Denis Duval
denis.duval13@gmail.com

Lab session # 6
ALTEGRAD 2022

12/13/22

1 Question 1

We are going to prove that $z_1^{(2)} = z_4^{(2)}$. The idea is that even though v_1 and v_4 don't have the same number of neighbours, they will receive the same information when we will normalize by the attention coefficients. Let's denote $z_i' = W^{(2)} z_i^{(1)}$ for all i. We have :

$$z_1^{(2)} = \alpha_{12} z_2' + \alpha_{13} z_3'$$

and

$$z_4^{(2)} = (\alpha_{42} + \alpha_{46}) z_2' + (\alpha_{43} + \alpha_{45}) z_3'$$

Moreover :

$$(\alpha_{42} + \alpha_{46}) = \frac{\exp(f(a[z_4'||z_2'])) + \exp(f(a[z_4'||z_6']))}{\exp(f(a[z_4'||z_2'])) + \exp(f(a[z_4'||z_6'])) + \exp(f(a[z_4'||z_3'])) + \exp(f(a[z_4'||z_5']))}$$

and using the equations at time (1):

$$(\alpha_{42} + \alpha_{46}) = \alpha_{12}$$

We can do the same for the second coefficient and obtain $(\alpha_{43} + \alpha_{45}) = \alpha_{13}$.So $z_1^{(2)} = z_4^{(2)}$.

2 Question 2

If we replace the features with constant ones then the model will behave like a dummy one and hence the accuracy will be bad. In fact, in a similar manner to the question 1 we can show that if all the features are equal then the message passing layer will again output equal representations for all the nodes (it comes from the fact that all the attention coefficients sum up to 1 for every node i). There is an exception if there are some nodes with degree 0. In that case, the resulting representation is just the zero vector. Hence we have a model that has to classify nodes that all have the same representation which is as good as having a dummy model (in the case of degree 0 nodes the model won't be more performing if the classification we want to achieve isn't closely related to this property of having some neighbour or not).

3 Question 3

(i) Sum :

$$z_{G_1} = [2.9 \quad 2.3 \quad 1.9]$$
$$z_{G_2} = [3.4 \quad 1.9 \quad 4.3]$$
$$z_{G_3} = [1.8 \quad 1.2 \quad 1.6]$$

(ii) Mean :

$$z_{G_1} = [0.96 \quad 0.76 \quad 0.63]$$
$$z_{G_2} = [0.85 \quad 0.475 \quad 1.075]$$
$$z_{G_3} = [0.6 \quad 0.4 \quad 0.53]$$

(iii) Max :

$$z_{G_1} = [2.2 \quad 1.8 \quad 1.5]$$
$$z_{G_2} = [2.2 \quad 1.8 \quad 1.5]$$
$$z_{G_3} = [2.2 \quad 1.8 \quad 1.5]$$

First we see that the max function isn't able to distinguish these graphs because their representations are all equal. For the sum and the mean the representations are also quite similar. Maybe the sum could be more discriminative here as for example the second component of z_{G_2} and the second component of z_{G_3} become (relatively) much closer than they were without the mean.

4 Question 4

$z_{G_1} = z_{G_2}$. In fact the adjacency matrices of the two graphs will have each row summing up to 2 (because of their structure). Then the matrices $\tilde{A}$ will have each row summing up to 3. This means that $\tilde{A}X$ is a vector full of 3 because X is full of ones and the sum on each of $\tilde{A}$ is 3. Hence the output after the first message passing layer is the same for the two graphs. Furthermore, the readout function is nothing but the scalar product by the vector full of ones. Using the symetricity of $\tilde{A}$ we can see that the output of the readout only depends on quantities that we have shown being equal for the two graphs.