1 Question 1

1. Directed Graphs

When performing random walks, decide whether to follow outgoing (`successors`) or incoming (`predecessors`) edges. To explore outgoing neighbors, use `G.successors(node)`, or for incoming neighbors, use `G.predecessors(node)`. It is important that node embeddings preserve the directional relationships (e.g., distinguishing roles as source or sink nodes).

2. Weighted Graphs

Bias random walks based on edge weights, treating them as probabilities. Normalize the weights and sample neighboring nodes in proportion to these weights. The probability of transitioning from node u to node v is given by:

$$P(v \mid u) = \frac{w_{u,v}}{\sum_{v' \in \text{Neighbors}(u)} w_{u,v'}}$$

where $w_{u,v}$ represents the edge weight from u to v.

2 Task 7

Spectral Clustering Features Accuracy: 1.0000
DeepWalk Features Accuracy: 1.0000
Both method have a perfect accuracy on this dataset.

3 Question 2

The two embedding matrices X_1 and X_2 are related by a **reflection** represented by the following rotation matrix:

$$R = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

This means X_2 is a mirrored version of X_1, where the x-coordinates are flipped. DeepWalk generates these embeddings by capturing the graph's structure, and this transformation preserves the relative relationships between nodes.

4 Question 3

In the GCN architecture, the receptive field of a node is determined by the number of message passing layers. For a given node i, the maximal distance (in terms of hops) to nodes considered in the calculation of $\hat{Y}_i$ is 2, the number of message passing layers.
In general, with k message passing layers, the maximal number of edges separating node i from the nodes whose features are used in the prediction $\hat{Y}_i$ is also k.

5 Task 11 and 12

Test set results: loss= 0.0062 accuracy= 1.0000
VS
Test set results: loss= 0.8458 accuracy= 0.2857
With a bad initialisation, there is a significant decrease in performances.

6 Question 4

For detailed computations, please refer to the screenshots in the zip file.

The structure of the solution aligns with the graph's topology. In the complete graph, the embeddings of all nodes are identical, reflecting the fully connected nature of the graph. In contrast, for the star graph, the central node has a distinct embedding, while the leaf nodes share the same embedding, highlighting their symmetric relationships to the central node.

If the input were initialized randomly, the embeddings would differ, but they would likely converge toward similar representations, maintaining the overall structure observed with the uniform input.

References