1 Question 1

For Erdos-Renyi random graph $G(50, 0.1)$, the expected number of edges is equal to the total number of possible edges times the probability that he is included (since the inclusion of edges are independent from each other) therefore, it is equal to $50 \times 49 \times 0.1 = 245$ edge. For a DeepWalk algorithm with 10 walks of length 20 from every node, we can cover in total $10 \times 20 \times 100 = 20000$ node in total, therefore, we can cover 100 times the total of nodes in our graph. The expected distances of embedding should be small in the same connected component, and greater for two nodes from different connected components, which should generate two aggregations of nodes in the embedding space.

2 Question 2

Since the DeepWalk algorithm maximizes the probability of the the neighbours given the embedding of a vertex, and the nodes v1 and v2 are the furthest away from each other, despite their similar structure, they will not be close to each other in the embedding space.

3 Question 3

The adjacency matrix of G is: $A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$

We first compute the matrix $\tilde{A} = A + I_4 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$

The matrix $\tilde{D}$ is defined as follows: $\tilde{D} = diag(3, 3, 3, 3)$

And from it, we can define the normalized adjacency matrix

$$\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

We compute the first message passing layer:

$$\hat{A}XW^0 = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (0.8 \quad -0.5) = \begin{pmatrix} 0.8 & -0.5 \\ 0.8 & -0.5 \\ 0.8 & -0.5 \\ 0.8 & -0.5 \end{pmatrix}$$

By passing it through the first activation function reLU, we get:

$$Z^0 = reLU(\hat{A}XW^0) = \begin{pmatrix} 0.8 & 0 \\ 0.8 & 0 \\ 0.8 & 0 \\ 0.8 & 0 \end{pmatrix}$$

We compute the second message passing layer:

$$\hat{A}Z^0W^1 = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0.8 & 0 \\ 0.8 & 0 \\ 0.8 & 0 \\ 0.8 & 0 \end{pmatrix} \begin{pmatrix} 0.9 & -1.2 & 0.4 \\ 1.4 & -0.3 & -0.2 \end{pmatrix} = \begin{pmatrix} 0.72 & -0.96 & 0.32 \\ 0.72 & -0.96 & 0.32 \\ 0.72 & -0.96 & 0.32 \\ 0.72 & -0.96 & 0.32 \end{pmatrix}$$

and we pass it through a reLU activation function to get:

$$Z^1 = reLU(\hat{A}Z^0W^1) = \begin{pmatrix} 0.72 & 0 & 0.32 \\ 0.72 & 0 & 0.32 \\ 0.72 & 0 & 0.32 \\ 0.72 & 0 & 0.32 \end{pmatrix}$$

We observe that the 4 nodes have the same representation, and they will overlap in the embedding.

4 Question 4

In this case, we observe that the accuracy drop drastically, from 0.85 in the initial case to 0.25. From the previous example, we see that if the features are initialised with the same values, the GNN learns the same representation for all the nodes, and therefore they overlap, which explains the drop in accuracy.