1 Question 1

The cosine similarity between two vectors x and y are given for all non-nul x, y by:

$$cossim(x,y) = \frac{x^T y}{||x||||y||} \in [-1;1].$$

For this metric, 1 indicates identical vectors (w.r.t their orientation), 0 indicates no similarity (orthogonal vectors), and -1 indicates opposite vectors.
Let's denote two nodes by X_1 and X_2, and their associated embedding x_1 and y_2.
If X_1 and X_2 are in the same K_2 graph, then any random walk starting from one of these nodes will go through the other node. And then, for random walks of length larger than 2, both nodes will be reached one after the other. For exemple a random walk from X_1 will be $X_1 \to X_2 \to X_1 \to \cdots \to X_i$ where $i \in 1, 2$. Due to this, the context (in terms of random walks) for both nodes in a single component will be extremely similar, if not identical. It implies that the Skip-Gram model will learn almost identical contexts for these nodes, leading to similar embeddings. Thus, $x^1 \sim x^2$ and $cossim(x_1, x_2) \sim 1$.
Otherwise, X^1 and X^2 will not have random walk in common, there will be no co-occurrence in the walks. And since each K_2 component is disconnected from the others, nodes from different components will never appear together in the random walks. There will be no overlap in their contexts, so x^1 and x^2 will be very different and $cossim(x_1, x_2)$ is expected to be low (near 0 or below).

2 Question 2

Time complexity of DeepWalk technique is $O(|V|) = O(n)$ where $|V| = n$ is the size of the vocabulary, that is to say the total number of nodes. [1].
Time complexity of Spectral embedding takes into account the computation of the adjacency matrix $A \in \mathcal{M}_{n\times n}$ which is $O(n^2)$ because the amount of pairs is $n(n-1)/2$, the computation of D (so the degree sequence) and D^{-1} is $O(n)$, and the computation of L_{rw} which needs matrix product, hence $O(n^2)$. Finally, the decomposition of $L_{rw} \in \mathcal{M}_{n\times n}$ to get the eigenvalues and eigenvectors needs $O(n^3)$. Thus the time complexity is $O(n^3)$ which is worse than $O(n)$.

3 Question 3

4 Question 4

We assume

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

4.1 Star graph S4

Consider a star graph S4. The adjacency matrix A of the star graph S4 is augmented by the identity matrix I:

$$\tilde{A} = A + I = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

The degree sequence is given by:

$$\text{Degrees} = [4, 2, 2, 2]$$

The diagonal matrix $\tilde{D}^{-1/2}$ is computed by taking the inverse of the square root of the node degrees:

$$\tilde{D}^{-1/2} = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$$

The normalized adjacency matrix $\hat{A}$ is obtained by the product:

$$\hat{A} = \tilde{D}^{-1/2} \times \tilde{A} \times \tilde{D}^{-1/2} = \begin{pmatrix} \frac{1}{4} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2\sqrt{2}} & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

We have

$$\hat{A}XW^0 = \begin{pmatrix} \frac{1+3\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \end{pmatrix} \begin{pmatrix} 0.5 & -0.2 \end{pmatrix} = \begin{pmatrix} \frac{1+3\sqrt{2}}{8} & -\frac{1+3\sqrt{2}}{20} \\ \frac{2+\sqrt{2}}{8} & -\frac{2+\sqrt{2}}{20} \\ \frac{2+\sqrt{2}}{8} & -\frac{2+\sqrt{2}}{20} \\ \frac{2+\sqrt{2}}{8} & -\frac{2+\sqrt{2}}{20} \end{pmatrix}$$

And then

$$Z^0 = f(\hat{A}XW^0) = \begin{pmatrix} \frac{1+3\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \end{pmatrix}$$

We consider now

$$\hat{A}Z^0W^1 = \begin{pmatrix} \frac{7+9\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \end{pmatrix} \begin{pmatrix} 0.3 & -0.4 & 0.8 & 0.5 \\ -1.1 & 0.6 & -0.1 & 0.7 \end{pmatrix}$$

$$= \begin{pmatrix} \left(\frac{7+9\sqrt{2}}{32} \times 0.3\right) & \left(\frac{7+9\sqrt{2}}{32} \times -0.4\right) & \left(\frac{7+9\sqrt{2}}{32} \times 0.8\right) & \left(\frac{7+9\sqrt{2}}{32} \times 0.5\right) \\ \left(\frac{10+3\sqrt{2}}{32} \times 0.3\right) & \left(\frac{10+3\sqrt{2}}{32} \times -0.4\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.8\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.5\right) \\ \left(\frac{10+3\sqrt{2}}{32} \times 0.3\right) & \left(\frac{10+3\sqrt{2}}{32} \times -0.4\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.8\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.5\right) \\ \left(\frac{10+3\sqrt{2}}{32} \times 0.3\right) & \left(\frac{10+3\sqrt{2}}{32} \times -0.4\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.8\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.5\right) \end{pmatrix}$$

And then

$$Z^1 = f(\hat{A}Z0W^1) = \begin{pmatrix} \left(\frac{7+9\sqrt{2}}{32} \times 0.3\right) & 0 & \left(\frac{7+9\sqrt{2}}{32} \times 0.8\right) & \left(\frac{7+9\sqrt{2}}{32} \times 0.5\right) \\ \left(\frac{10+3\sqrt{2}}{32} \times 0.3\right) & 0 & \left(\frac{10+3\sqrt{2}}{32} \times 0.8\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.5\right) \\ \left(\frac{10+3\sqrt{2}}{32} \times 0.3\right) & 0 & \left(\frac{10+3\sqrt{2}}{32} \times 0.8\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.5\right) \\ \left(\frac{10+3\sqrt{2}}{32} \times 0.3\right) & 0 & \left(\frac{10+3\sqrt{2}}{32} \times 0.8\right) & \left(\frac{10+3\sqrt{2}}{32} \times 0.5\right) \end{pmatrix}$$

We observe that

- all the rows are equals except the first one. Knowing the uniformity in the features, this is due because the central node in the star graph (represented by the first row of Z^1) has a distinct structural role compared to the leaf nodes, but the three leaf nodes are structurally identical.
- there is a column filled by 0 (due to the weight matrices we chose).

4.2 Cycle graph C4

Consider a star graph C4. The adjacency matrix A of the star graph S4 is augmented by the identity matrix I

$$\tilde{A} = A + I = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

The degree sequence is given by

$$\text{Degrees} = [3, 3, 3, 3]$$

and then we have

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

The normalized adjacency matrix $\hat{A}$ is obtained by the product:

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

We have

$$\hat{A}XW^0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0.5 & -0.2 \end{pmatrix} = \begin{pmatrix} 0.5 & -0.2 \\ 0.5 & -0.2 \\ 0.5 & -0.2 \\ 0.5 & -0.2 \end{pmatrix}$$

And then

$$Z^0 = f(\hat{A}XW^0) = \begin{pmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \end{pmatrix}$$

We finally, we have

$$\begin{aligned} \hat{A}Z^0W^1 &= \frac{1}{6} \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \end{pmatrix} \begin{pmatrix} 0.3 & -0.4 & 0.8 & 0.5 \\ -1.1 & 0.6 & -0.1 & 0.7 \end{pmatrix} \\ &= \begin{pmatrix} 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \end{pmatrix} \end{aligned}$$

And then

$$Z^1 = f(\hat{A}Z0W^1) = \begin{pmatrix} 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \end{pmatrix}$$

We observe that

- All the rows are equals. Knowing the uniformity of the features, this is because all the nodes of the cycle are structurally equivalent. They are surrounded by the same topology and equivalent neighborhood.
- there is a column filled by 0 (due to the weight matrices we chose).

References

[1] 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.