1 Question 1

The nodes within the same component are connected by an edge and their embeddings are are likely to have a high cosine similarity.
Indeed, if we look at the formula $\underset{\phi}{\text{minimize}} - \log \prod_{\substack{j=i-w \\ j\neq i}}^{i+w} P\left(v_j \mid \phi(v_i)\right)$, for each i such that v_i is a vertex, and $\phi(v_i)$ its embedding, we have that the probability of having neighbours of v_i knowing $\phi(v_i)$ must be high, and so the embeddings of the neighbours of v_i are similar to the embedding of v_i.
In the same way, nodes in two different components are likely to have embeddings that have a low cosine similarity.

2 Question 2

Let's take a graph $G = (V, E)$ with N nodes. We are looking fot the complexity of the deepwalk algorithm.

1. **First step - Generate walks:** For each node, we generate a random walks of size t, so we have a complexity of $O(Nat)$
2. **Second step - Word2Vec:** We use continuous skipgram model. The complexity of one training step is $O(2w(n_{dim} + n_{dim} log_2(N))$ (detailed in [1]).
   With a number of epochs n_{epochs} and Na sequences (one random walk is treated as a sentence), we have a complexity of $O(n_{epochs} Na 2w(n_{dim} + n_{dim} log_2(N))$
3. **Final complexity of deepwalk:** We finally have a complexity of $O(Na(t+2wn_{epochs}(n_{dim}+n_{dim}log_2(N)))$, and if we keep only the variables linked to the graph, we have a complexity of $O(N + Nlog_2(N))$.

   The complexity of spectral embedding is $O(N^3)$ (comes from the research of eigenvectors and eigenvalues).

3 Question 3

Without self-loop, we won't use the node self-information in the first layer of the GNN. The evolution of hidden states is using the information of neighboring nodes.
So we may have same (or similar) hidden states for different nodes. If we use two layers, we'll reduce this phenomenon as we may be able to retrieve information of one node using hidden state of other nodes.
Finally, it remains better to use self-loops and that's why we add them to the graphs.

4 Question 4

Star graph C_4

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

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

and with $X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, W^0 = \begin{bmatrix} 0.5 & -0.2 \end{bmatrix}$, we have

$$\hat{A}XW^0 = \begin{bmatrix} \frac{3\sqrt(2)+1}{8} & -\frac{3\sqrt(2)+1}{20} \\ \frac{\sqrt(2)+2}{8} & -\frac{\sqrt(2)+2}{20} \\ \frac{\sqrt(2)+2}{8} & -\frac{\sqrt(2)+2}{20} \\ \frac{\sqrt(2)+2}{8} & -\frac{\sqrt(2)+2}{20} \end{bmatrix}$$

With ReLU activation function, we have $Z^0 = \begin{bmatrix} \frac{3\sqrt(2)+1}{8} & 0 \\ \frac{\sqrt(2)+2}{8} & 0 \\ \frac{\sqrt(2)+2}{8} & 0 \\ \frac{\sqrt(2)+2}{8} & 0 \end{bmatrix}$

Then, we have, after multiplying the three matrices with $W^1 = \begin{bmatrix} 0.3 & -0.4 & 0.8 & 0.5 \\ -1.1 & 0.6 & -0.1 & 0.7 \end{bmatrix}$,

$$\hat{A}Z^0W^1 = \begin{bmatrix} 0.185 & -0.247 & 0.493 & 0.308 \\ 0.134 & -0.178 & 0.356 & 0.223 \\ 0.134 & -0.178 & 0.356 & 0.223 \\ 0.134 & -0.178 & 0.356 & 0.223 \end{bmatrix}$$

And finally, after applying ReLU activation function, we have:

$$Z^1 = \begin{bmatrix} 0.185 & 0 & 0.493 & 0.308 \\ 0.134 & 0 & 0.356 & 0.223 \\ 0.134 & 0 & 0.356 & 0.223 \\ 0.134 & 0 & 0.356 & 0.223 \end{bmatrix}$$

Cycle graph C_4

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

$$\hat{A} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \approx \begin{bmatrix} 0.25 & 0.35 & 0.35 & 0.35 \\ 0.35 & 0.5 & 0 & 0 \\ 0.35 & 0 & 0.5 & 0 \\ 0.35 & 0 & 0 & 0.5 \end{bmatrix}$$

$$\hat{A}XW^0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0.5 & -0.2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{5} \\ \frac{1}{2} & -\frac{1}{5} \\ \frac{1}{2} & -\frac{1}{5} \\ \frac{1}{2} & -\frac{1}{5} \end{bmatrix}$$

And with ReLU activation function, we have:

$$Z^0 = \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \end{bmatrix}$$

And then,

$$\hat{A}Z^0W^1 = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \end{bmatrix} W^1 = \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \end{bmatrix} W^1 = \begin{bmatrix} 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \end{bmatrix}$$

And with ReLU activation function, we have

$$Z^1 = \begin{bmatrix} 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \end{bmatrix}$$

Observations

We notice that in the cycle all the nodes have the same features, which makes sense as input features were the

same and all the nodes are playing an identical role in the graph.
Same applies to star graph, where all the "external" roles have same role in the graph and their features are the same, whereas the central node has a different feature. We can then separate central node and external node looking at the features and we can apply this idea to more complex graphs.
We also notice that the second columns of Z^0 and Z^1 are null in both cases.

References

[1] Greg Corrado Jeffrey Dean Tomas Mikolov, Kai Chen. Efficient estimation of word representations in vector space. *arXiv:1301.3781v3*, 2013.