Kevin Wardakhan
kevin.wardakhan@icloud.com

Lab session # 6
ALTEGRAD 2025

12/01/25

1 Question 1

In DeepWalk, the window parameter w sets how many neighbors in the random walk are treated as context for a vertex v_i.
A small w means each node is trained mainly with its immediate neighbors, so the embeddings capture very local structure with low bias to the local graph, but high variance because they rely on few, specific co-occurrences.
A large w includes nodes several steps away in the walk, so the embeddings reflect higher order proximity and global communities, at the cost of blurring local neighborhoods and increasing bias while reducing variance.
This is reminiscent of the bias variance trade-off in kernel methods such as the **Nadaraya–Watson** estimator: a small bandwidth yields very local estimates with low bias but high variance, whereas a large bandwidth produces smoother, more global estimates with higher bias and lower variance.

2 Question 2

We can write the two embedding matrices as

$$X_1 = \begin{bmatrix} -1.0 & 1.0 \\ -1.0 & 1.0 \\ -0.5 & 0.5 \\ 0.5 & -0.5 \\ 1.0 & -1.0 \\ 1.0 & -1.0 \end{bmatrix}, \qquad X_2 = \begin{bmatrix} -1.0 & -1.0 \\ -1.0 & -1.0 \\ -0.5 & -0.5 \\ 0.5 & 0.5 \\ 1.0 & 1.0 \\ 1.0 & 1.0 \end{bmatrix}.$$

Define

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

Then we have

$$X_2 = X_1 R.$$

Since $R^\top R = I$, the matrix R is orthogonal, the two embedding results are the same up to an axial symmetry with respect to the horizontal axis.

3 Question 3

We consider a single GNN layer of the form

$$\text{GNN}(A, X) = f(\hat{A}XW),$$

where

$$\hat{A} = \tilde{D}^{-1/2}(A + I)\tilde{D}^{-1/2},$$

X is the node feature matrix, W the weight matrix, and f an elementwise activation function such as $\tanh$, ReLU or sigmoid.

Key property of the activation function.

Let P be a permutation matrix and $Z \in \mathbb{R}^{n\times d}$ an arbitrary matrix. Since P only permutes the rows of Z, and f is applied independently and identically on each row (and each coordinate), permuting before or after applying f gives the same result:

$$f(PZ) = Pf(Z), \qquad \forall Z \in \mathbb{R}^{n\times d}. \tag{$*$}$$

Let

$$A' = PAP^\top, \qquad \tilde{A} = A + I, \qquad \tilde{A}' = A' + I.$$

Then
$$\tilde{A}' = A' + I = PAP^\top + I = P(A+I)P^\top = P\tilde{A}P^\top.$$
Let $\tilde{D}$ and $\tilde{D}'$ be the diagonal matrices associated with $\tilde{A}$ and $\tilde{A}'$, defined componentwise by
$$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij} \quad \text{and} \quad \tilde{D}'_{ii} = \sum_j \tilde{A}'_{ij}$$
Using $\tilde{A}' = P\tilde{A}P^\top$, we obtain
$$\tilde{D}'_{ii} = \sum_j \tilde{A}'_{ij} = \sum_j (P\tilde{A}P^\top)_{ij} = \sum_j \tilde{A}_{\pi(i)\,\pi(j)} = \sum_\ell \tilde{A}_{\pi(i)\,\ell} = \tilde{D}_{\pi(i)\,\pi(i)}$$
where π is the permutation associated with P. Thus the diagonal entries of $\tilde{D}'$ are obtained by permuting those of $\tilde{D}$, which can be written in matrix form as
$$\tilde{D}' = P\tilde{D}P^\top$$
Since $\tilde{D}$ is diagonal with positive entries, its inverse square root is well defined and we also have
$$\tilde{D}'^{-1/2} = (P\tilde{D}P^\top)^{-1/2} = P\tilde{D}^{-1/2}P^\top$$
Therefore the normalized adjacency for the permuted graph is
$$\hat{A}' = \tilde{D}'^{-1/2}\tilde{A}'\tilde{D}'^{-1/2} = (P\tilde{D}^{-1/2}P^\top)(P\tilde{A}P^\top)(P\tilde{D}^{-1/2}P^\top) = P\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}P^\top = P\hat{A}P^\top$$
We now compute
$$\text{GNN}(PAP^\top, PX) = f\big(\hat{A}'\,PXW\big) = f\big(P\hat{A}P^\top PXW\big) = f\big(P\hat{A}XW\big)$$
Applying property $(*)$ gives,
$$f\big(P\hat{A}XW\big) = Pf(\hat{A}XW) = P \cdot \text{GNN}(A, X)$$
Hence
$$\boxed{\text{GNN}(PAP^\top, PX) = P \cdot \text{GNN}(A, X)}$$

4 Question 4

4.1 Question 4.1

We consider the normalized adjacency matrix
$$\hat{A} = \tilde{D}^{-1/2}\,\tilde{A}\,\tilde{D}^{-1/2},$$
where $\tilde{A} = A + I$ and $\tilde{D}$ is the diagonal degree matrix of $\tilde{A}$. More precisely,
$$\tilde{D}_{ii} = \tilde{d}_i = \sum_j \tilde{A}_{ij},$$
that is, $\tilde{d}_i$ is the degree of node i including the self-loop. Let $u \in \mathbb{R}^n$ be defined by
$$u_i = \sqrt{\tilde{d}_i}, \quad i = 1, \ldots, n$$
We want to show that $\hat{A}u = u$.

First compute $\tilde{D}^{-1/2}u$. We have
$$\tilde{D}^{-1/2} = \begin{pmatrix} \tilde{d}_1^{-1/2} & & 0 \\ & \ddots & \\ 0 & & \tilde{d}_n^{-1/2} \end{pmatrix}, \qquad u = \begin{pmatrix} \sqrt{\tilde{d}_1} \\ \vdots \\ \sqrt{\tilde{d}_n} \end{pmatrix},$$
hence
$$\tilde{D}^{-1/2}u = \begin{pmatrix} \tilde{d}_1^{-1/2}\sqrt{\tilde{d}_1} \\ \vdots \\ \tilde{d}_n^{-1/2}\sqrt{\tilde{d}_n} \end{pmatrix} = \mathbf{1}_n$$

By definition of the degrees, for each i we have

$$\tilde{d}_i = \sum_j \tilde{A}_{ij},$$

which in vector form gives

$$\tilde{A}\mathbf{1} = \begin{pmatrix} \sum_j \tilde{A}_{1j} \\ \vdots \\ \sum_j \tilde{A}_{nj} \end{pmatrix} = \begin{pmatrix} \tilde{d}_1 \\ \vdots \\ \tilde{d}_n \end{pmatrix} =: \tilde{d}$$

We can now compute $\hat{A}u$:

$$\hat{A}u = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}u = \tilde{D}^{-1/2}\tilde{A}\mathbf{1}_n = \tilde{D}^{-1/2}\tilde{d}$$

Write this last product explicitly:

$$\tilde{D}^{-1/2}\tilde{d} = \begin{pmatrix} \tilde{d}_1^{-1/2} & & 0 \\ & \ddots & \\ 0 & & \tilde{d}_n^{-1/2} \end{pmatrix} \begin{pmatrix} \tilde{d}_1 \\ \vdots \\ \tilde{d}_n \end{pmatrix} = \begin{pmatrix} \tilde{d}_1^{1/2} \\ \vdots \\ \tilde{d}_n^{1/2} \end{pmatrix} = u$$

Thus $\boxed{\hat{A}u = u}$, which proves that u is an eigenvector of $\hat{A}$ associated with the eigenvalue $\lambda = 1$.

4.2 Question 4.2

We are given the linear GCN with k layers

$$Z^{(k)} = \hat{A}^k XW,$$

where $\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ is real symmetric. Hence $\hat{A}$ admits an orthonormal eigendecomposition

$$\hat{A} = U\Lambda U^\top = \sum_{i=1}^n \lambda_i\, u_i u_i^\top,$$

where λ_i are the eigenvalues and u_i the corresponding orthonormal eigenvectors.
For a connected, non-bipartite graph it is known that

$$\lambda_1 = 1 \quad \text{with multiplicity } 1, \qquad |\lambda_i| < 1 \ \text{ for } i = 2, \dots, n.$$

We choose u_1 to be the eigenvector associated with $\lambda_1 = 1$, and u_i, λ_i for $i \geq 2$ the remaining eigenpairs.
Using the spectral decomposition,

$$\hat{A}^k = U\Lambda^k U^\top = \sum_{i=1}^n \lambda_i^k\, u_i u_i^\top.$$

Therefore

$$Z^{(k)} = \hat{A}^k XW = \sum_{i=1}^n \lambda_i^k\, u_i u_i^\top XW.$$

Since $|\lambda_i| < 1$ for all $i \geq 2$, we have $\lambda_i^k \to 0$ as $k \to \infty$ for $i \geq 2$, while $\lambda_1^k = 1$. Hence

$$\boxed{\lim_{k\to\infty} Z^{(k)} = \lim_{k\to\infty} \sum_{i=1}^n \lambda_i^k\, u_i u_i^\top XW = u_1 u_1^\top XW}$$

4.3 Question 4.3

From 4.2, as $k \to \infty$ we have $Z^{(k)} \to u_1 u_1^\top XW$, so the representation of node i converges to $u_{1,i}(u_1^\top XW)$.

Therefore, any two nodes with the same degree (hence the same $u_{1,i} = \sqrt{\tilde{d}_i}$) end up with exactly the same limit representation, and this deep architecture cannot distinguish them regardless of the initial features X.