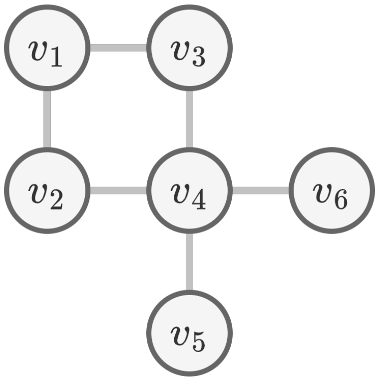
Matias Etcheverry
matias.etcheverry9@gmail.com

Lab session # 6
ALTEGRAD 2022

12/04/22

1 Question 1

We are studying the following graph:



By construction, $z_5^{(1)} = z_6^{(1)}$, as the hidden features of v_5 and v_6 was updated with respect to their same unique neighbor v_4. Moreover, $z_2^{(1)} = z_3^{(1)}$ using the same argument: the hidden features of v_2 and v_3 will be update using their 2 common neighbors which have the same hidden features. Moreover, we know that $z_2^{(1)} = z_6^{(1)}$. Thus, we end up with $z_2^{(1)} = z_3^{(1)} = z_5^{(1)} = z_6^{(1)}$.
Then, at 2$^{\text{nd}}$ step, $z_4^{(2)}$ will be a weighted average of $[z_2^{(1)}, z_3^{(1)}, z_5^{(1)}, z^{(1)}]$, which are all equal. Similarly, $z_1^{(2)}$ will be a weighted average of $[z_2^{(1)}, z_3^{(1)}]$, wich are all equal. Thus, $\boxed{z_1^{(2)} = z_4^{(2)}}$

2 Question 2

If the nodes are annotated with identical features, their outputs from a graph attention layer will be all the same. Indeed, this layer will make a weighted average of a node's neighbor features, which won't change if the features are all the same.
Thus, the 2 first layers of the Graph Neural Network will be useless. In the end, the predicted labels will be all the same.

3 Visualization of Attention Scores

fig. 1 shows the weighted edges after training the graph attention layers.

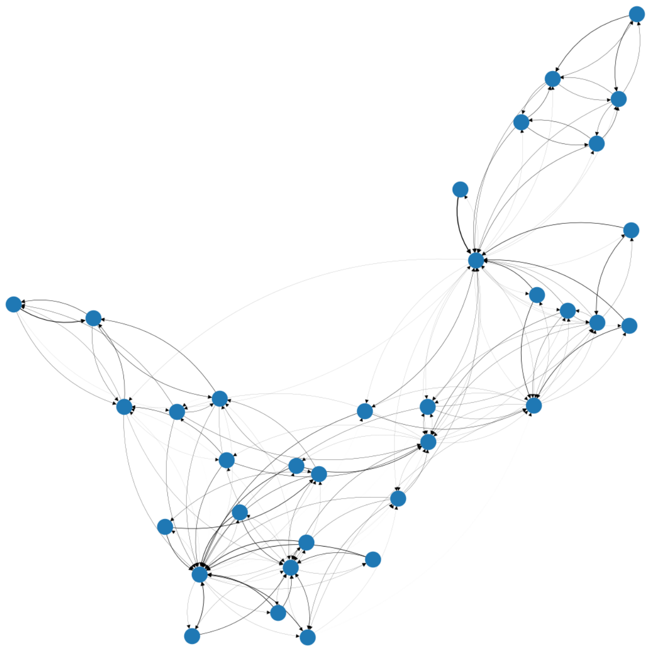


Figure 1: Attention scores in the Karate graph.

4 Question 3

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

Readout function	z_{G_1}	z_{G_2}	z_{G_3}
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 notice that with the max readout function, the hidden features are all the same. Thus our network won't be able to classify the graphs. With the mean function, the values are all shrunk, making it harder for the network to separate them. Finally, the sum function provides the best hidden features for each graph.

5 Question 4

Below, are the 2 graphs we would like apply the graph classification model on:

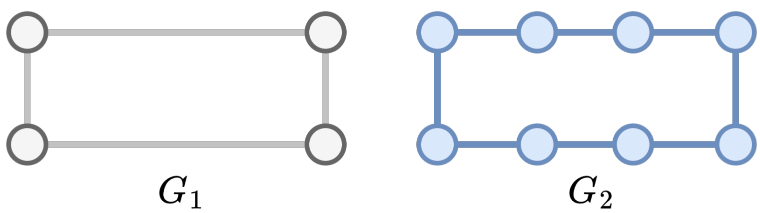


Figure 2: The C_4 and C_8 graphs where C_n denotes a cycle consisting of n nodes.

They share the same adjacency matrix, where each node is linked to its 2 neighbors.

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

Thus, the 2 graphs will have the same output in the 2 message passing layers ie $\mathbf{Z}^{(2)}(G_1) = \mathbf{Z}^{(2)}(G_2)$. The only difference will occur when passing the readout function. Indeed, a sum readout function will sum on twice more nodes in G_2. Thus, we end up with $z_{G_2} = 2z_{G_1}$.