1 Question 1

In Eq. (1), the parameter w controls how many steps of the random walk around a node v_i are treated as its "context". If w is very small, each node is trained mainly to predict its immediate neighbours, so the embeddings capture very *local* structure, but may miss larger communities.
If w is large, nodes that are several steps apart along the walk are also treated as context. Then the embeddings reflect more *global* or high-order proximity (nodes in the same community are pulled closer), but if w is too large, many weakly related nodes are included, which can blur fine local differences and add noise to the representations.

2 Question 2

In general, node embeddings are only meaningful through the geometry they induce: what matters are pairwise distances and similarities, not the absolute coordinates. Any embedding that can be obtained from another by an isometry (e.g. a rotation or reflection) encodes the same information, because all pairwise distances are preserved. More generally, if we allow distances to be measured with a norm induced by a positive definite matrix M (a Mahalanobis norm $\|x-y\|_M = \sqrt{(x-y)^\top M(x-y)}$), then embeddings related by a linear change of basis have the same relative structure under the corresponding metric.
In our example, the two DeepWalk runs produce X_1 and X_2, which induce exactly the same pairwise distances between nodes under the usual Euclidean norm. Thus they define the same neighborhoods and clusters; one configuration is simply a mirrored version of the other in the 2D space. Consequently, X_1 and X_2 should be regarded as equivalent embeddings that differ only by a change of coordinates.

3 Question 3

We consider one GNN layer

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

with $\tilde{A} = A + I$ and

$$\tilde{D} = \mathrm{diag}(\tilde{A}\mathbf{1}),$$

i.e. $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. Let P be an $n \times n$ permutation matrix.
let s define

$$A' = PAP^\top.$$

For the normalized GNN layer we use $\tilde{A}' = A' + I$, so

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

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

so $(\tilde{D}')^{-1/2} = P\tilde{D}^{-1/2}P^\top$ as well.
Hence the normalized adjacency for the permuted graph is

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

Now consider the permuted inputs $(A', PX) = (PAP^\top, PX)$:

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

The nonlinearity f is applied elementwise, so for any matrix Y, permuting the rows before or after f gives the same result: $f(PY) = Pf(Y)$. Therefore

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

which proves permutation equivariance of the GNN layer.

4 Question 4

1) Let $\tilde{A} = A + I$ and $\tilde{D} = \operatorname{diag}(\tilde{A}\mathbf{1})$, and define

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

Let $u \in \mathbb{R}^n$ be defined by

$$u_i = \sqrt{\tilde{d}_i}, \qquad \tilde{d}_i = \sum_j \tilde{A}_{ij}.$$

Then

$$\tilde{D}^{-\frac{1}{2}}u = \mathbf{1},$$

since $(\tilde{D}^{-\frac{1}{2}}u)_i = \tilde{d}_i^{-1/2}\sqrt{\tilde{d}_i} = 1$. Therefore

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

because $(\tilde{D}^{-\frac{1}{2}}\tilde{d})_i = \tilde{d}_i^{-1/2}\tilde{d}_i = \sqrt{\tilde{d}_i} = u_i$. Thus u is an eigenvector of $\hat{A}$ with eigenvalue $\lambda = 1$.

2) Since $\hat{A}$ is real symmetric, by the spectral theorem there exists an orthonormal basis of eigenvectors $v^{(1)}, \dots, v^{(n)}$ and real eigenvalues $\lambda_1, \dots, \lambda_n$ such that

$$\hat{A}v^{(\ell)} = \lambda_\ell v^{(\ell)}, \qquad \ell = 1, \dots, n,$$

and

$$\hat{A} = \sum_{\ell=1}^{n} \lambda_\ell \, v^{(\ell)}(v^{(\ell)})^\top.$$

We know from the statement that $\lambda_1 = 1$ (with multiplicity 1) and $|\lambda_\ell| < 1$ for all $\ell \geq 2$. Moreover, the eigenvector associated with $\lambda_1 = 1$ is proportional to u, so we can choose

$$v^{(1)} = \frac{u}{\|u\|}.$$

For the k-th power, we get

$$\hat{A}^k = \left(\sum_{\ell=1}^{n} \lambda_\ell \, v^{(\ell)}(v^{(\ell)})^\top\right)^k = \sum_{\ell=1}^{n} \lambda_\ell^k \, v^{(\ell)}(v^{(\ell)})^\top,$$

since the $v^{(\ell)}$ form an orthonormal eigenbasis. As $k \to \infty$,

$$\lambda_1^k = 1, \qquad \lambda_\ell^k \to 0 \text{ for } \ell \geq 2,$$

so

$$\lim_{k\to\infty} \hat{A}^k = v^{(1)}(v^{(1)})^\top = \frac{u}{\|u\|}\frac{u^\top}{\|u\|} = \frac{uu^\top}{\|u\|^2}.$$

Now for the k-layer linear GCN

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

we obtain

$$\lim_{k\to\infty} Z^{(k)} = \left(\frac{uu^\top}{\|u\|^2}\right) XW = \frac{u}{\|u\|^2}\left(u^\top XW\right).$$

3) The limiting representation of node i is

$$Z_i^{(\infty)} = \frac{u_i}{\|u\|^2}\left(u^\top XW\right), \qquad u_i = \sqrt{\tilde{d}_i},$$

so nodes with the same degree receive exactly the same embedding vector, regardless of their initial features X.