Megi Dervishi
megi.dervishi@ens.fr

Lab session # 7
ALTEGRAD 2021

02/09/22

Question 1

We see that for the problem at hand the model can simply learn $\rho = \mathrm{Id}$ and ϕ to be a permutation. Since then

$$\rho\left(\sum_{x \in X} \phi(x)\right) = \sum_{x \in X} x \tag{1}$$

which is the desired output. Hence we expect that the biases are zero and the weights can be any permutation matrix.

Question 2

The difference between DeepSets and GNN seen in Lab 6 is that the first MLP layers of DeepSets is a message passing layer in the GNN architecture, and the readout function in the GNN architecture doesn't necessarily have to be a sum. However if we consider a graph with no edges i.e. only isolated nodes then $A = 0$ so $\tilde{A} = \mathrm{Id}$ so a message passing layer becomes completely equivalent to a fully connected layer. The only difference then comes from how we choose to do the node embeddings in the GNN compared to how we choose to do the item embeddings in DeepSets.

Question 3

Using an embedding layer is simply more efficient, since to use a fully connected layer we would first have to transform the input words into a one-hot encoding and then pass it through the layer which is extremely wasteful especially since if the vocabulary is large the one-hot encoding will be particularly costly.

Question 4

Using DeepSets is probably a bad idea because by doing so we loose a crucial information of the problem/dataset which are the transitions from one state to the next. For example, one might think that along a 'listening session' a user could start listening to electronic music and transition to pop music. Then the SR-GNN would be able to understand that currently the user is most likely looking for more pop music. However DeepSets will have no idea of this 'evolution' and would simply recommend pop music and electronic music equally likely, which is of course sub-optimal.