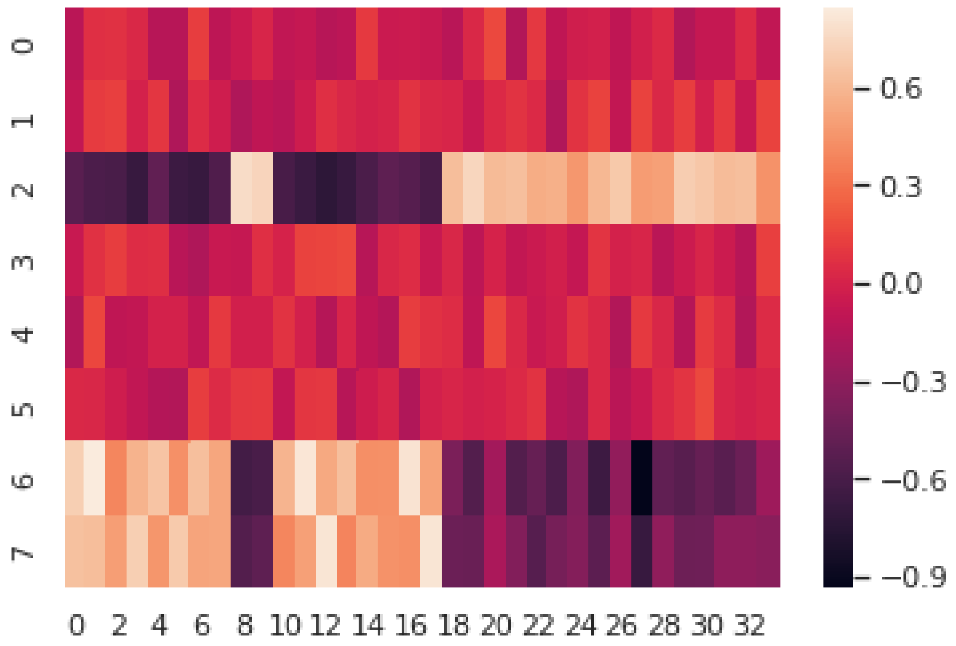

Figure 1: Heatmap when I_n

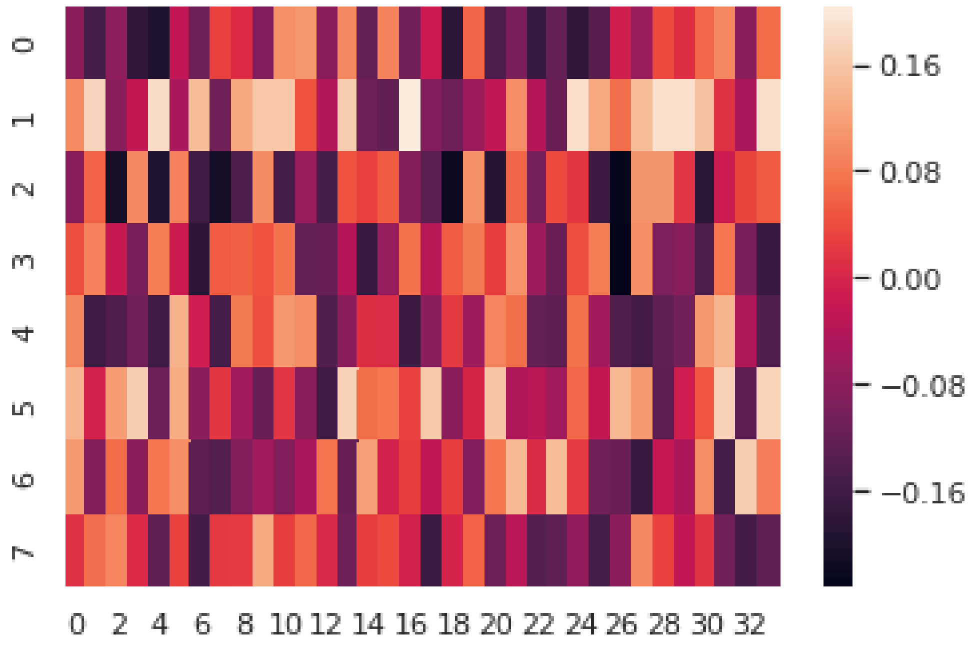

Figure 2: Heatmap when I_n