Question 1

$z_4^{(2)} = 2 * z_1^{(2)}$. Proof as follows:

$$(z_2^{(1)} = z_6^{(1)}) \;\&\; (z_3^{(1)} = z_5^{(1)}) \;\&\; (z_1^{(1)} = z_4^{(1)}) \;\rightarrow\; (\alpha_{12}^{(2)} = \alpha_{42}^{(2)} = \alpha_{46}^{(2)}) \;\&\; (\alpha_{13}^{(2)} = \alpha_{43}^{(2)} = \alpha_{45}^{(2)})$$

Moreover, we have:

$$z_1^{(2)} = \alpha_{12}^{(2)} W^{(2)} z_2^{(1)} + \alpha_{13}^{(2)} W^{(2)} z_3^{(1)}$$

Using the equalities between attention weights and the fact $(z_2^{(1)} = z_6^{(1)})$ & $(z_3^{(1)} = z_5^{(1)})$, we deduce that:

$$z_4^{(2)} = \alpha_{42}^{(2)} W^{(2)} z_2^{(1)} * 2 + \alpha_{43}^{(2)} W^{(2)} z_3^{(1)} * 2 = 2 * z_1^{(2)}$$

Question 2

Just as we discussed in the last lab, if nodes are annotated with identical features, the effect is the same as the case where there is only one feature. There would be a huge decrease of accuracy.

Question 3

- Option 1 of readout function: sum. In this case, the representations are:

$$z_{G1} = [2.9, 2.3, 1.9];\; z_{G2} = [3.4, 1.9, 4.3];\; z_{G3} = [1.8, 1.2, 1.6]$$

- Option 2 of readout function: mean

$$z_{G1} = [0.97, 0.77, 0.63];\; z_{G2} = [0.85, 0.475, 1.075];\; z_{G3} = [0.9, 0.6, 0.8]$$

- Option 3 of readout function: max

$$z_{G1} = [2.2, 1.8, 1.5];\; z_{G2} = [2.2, 1.8, 1.5];\; z_{G3} = [2.2, 1.8, 1.5]$$

In this case, we can use **sum** as readout function which distinguishes best these three graphs.

Question 4

Since the features of all nodes are initialized to 1 and all nodes are equivalent in graph structure (degree is 2), we can deduce that the representation of all nodes (in G_1 and G_2) are the same. Moreover, we use **sum** as readout function. Therefore, $z_{G2} = 2 * z_{G1}$ because there are two times of nodes in G_2 than in G_1.