Julien Delavande
juliendelavande@gmail.com

Question 1

Given that:

$$z_1^{(1)} = z_4^{(1)}, \quad z_2^{(1)} = z_6^{(1)}, \quad z_3^{(1)} = z_5^{(1)},$$

and considering the neighborhoods:

$$N(v_1) = \{v_2, v_3\}, \quad N(v_4) = \{v_2, v_3, v_5, v_6\},$$

the update rule of the message passing layer implies:

$$z_1^{(2)} = [z_2^{(1)} || z_3^{(1)}],$$

$$z_4^{(2)} = [z_2^{(1)} || z_3^{(1)} || z_5^{(1)} || z_6^{(1)}].$$

Substituting the equalities:

$$z_5^{(1)} = z_3^{(1)}, \quad z_6^{(1)} = z_2^{(1)},$$

we find:

$$z_4^{(2)} = [z_2^{(1)} || z_3^{(1)} || z_3^{(1)} || z_2^{(1)}].$$

Since $z_4^{(2)}$ includes repeated contributions from $z_2^{(1)}$ and $z_3^{(1)}$, we conclude:

$$z_4^{(2)} = [z_1^{(2)} || z_1^{(2)}].$$

This results from the doubled contribution of each neighbor in $z_4^{(2)}$ compared to $z_1^{(2)}$.

Question 2

If all rows of matrix X are identical, the features of all nodes are the same. In this case:

- The message passing mechanism propagates identical information across the graph, and the node representations $z_i^{(t)}$ will depend solely on the structure of the graph (i.e., the adjacency matrix) and not on distinct node features.
- As the adjacency matrix encodes graph connectivity, the model may still differentiate nodes based on their neighborhoods.
- However, the absence of distinguishing node features can limit the model's ability to achieve high classification accuracy, especially if nodes with similar neighborhoods belong to different classes.

Conclusion: The model can achieve some level of classification accuracy based on graph structure alone, but identical node features reduce its expressive power, potentially leading to lower accuracy compared to using random or meaningful features.

Question 3

We compute the graph-level representations $z_{G_1}, z_{G_2}, z_{G_3}$ for each readout function:

- **Sum**

$$z_{G_1} = \sum_{i \in G_1} Z[i] = \begin{bmatrix} 2.9 & 2.3 & 1.9 \end{bmatrix}$$

$$z_{G_2} = \sum_{i \in G_2} Z[i] = \begin{bmatrix} 3.4 & 2.8 & 4.3 \end{bmatrix}$$

$$z_{G_3} = \sum_{i \in G_3} Z[i] = \begin{bmatrix} 1.8 & -0.6 & 1.6 \end{bmatrix}.$$

- **Mean**

$$z_{G_1} = \frac{1}{3} \sum_{i \in G_1} Z[i] = \begin{bmatrix} 0.967 & 0.767 & 0.633 \end{bmatrix}$$

$$z_{G_2} = \frac{1}{4} \sum_{i \in G_2} Z[i] = \begin{bmatrix} 0.85 & 0.7 & 1.075 \end{bmatrix}$$

$$z_{G_3} = \frac{1}{2} \sum_{i \in G_3} Z[i] = \begin{bmatrix} 0.9 & -0.3 & 0.8 \end{bmatrix}.$$

- **Max**

$$z_{G_1} = \max_{i \in G_1} Z[i] = \begin{bmatrix} 2.2 & 1.8 & 1.5 \end{bmatrix}$$

$$z_{G_2} = \max_{i \in G_2} Z[i] = \begin{bmatrix} 2.2 & 1.8 & 1.5 \end{bmatrix}$$

$$z_{G_3} = \max_{i \in G_3} Z[i] = \begin{bmatrix} 2.2 & -0.6 & 1.5 \end{bmatrix}.$$

Best Readout Function

The **sum** and **mean** functions provide distinct representations for G_1, G_2, and G_3. However, the **max** function results in identical representations for G_1 and G_2, making it less effective for distinguishing these graphs. Thus, the **sum** or **mean** functions are the best choices for this task.

Question 4

Given the two graphs $G_1 = C_4$ and $G_2 = C_8$, where C_n represents a cycle graph with n nodes, and assuming the node features are initialized to 1, the model computes graph-level representations z_{G_1} and z_{G_2} as follows:

- **Message Passing Layers:** Since all node features are initialized identically and the graphs are regular (same degree for all nodes), the propagation mechanism will result in identical node embeddings for all nodes in each graph. Thus, the node embeddings are indistinguishable within each graph.
- **Readout Phase**:
  - For the sum readout, $z_{G_1} = 4 \cdot h$ and $z_{G_2} = 8 \cdot h$, where h is the common node embedding.
  - For the mean readout, $z_{G_1} = z_{G_2} = h$, as the average of identical embeddings is the same for both graphs.
  - For the max readout, $z_{G_1} = z_{G_2} = h$, since the maximum embedding is the same in both cases.

Conclusion: The sum readout distinguishes G_1 and G_2, as it is sensitive to the number of nodes, while the mean and max readouts fail to distinguish the graphs due to the identical node embeddings.

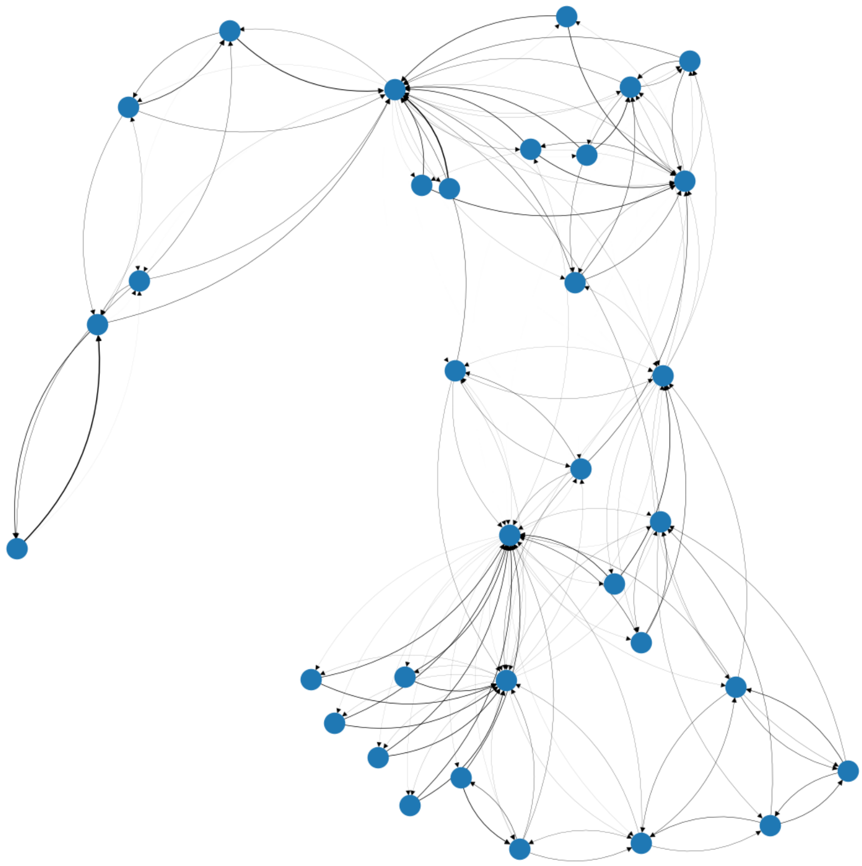

Figure 1: Task 4: Visualizing the attention weights on the karate graph.