1 Question 1

A random walk should happen only between 2 nodes in the same connected component. At each iteration, the next node to be chosen for the walk is a neighbor of the current node, which is the only other node of same connected component. Therefore, the resulting DeepWalk embeddings of the nodes within the same connected component should be very similar in terms of cosine similarity, whose value should be close to 1. In contrast, the cosine similarity of the embeddings of nodes in different connected components should be close to 0 as they should not share any similarity because each of the random walks only stay within a component.

2 Question 2

For DeepWalk, we generate n random walks of length l for each of the v vertices. Thus, the time complexity is $O(v \times n \times l)$ or $O(v)$ for small n and l and large v, which also corresponds to what was given in [2]. For spectral embedding, we first compute Laplacian matrix, with a complexity of $O(v^2)$ since we have to do a sparse matrix multiplication with a diagonal matrix. Then, computing the eigenvectors of the Laplacian gives a complexity of $O(v^3)$ according to [1], bringing a total complexity for the embedding of $O(v^3 + v^2) = O(v^3)$.

3 Question 3

If no self-loops were added to the graph, the hidden state of a node is only influenced by its neighbors without the node itself. Let's take the architecture of the GNN in the lab paper as an example. The first layer is given by $\mathbf{Z}^0 = f(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^0)$. Since $\hat{\mathbf{A}}$ has zero diagonal, the hidden state of a node is the aggregation of only the features of the node's neighbors. Furthermore, the second layer is given by $\mathbf{Z}^1 = f(\hat{\mathbf{A}}\mathbf{Z}^0\mathbf{W}^1)$ and once again $\hat{\mathbf{A}}$ has zero diagonal. Therefore, the hidden state of a node in this layer is the aggregation of the hidden states of the node's neighbors from the previous layer. The hidden state of the node itself from the previous layer is not included in the aggregation and thus some information is lost.

4 Question 4

$$\mathbf{W}^0 = \begin{bmatrix} 0.5 & -0.2 \end{bmatrix}, \ \mathbf{W}^1 = \begin{bmatrix} 0.3 & -0.4 & 0.8 & 0.5 \\ -1.1 & 0.6 & -0.1 & 0.7 \end{bmatrix}$$

4.1 Star graph S_4

Adjacency matrix:

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

Normalized adjacency matrix:

$$\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ and $\tilde{\mathbf{D}}$ is a diagonal matrix with $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$.

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \ \tilde{\mathbf{D}} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \Rightarrow \tilde{\mathbf{D}}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\hat{\mathbf{A}} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2\sqrt{2}} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

Feature matrix:

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

First message passing layer:

$$\mathbf{Z}^0 = \mathrm{ReLU}(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^0) = \mathrm{ReLU}\left(\begin{bmatrix} 0.6553 & -0.2621 \\ 0.4267 & -0.1707 \\ 0.4267 & -0.1707 \\ 0.4267 & -0.1707 \end{bmatrix}\right) = \begin{bmatrix} 0.6553 & 0 \\ 0.4267 & 0 \\ 0.4267 & 0 \\ 0.4267 & 0 \end{bmatrix}$$

Second message passing layer:

$$\mathbf{Z}^1 = \mathrm{ReLU}(\hat{\mathbf{A}}\mathbf{Z}^0\mathbf{W}^1) = \mathrm{ReLU}\left(\begin{bmatrix} 0.1849 & -0.2465 & 0.4931 & 0.3082 \\ 0.1335 & -0.1780 & 0.3560 & 0.2225 \\ 0.1335 & -0.1780 & 0.3560 & 0.2225 \\ 0.1335 & -0.1780 & 0.3560 & 0.2225 \end{bmatrix}\right) = \begin{bmatrix} 0.1849 & 0 & 0.4931 & 0.3082 \\ 0.1335 & 0 & 0.3560 & 0.2225 \\ 0.1335 & 0 & 0.3560 & 0.2225 \\ 0.1335 & 0 & 0.3560 & 0.2225 \end{bmatrix}$$

4.2 Circle graph C_4

Adjacency matrix:

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

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \; \tilde{\mathbf{D}} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \Rightarrow \tilde{\mathbf{D}}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$$

Normalized adjacency matrix:

$$\hat{\mathbf{A}} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

First message passing layer:

$$\mathbf{Z}^0 = \mathrm{ReLU}(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^0) = \mathrm{ReLU}\left(\begin{bmatrix} 0.5 & -0.2 \\ 0.5 & -0.2 \\ 0.5 & -0.2 \\ 0.5 & -0.2 \end{bmatrix}\right) = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \end{bmatrix}$$

Second message passing layer:

$$\mathbf{Z}^1 = \mathrm{ReLU}(\hat{\mathbf{A}}\mathbf{Z}^0\mathbf{W}^1) = \mathrm{ReLU}\left(\begin{bmatrix} 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \end{bmatrix}\right) = \begin{bmatrix} 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \end{bmatrix}$$

After the computation of $\mathbf{Z}^1$, for the case of C_4, we can see that all rows of the matrix have the same value. This means that all nodes have the same resulting features, and that is because of the same feature initialization of all nodes with the matrix $\mathbf{X}$ and the fact that all nodes have the same adjacency. For the case of S_4, the first row, which corresponds to the center node, has a different value due to its different adjacency.

References

[1] Mu Li, Xiao-Chen Lian, James T. Kwok, and Bao-Liang Lu. Time and space efficient spectral clustering via column sampling. In *CVPR 2011*, pages 2297–2304, 2011.

[2] Tiago Pimentel, Rafael Castro, Adriano Veloso, and Nivio Ziviani. Efficient estimation of node representations in large graphs using linear contexts. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2019.