1 Question 1

Here in this task we are trying to learn and re-construct the adjacency matrix of a graph with an auto-encoder. While it fits naturally to compare the learned representation $\hat{A} = \sigma(ZZ^T)$ to the normalized adjacency matrix $\bar{A}$ using an MSE function, the problem can also be treated a multi-class classification.

We can take the first layer of the encoder model: $H = ReLU(\bar{A}XW^0)$ (omitting bias for clarity), and having a decoder model that is different from a single dot product, so instead of having $\hat{A} = \sigma(ZZ^T)$, we would have:

$$\hat{A} = \sigma(\bar{A}HW^{(a)}) \tag{1}$$

Where $W^{(a)}$ is a different matrix from $W^{(1)}$ in the handout, such that here here $W^{(a)}$ is of size ($hidden\_dim\_1 * n$), where n is the number of nodes in the graph. We can then measure the auto-encoder accuracy by the cross-entropy loss instead of the MSE loss, as such:

$$\mathcal{L} = -\frac{1}{n^2}\sum_{i,j=1}^{n} \bar{A}_{ij} log \hat{A}_{ij} \tag{2}$$

2 Question 2

Currently, the given graph auto-encoder network is able to learn latent representations of the graph construction and node features through the Z matrices, with the following recurrent equation, for any given layer number l:

$$Z^{(l+1)} = ReLU(\bar{A}Z^{(l)}W^{(l)}) \tag{3}$$

In our current architecture, even if there could be L number of layers in the encoder part, there is only 1 dot product layer composed with a sigmoid to predict $\bar{A}$. What we can change to this architecture in order to predict the node features, is to have both a single dot product for the reconstruction of $\bar{A}$, and a second decoder part with L layers (the same number as for encoding), that mirrors the encoder structure (graph convolutions followed by ReLU activations), and we would calculate the loss for $\bar{X}$ as:

$$\mathcal{L}_{\mathcal{X}} = ||Z^{(0)} - Z^{(L)}||_F^2 \tag{4}$$

Since we are reconstructing both the adjacency matrix $\bar{A}$ and the node features $\bar{X}$ with the latent representation Z, we would have to optimise on a combined loss on our encoder-bidecoder network:

$$\mathcal{L} = \mathcal{L}_{\mathcal{A}} + \mathcal{L}_{\mathcal{X}} \tag{5}$$

3 Question 3

Let's compute the representation of the 3 graphs for 3 readout functions and

$$Z = \begin{bmatrix} 0.77 & -1.26 & -0.63 \\ 1.15 & -1.90 & -0.94 \\ 0.77 & -1.26 & -0.63 \\ 1.15 & -1.90 & -0.94 \\ 1.15 & -1.90 & -0.94 \\ 1.15 & -1.90 & -0.94 \\ 1.15 & -1.90 & -0.94 \\ 0.77 & -1.26 & -0.63 \\ 0.77 & -1.26 & -0.63 \end{bmatrix}; \tag{6}$$

Note that Z_{G_1} indices range from 0 to 2, indices for Z_{G_2} range from 3 to 6 and indices for Z_{G_3} range from 7 to 8.

- For the **sum** readout function, we have

$$Z_{G_1} = [2.69, -4.42, -2.2]; Z_{G_2} = [4.60, -7.6, -3.76]; Z_{G_3} = [1.54, -2.52, -1.26]; \tag{7}$$

- For the **mean** readout function, we have

$$Z_{G_1} = [0.896, -1.473, -0.733]; Z_{G_2} = [1.15, -1.90, -0.94]; Z_{G_3} = [0.77, -1.26, -0.63]; \tag{8}$$

- For the **max** readout function, we have

$$Z_{G_1} = [1.15, -1.26, -0.63]; Z_{G_2} = [1.15, -1.9, -0.94]; Z_{G_3} = [0.77, -1.26, -0.63]; \tag{9}$$

4 Question 4

If we observe the difference in representation of G_1 and G_2 in Z the previous question, we can see that they differ by one row, most likely because G_1 has 2 nodes with 1 edge and 1 node with 2 edges and G_2 has all nodes with 2 edges (regular graph). Moreover, we can also see that the nodes of G_3, each having only 1 edge, and not connected to G_1, have also the same Z values $[0.77, -1.26, -0.63]$ as the nodes of G_1 with only 1 edge.

Noticing that the 2-regular graphs have exactly the same number of nodes, it means that the regularized version $\hat{A}$ will have the same amplitude in values. Following this logic, for G_1 and G_2 of this question, every node has exactly 2 edges, and we can suppose that the Z matrix would have 6+6 vectors of same values, for each of the rows.

Since the readout function of the model is the sum() function, this will lead to having the same representation for Z_{G_1} and Z_{G_2}, which will be the sum of $Z[0:6,\ :]$, identical to the sum of $Z[6:12,\ :]$.