1 Question 1

$$z_1^{(2)} = \sum_{j \in \mathcal{N}(v_1)} \alpha_{ij} W^{(2)} z_j^{(1)},$$

$$z_4^{(2)} = \sum_{j \in \mathcal{N}(v_4)} \alpha_{ij} W^{(2)} z_j^{(1)} = \sum_{j \in \mathcal{N}(v_1)} \alpha_{ij} W^{(2)} z_j^{(1)} + \sum_{j \in \{v_5, v_6\}} \alpha_{ij} W^{(2)} z_j^{(1)}.$$

But:

$$\begin{cases} z_5^{(1)} = z_3^{(1)} \\ z_6^{(1)} = z_2^{(1)} \end{cases}$$

Let's look at α_{4j}:

$$\alpha_{4j} = \frac{\exp\left(\text{leakyReLU}\left[a^\top \left[W^{(2)} z_4^{(2)} || W^{(2)} z_j^{(2)}\right]\right]\right)}{\sum_{k \in \mathcal{N}(4)} \exp\left(\text{leakyReLU}\left[a^\top \left[W^{(2)} z_4^{(2)} || W^{(2)} z_k^{(2)}\right]\right]\right)}.$$

$$\alpha_{4j} = \frac{\exp\left(\text{leakyReLU}\left[a^\top \left[W^{(2)} z_1^{(2)} || W^{(2)} z_j^{(2)}\right]\right]\right)}{\sum_{k \in \mathcal{N}(4)} \exp\left(\text{leakyReLU}\left[a^\top \left[W^{(2)} z_1^{(2)} || W^{(2)} z_k^{(2)}\right]\right]\right)}.$$

$$\alpha_{4j} = \frac{1}{2} \alpha_{1j}$$

Thus,

$$z_4^{(2)} = \frac{1}{2} z_1^{(2)} + \frac{1}{2} z_1^{(2)} = z_1^{(2)}.$$

2 Task 3

The accuracy rapidly reaches 1.0, which mean that the GNN is able to clearly indentify two clusters of nodes.

3 Question 2

As we have seen in Q1, if all the features are equal at one stage, all the features will stay at the same encoding at the next step. Indeed, the initial encodings are equal, thus at the next step, the encoding will be that value multiply by a weight matrix shared among nodes. Since the number of neighbor intervene in the numerator with the sum over neighbors and at the denominator with the attention coefficient, the second layer will also have the same encoding for all nodes. Thus, the accuracy will be bad.

4 Task 4

The attention visualization looks like this :

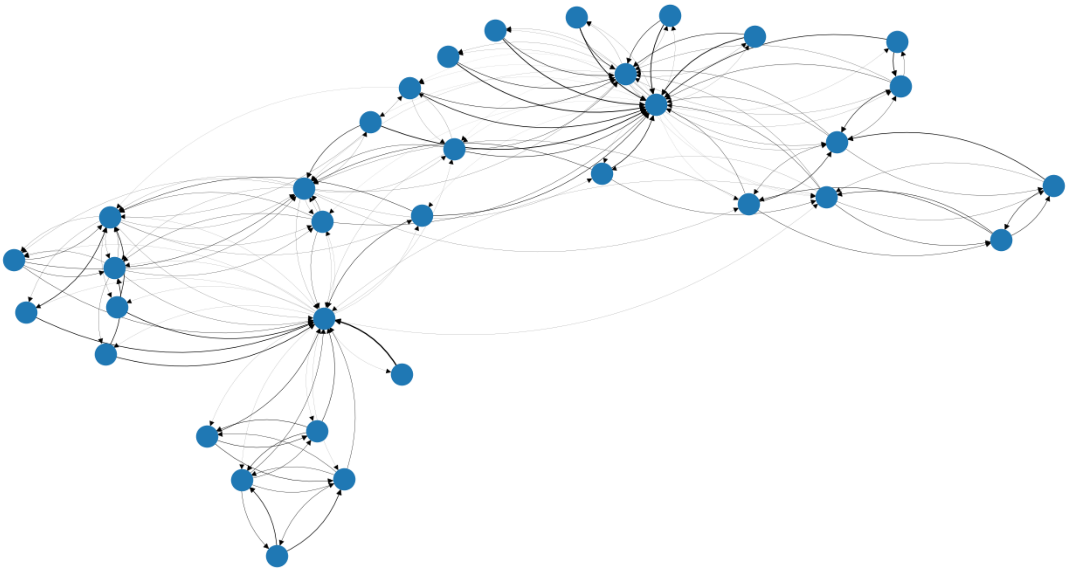

Figure 1: Attention visualization for the karate graph dataset.

5 Task 7

Here we also rapidly managed to get a 1.0 accuracy.

6 Question 3

(i)

$$Z_{G_1} = \begin{bmatrix} 2.8 & 1.9 & 1.9 \end{bmatrix}, \quad Z_{G_2} = \begin{bmatrix} 3.4 & 7.9 & 4.3 \end{bmatrix}, \quad Z_{G_3} = \begin{bmatrix} 1.8 & 3.0 & 2.9 \end{bmatrix}.$$

(ii)

$$Z_{G_1} = \frac{1}{3}\begin{bmatrix} 2.8 & 1.9 & 1.9 \end{bmatrix}, \quad Z_{G_2} = \frac{1}{4}\begin{bmatrix} 3.4 & 7.9 & 4.3 \end{bmatrix}, \quad Z_{G_3} = \frac{1}{2}\begin{bmatrix} 1.8 & 3.0 & 2.9 \end{bmatrix}.$$

(iii)

$$Z_{G_1} = \begin{bmatrix} 2.2 & 1.8 & 1.5 \end{bmatrix}, \quad Z_{G_2} = \begin{bmatrix} 2.2 & 1.8 & 1.5 \end{bmatrix}, \quad Z_{G_3} = \begin{bmatrix} 2.2 & 1.8 & 1.5 \end{bmatrix}.$$

The sum allows to distinguish the graphs the best, then it's the average, and finally, the max.

7 Question 4

On the node level, it is the same operation since local configuration is the same. However, since we use a sum in the readout and there is twice as much node in the second graph:

$$Z_{G_2} = 2 \times Z_{G_1}.$$

References