# 1 Question 1

On one hand, we have,

$$\begin{aligned} z_1^{(2)} &= \sum_{j \in \mathcal{N}_1} \alpha_{1j}^{(2)} W^{(2)} z_j^{(1)} \\ &= \alpha_{12}^{(2)} W^{(2)} z_2^{(1)} + \alpha_{13}^{(2)} W^{(2)} z_3^{(1)} \end{aligned}$$

On the other hand, following the assumptions after the first message passing layer,

$$\begin{aligned} z_4^{(2)} &= \sum_{j \in \mathcal{N}_4} \alpha_{4j}^{(2)} W^{(2)} z_j^{(1)} \\ &= \alpha_{42}^{(2)} W^{(2)} z_2^{(1)} + \alpha_{43}^{(2)} W^{(2)} z_3^{(1)} + \alpha_{45}^{(2)} W^{(2)} z_5^{(1)} + \alpha_{46}^{(2)} W^{(2)} z_6^{(1)} \\ &= \alpha_{42}^{(2)} W^{(2)} z_2^{(1)} + \alpha_{43}^{(2)} W^{(2)} z_3^{(1)} + \alpha_{45}^{(2)} W^{(2)} z_3^{(1)} + \alpha_{46}^{(2)} W^{(2)} z_2^{(1)} \\ &= (\alpha_{42}^{(2)} + \alpha_{46}^{(2)}) W^{(2)} z_2^{(1)} + (\alpha_{43}^{(2)} + \alpha_{45}^{(2)}) W^{(2)} z_3^{(1)} \end{aligned}$$

Moreover, we observe, thanks to the same assumption that:

$$\begin{aligned} \alpha_{42}^{(2)} = \frac{\alpha_{12}^{(2)}}{2} \text{ and } \alpha_{46}^{(2)} = \frac{\alpha_{12}^{(2)}}{2} \\ \alpha_{43}^{(2)} = \frac{\alpha_{13}^{(2)}}{2} \text{ and } \alpha_{45}^{(2)} = \frac{\alpha_{13}^{(2)}}{2} \end{aligned}$$

Finally, we find that $z_1^{(2)} = z_4^{(2)}$!

# 2 Question 2

If nodes are assigned to identical features, our GNN won't be able to differentiate nodes and classify them correctly, it will conduct to a less good accuracy result for sure. Indeed, if node features are the same, the node representations from the graph attention layer will be the same for every node too, as each node representation is a normalized weighted sum of the previous neighbors representation (from the graph attention layer). Therefore, as our GNN is composed of 2 graph attention layers, it will not help us to classify the nodes with a high classification accuracy as with random features.

# Karate node classification

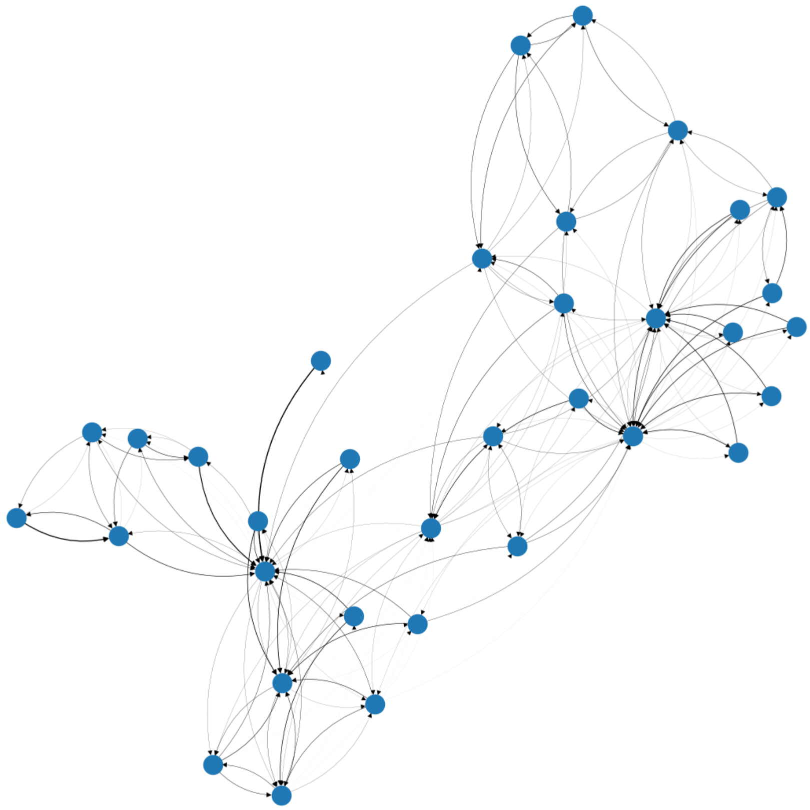

Figure 1: Visualization of the attention weights of the edges

# 3 Question 3

- For the SUM readout function here are the three graphs representations:

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

- For the MEAN readout function here are the three graphs representations:

$$z_{G_1} = [0.97, 0.77, 0.63]$$
$$z_{G_2} = [0.85, 0.47, 1.07]$$
$$z_{G_3} = [0.9, 0.6, 0.8]$$

- For the MAX readout function here are the three graphs representations:

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

For sure the MAX readout function is not a function that enable to distinguish the graph. Indeed, for the three graph $G_1, G_2, G_3$, through the MAX function, we get the same representation for each graph which seems to be not correct and will not help us to create clusters of graphs.

For the MEAN readout function it's better, the representation of each graph are different but very similar one from each other, which makes difficult to distinguish correctly graphs.

Finally, the best readout function is the SUM function. Indeed, the representation of each graph are different and much more separated. This function can therefore distinguish graphs better than the two others.

# 4 Question 4

For $G_1$, the adjadency matrix looks like that:

$$A_{G_1} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

For $G_2$, this is exactly the same pattern in a higher dimension:

$$A_{G_2} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Therefore, passing it through the two message passing layers, it will give exactly the same output except the fact that the dimension are different.

As we use the SUM readout function, it will sum over 8 lines for $G_2$ and only 4 lines for $G_1$. Therefore $z_{G_2} = 2z_{G_1}$
Therefore, the representation of both graph will be collinear, it's coherent with the fact that $G_1$ and $G_2$ can belong to a same clusters as they are cycles.