Fotios Kapotos
fotiskapotos@gmail.com

1 Question 1

The window parameter w controls the scale of graph structure captured by DeepWalk by defining how far apart two nodes can be in a random walk and still be treated as contextual neighbors. For small w, only immediate neighbors contribute to the objective, leading to embeddings that emphasize local connectivity and fine-grained graph structure. For large w, nodes that are farther apart along the walk are paired, which captures more global relationships and community-level structure but may blur local distinctions by combining heterogeneous contexts.

2 Question 2

The two embedding matrices are equivalent up to an orthogonal transformation. Specifically, $X_2 = X_1 R$ with $R = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, corresponding to a reflection across one coordinate axis. This illustrates that DeepWalk embeddings are not unique and are defined only up to rotations and reflections of the embedding space.

3 Question 3

We write a single GNN layer as

$$\mathrm{GNN}(A, X) \ = \ f(\hat{A}XW), \qquad \hat{A} = \tilde{D}^{-1/2}(A + I)\tilde{D}^{-1/2},$$

where f is applied elementwise (e.g. ReLU).
Let P be a permutation matrix and define

$$A' = PAP^\top, \qquad X' = PX.$$

Then

$$\tilde{A}' = A' + I = PAP^\top + I = P(A + I)P^\top = P\tilde{A}P^\top.$$

Let $\mathbf{1}$ be the all-ones vector. The degree matrix satisfies

$$\tilde{D} = \mathrm{diag}(\tilde{A}\mathbf{1}), \qquad \tilde{D}' = \mathrm{diag}(\tilde{A}'\mathbf{1}) = \mathrm{diag}(P\tilde{A}P^\top\mathbf{1}) = \mathrm{diag}(P\tilde{A}\mathbf{1}) = P\tilde{D}P^\top.$$

Since P is a permutation matrix and $\tilde{D}$ is diagonal,

$$(\tilde{D}')^{-1/2} = (P\tilde{D}P^\top)^{-1/2} = P\tilde{D}^{-1/2}P^\top.$$

Hence the normalized adjacency after permutation is

$$\hat{A}' = (\tilde{D}')^{-1/2}\tilde{A}'(\tilde{D}')^{-1/2} = P\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}P^\top = P\hat{A}P^\top.$$

Now

$$\mathrm{GNN}(PAP^\top, PX) = f(\hat{A}'X'W) = f(P\hat{A}P^\top PXW) = f(P\hat{A}XW).$$

Since f is applied elementwise (rowwise), it commutes with the permutation:

$$f(PY) = Pf(Y) \quad \text{for all } Y.$$

Therefore

$$\mathrm{GNN}(PAP^\top, PX) = Pf(\hat{A}XW) = P\,\mathrm{GNN}(A, X),$$

which proves permutation equivariance.

4 Question 4

1. u is an eigenvector of $\hat{A}$ with eigenvalue 1.
Let $\tilde{A} = A + I$ and $\tilde{D} = \mathrm{diag}(\tilde{d}_1, \dots, \tilde{d}_n)$ with $\tilde{d}_i = \sum_j \tilde{A}_{ij}$. The normalized adjacency is

$$\hat{A} = \tilde{D}^{-1/2}\, \tilde{A}\, \tilde{D}^{-1/2}.$$

Define $u \in \mathbb{R}^n$ by $u_i = \sqrt{\tilde{d}_i}$. Then

$$(\hat{A}u)_i = \sum_j \tilde{D}_{ii}^{-1/2}\, \tilde{A}_{ij}\, \tilde{D}_{jj}^{-1/2}\, u_j = \frac{1}{\sqrt{\tilde{d}_i}} \sum_j \tilde{A}_{ij} \frac{1}{\sqrt{\tilde{d}_j}} \sqrt{\tilde{d}_j} = \frac{1}{\sqrt{\tilde{d}_i}} \sum_j \tilde{A}_{ij} = \frac{\tilde{d}_i}{\sqrt{\tilde{d}_i}} = \sqrt{\tilde{d}_i} = u_i.$$

Hence $\hat{A}u = u$, i.e. u is an eigenvector with eigenvalue $\lambda = 1$.

2. Limit of $Z^{(k)}$ as $k \to \infty$.
Since the graph is connected and non-bipartite, $\hat{A}$ is symmetric with eigenvalues $1 = \lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_n|$ and an orthonormal basis of eigenvectors $q_1, \dots, q_n$ with $q_1 \propto u$. Write the spectral decomposition

$$\hat{A} = Q \Lambda Q^\top, \quad Q = [q_1 \ \dots \ q_n], \quad \Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n).$$

Then

$$\hat{A}^k = Q \Lambda^k Q^\top \xrightarrow[k \to \infty]{} Q\, \mathrm{diag}(1, 0, \dots, 0)\, Q^\top = q_1 q_1^\top.$$

Thus

$$Z^{(k)} = \hat{A}^k X W \xrightarrow[k \to \infty]{} (q_1 q_1^\top) X W = q_1 \big(q_1^\top X W\big),$$

so every node embedding becomes a scalar multiple of the same global vector $v := q_1^\top X W$.

3. Why nodes with the same degree become indistinguishable.
The first eigenvector q_1 is proportional to u, hence $q_{1,i} \propto \sqrt{\tilde{d}_i}$. In the limit, the representation of node i is

$$Z_i^{(\infty)} = q_{1,i}\, v,$$

which depends on i only through $\sqrt{\tilde{d}_i}$. Therefore, any two nodes with the same degree (same $\tilde{d}_i$) have identical limiting representations, regardless of their initial features X, illustrating oversmoothing.

References