Ruben Cardoso
ruben.cardoso@dauphine.eu

Lab Session 6
ALTEGRAD 2025

For 12/2/25

QUESTION 1

The window parameter w controls how many neighboring vertices (in the random walk sequence) are used as context for learning the embedding of a vertex.

- A very small w focuses only on immediate neighbors, producing highly local embeddings that capture fine structural details but miss global relationships..
- Conversely : a large w incorporates distant nodes in the walk, yielding more global and smoother embeddings, at the cost of introducing noise and losing local precision.

Thus, w determines the trade-off between local and global structural information !

QUESTION 2

The two embeddings X_1 and X_2 differ only by a global orthogonal transformation (rotation or reflection): there exists an orthogonal matrix R such that

$$X_2 = X_1 R.$$

The distances and the relative structure remain identical, only the orientation of the coordinate system changes.

QUESTION 3

We want to show that

$$\mathrm{GNN}(PAP^T,\, PX) = P\,\mathrm{GNN}(A, X),$$

for any permutation matrix P.

Let $\widetilde{A} = A + I$ and let $\widetilde{D}$ be the diagonal matrix with $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$. The normalized adjacency is

$$\widehat{A} = \widetilde{D}^{-1/2}\,\widetilde{A}\,\widetilde{D}^{-1/2}.$$

For the permuted graph,

$$A' = PAP^T, \qquad X' = PX,$$

and therefore

$$\widetilde{A}' = A' + I = P(A + I)P^T = P\widetilde{A}P^T.$$

Since P only permutes diagonal entries, the degree matrix transforms as

$$\widetilde{D}' = P\widetilde{D}P^T, \qquad (\widetilde{D}')^{-1/2} = P\widetilde{D}^{-1/2}P^T.$$

Hence the normalized adjacency satisfies

$$\widehat{A}' = (\widetilde{D}')^{-1/2}\,\widetilde{A}'\,(\widetilde{D}')^{-1/2} = P\,\widehat{A}\,P^T.$$

A single GNN layer is defined by

$$\mathrm{GNN}(A, X) = f(\widehat{A}\,X\,W).$$

Thus for the permuted graph,

$$\mathrm{GNN}(A', X') = f(\widehat{A}'X'W) = f((P\widehat{A}P^T)(PX)W) = f(P(\widehat{A}XW)).$$

Since a pointwise nonlinearity f commutes with the row-permutation P,

$$f(PY) = Pf(Y),$$

we obtain

$$\mathrm{GNN}(A', X') = P\, f(\widehat{A}XW) = P\,\mathrm{GNN}(A, X).$$

QUESTION 4:

1. We have $\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$, where $\tilde{D}_{ii} = \tilde{d}_i$ and $\tilde{A}$ is the adjacency matrix with self-loops. Consider the vector $u \in \mathbb{R}^n$ defined by $u_i = \sqrt{\tilde{d}_i}$. Then, for each node i,

$$(\hat{A}u)_i = \sum_{j=1}^{n} \hat{A}_{ij}u_j = \sum_{j=1}^{n} \frac{\tilde{A}_{ij}}{\sqrt{\tilde{d}_i\tilde{d}_j}}\sqrt{\tilde{d}_j} = \frac{1}{\sqrt{\tilde{d}_i}}\sum_{j=1}^{n}\tilde{A}_{ij} = \frac{\tilde{d}_i}{\sqrt{\tilde{d}_i}} = \sqrt{\tilde{d}_i} = u_i.$$

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

2. Let the spectral decomposition of $\hat{A}$ be

$$\hat{A} = \sum_{\ell=1}^{n} \lambda_\ell v_\ell v_\ell^\top,$$

   where (v_ℓ) is an orthonormal basis of eigenvectors, with $\lambda_1 = 1$ (simple) and $|\lambda_\ell| < 1$ for all $\ell \geq 2$. Then

$$\hat{A}^k = \sum_{\ell=1}^{n} \lambda_\ell^k v_\ell v_\ell^\top.$$

   As $k \to \infty$, we have $\lambda_1^k = 1$ while $\lambda_\ell^k \to 0$ for $\ell \geq 2$. Therefore,

$$\hat{A}^k \xrightarrow[k\to\infty]{} v_1 v_1^\top.$$

   Since v_1 is the normalized version of u (i.e. $v_1 = u/\|u\|_2$), we obtain

$$Z^{(k)} = \hat{A}^k XW \xrightarrow[k\to\infty]{} v_1 v_1^\top XW = \frac{uu^\top}{\|u\|_2^2}XW.$$

3. In the limit, the representation of node i is

$$Z_{i,:}^{(\infty)} = \frac{u_i}{\|u\|_2^2}(u^\top XW),$$

   which is a scalar multiple of the *same* vector for all nodes. Since $u_i = \sqrt{\tilde{d}_i}$, all nodes with the same degree receive exactly the same representation, regardless of their initial features. Thus, deep GCNs cannot distinguish between nodes that share the same degree.