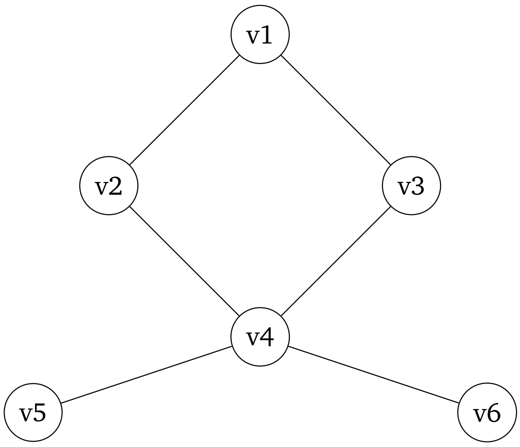
GRAVIER Thomas
thomas.gravier@gadz.org

1 Question 1

We are examining the following graph:



By design, $z_5^{(1)} = z_6^{(1)}$, since the hidden features of v_5 and v_6 were updated based on their shared neighbor v_4. Similarly, $z_2^{(1)} = z_3^{(1)}$, using the same reasoning: the hidden features of v_2 and v_3 are updated according to their two common neighbors, which share identical hidden features. Additionally, we have $z_2^{(1)} = z_6^{(1)}$. Therefore, we obtain that $z_2^{(1)} = z_3^{(1)} = z_5^{(1)} = z_6^{(1)}$.
In the second step, $z_4^{(2)}$ will be the weighted average of $[z_2^{(1)}, z_3^{(1)}, z_5^{(1)}, z_1^{(1)}]$, all of which are equal. Similarly, $z_1^{(2)} = z_4^{(2)}$.

2 Question 2

If the nodes are assigned identical features, the outputs from a graph attention layer will also be identical. Indeed, this layer computes a weighted average of a node's neighbor features, which will not change if the features are all the same. Therefore, the first two layers of the Graph Neural Network will be ineffective. In the end, the predicted labels will all be the same.

3 Visualization of Attention Scores

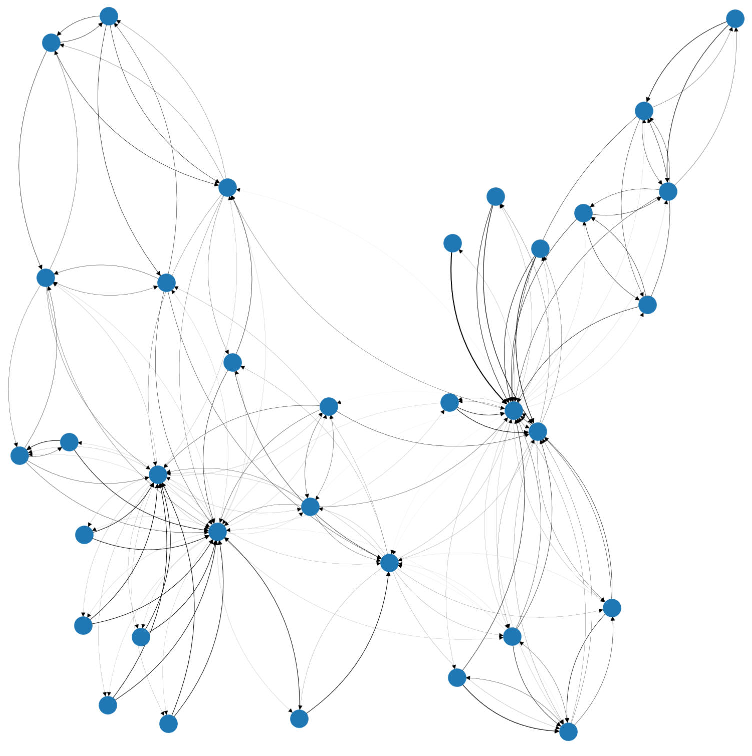


Figure 1: Weighted edges after training the graph attention layers.

4 Question 3

With the given Z matrix, we can compute the representation of the three graphs using different readout functions:

Readout function	z_{G1}	z_{G2}	z_{G3}
Sum	[2.9, 2.3, 1.9]	[3.4, 1.9, 4.3]	[1.8, 1.2, 1.6]
Mean	[0.97, 0.77, 0.63]	[0.85, 0.48, 1.08]	[0.9, 0.6, 0.8]
Max	[2.2, 1.8, 1.5]	[2.2, 1.8, 1.5]	[2.2, 1.8, 1.5]

Table 1: Hidden features with respect to the readout function.

We can observe that with the max readout function, the hidden features are identical for all graphs. As a result, the network won't be able to distinguish between the graphs. With the mean function, the values are all reduced, making it more difficult for the network to separate them. Finally, the sum function provides the best hidden features for each graph.

5 Question 4

Below are the two graphs on which we would like to apply the graph classification model:
They share the same adjacency matrix, where each node is connected to its two neighbors.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

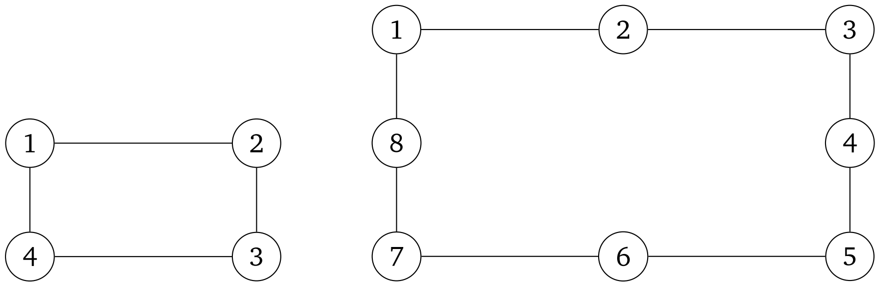


Figure 2: The C_4 and C_8 graphs, where C_n denotes a cycle consisting of n nodes.

Thus, the two graphs will produce the same output after two message-passing layers, i.e., $Z^{(2)}(G_1) = Z^{(2)}(G_2)$. The only difference will arise when applying the readout function. Specifically, the sum readout function will accumulate twice as many nodes in G_2. As a result, we obtain $z_{G_2} = 2z_{G_1}$.