1 Question 1

- (1) In the case of directed graphs, the DeepWalk architecture can be adapted easily: we can still randomly select a neighbor of the current node (but i being a neighbor of j does not imply that j is a neighbor of i).
- (2) In the case of weighted graphs, the DeepWalk architecture can be adapted by picking a neighbor randomly, with the probability of a neighbor being selected proportional to the weight between the current node and this neighbor. That is, if w_{ij} is the (positive) weight between i and j, and $N(i)$ the set of neighbors of i, then we go from i to j with probability $p_{ij} = \frac{w_{ij}}{\sum\limits_{k \in N(i)} w_{ik}}$.

2 Question 2

We observe that the embeddings in $\mathbf{X}_2$ are the image of the embeddings in $\mathbf{X}_1$ by a a **reflection**: the second coordinate is multiplied by -1.
If we set $R = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, then $\mathbf{X}_2 = \mathbf{X}_1 R$.

My interpretation is that though the learning process is stochastic, the embeddings will essentially be invariant up to orthogonal transformations (the distance between two vectors not being modified by such transformations).

3 Question 3

For nodes i, j, $\hat{\mathbf{A}}_{ij} \neq 0$ if and only if i and j are neighbors.
So when computing $\hat{\mathbf{A}}\mathbf{X}$, the features of a given node are replaced by an average of the features of its neighbors. Thus, for $\mathbf{Z}^0 = f(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^0)$, the row corresponding to a node i uses the features of its neighbors only.

The same way, when computing $\hat{\mathbf{A}}\mathbf{Z}^0$, for a node i, the rows of $\mathbf{Z}^0$ that are used are the ones corresponding to the neighbors of j; from the previous point, they then use i's neighbors' features, as well as i's neighbors' neighbors' features. So for $\mathbf{Z}^1 = f(\hat{\mathbf{A}}\mathbf{Z}^0\mathbf{W}^1)$, the row corresponding to a node i uses the features of the nodes with distance at most 2.

And $\hat{\mathbf{Y}} = \text{softmax}(\mathbf{Z}^1\mathbf{W}^2)$ does not change that (right-multiplication by a matrix).

So $\boxed{\text{for the GCN architecture implemented in Task 10, the receptive field is } 2}$.

And the same way (proof by induction), $\boxed{\text{for } k \text{ message passing layers, the receptive field is } k.}$

4 Question 4

Both K_4 and S_4 are represented in Fig. 1.

We use $\mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ and $f = \text{ReLU}$.

For K_4, $\boxed{\mathbf{Z}_1 = \begin{pmatrix} 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \end{pmatrix}}$, and for S_4, $\boxed{\mathbf{Z}_1 \approx \begin{pmatrix} 0 & 0.37 & 0.31 \\ 0 & 0.27 & 0.22 \\ 0 & 0.27 & 0.22 \\ 0 & 0.27 & 0.22 \end{pmatrix}}$. The computations are detailed below.

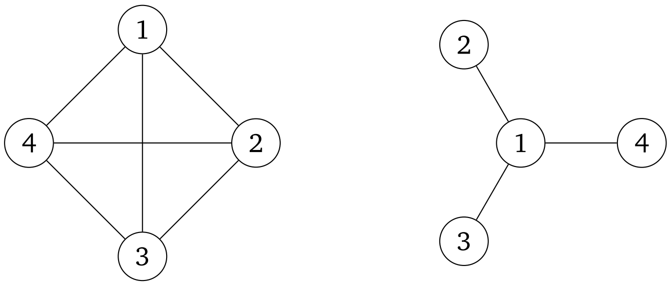

Figure 1: Left: Complete graph K_4. Right: Star graph S_4.

For K_4:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \tilde{\mathbf{A}} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \tilde{\mathbf{D}} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}, \hat{\mathbf{A}} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}.$$

So $\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^0 = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} -0.8 & 0.5 \end{pmatrix} = \begin{pmatrix} -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \end{pmatrix}$, and $\mathbf{Z}^0 = f(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^0) = \begin{pmatrix} 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \end{pmatrix}$.

Then $\hat{\mathbf{A}}\mathbf{Z}^0\mathbf{W}^1 = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix} \begin{pmatrix} 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} 0.1 & 0.3 & -0.05 \\ -0.4 & 0.6 & 0.5 \end{pmatrix} = \begin{pmatrix} -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \\ -0.2 & 0.3 & 0.25 \end{pmatrix}$, so

$$\mathbf{Z}^1 = f(\hat{\mathbf{A}}\mathbf{Z}^0\mathbf{W}^1) = \begin{pmatrix} 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \\ 0 & 0.3 & 0.25 \end{pmatrix}.$$

The computations for S_4 are similar:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \tilde{\mathbf{A}} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \tilde{\mathbf{D}} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

$$\hat{\mathbf{A}} = \begin{pmatrix} 0.25 & \sqrt{2}/4 & \sqrt{2}/4 & \sqrt{2}/4 \\ \sqrt{2}/4 & 0.5 & 0 & 0 \\ \sqrt{2}/4 & 0 & 0.5 & 0 \\ \sqrt{2}/4 & 0 & 0 & 0.5 \end{pmatrix} \approx \begin{pmatrix} 0.25 & 0.35 & 0.35 & \sqrt{2}/4 \\ 0.35 & 0.5 & 0 & 0 \\ 0.35 & 0 & 0.5 & 0 \\ 0.35 & 0 & 0 & 0.5 \end{pmatrix}.$$

So $\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^0 \approx \begin{pmatrix} -1.05 & 0.65 \\ -0.68 & 0.43 \\ -0.68 & 0.43 \\ -0.68 & 0.43 \end{pmatrix}$, and $\mathbf{Z}^0 = f(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^0) \approx \begin{pmatrix} 0 & 0.65 \\ 0 & 0.43 \\ 0 & 0.43 \\ 0 & 0.43 \end{pmatrix}$.

Then $\hat{\mathbf{A}}\mathbf{Z}^0\mathbf{W}^1 \approx \begin{pmatrix} -0.25 & 0.37 & 0.31 \\ -0.18 & 0.27 & 0.22 \\ -0.18 & 0.27 & 0.22 \\ -0.18 & 0.27 & 0.22 \end{pmatrix}$, and $\mathbf{Z}^1 = f(\hat{\mathbf{A}}\mathbf{Z}^0\mathbf{W}^1) \approx \begin{pmatrix} 0 & 0.37 & 0.31 \\ 0 & 0.27 & 0.22 \\ 0 & 0.27 & 0.22 \\ 0 & 0.27 & 0.22 \end{pmatrix}$.

We see that for K_4, all rows of $\mathbf{Z}^1$ are equal, and for S_4, the second, third and fourth row are equal. This makes sense: **if two nodes have the same neighbors** (considering that a node is its own neighbor, since we consider $\hat{A}$), **and same embedding**, then the corresponding rows in $\mathbf{Z}^1$ are equal.

If $\mathbf{X}$ had been randomly sampled from a random uniform distribution, then the rows of $\mathbf{Z}^1$ would be different, and there would be **no particular pattern** in $\mathbf{Z}^1$.