Paul Londres
paul.londres@free.fr

Lab session # 5
ALTEGRAD 2024

11/26/24

Question 1

Here is how to adapt the DeepWalk architecture to apply to both situation :

(a). In the case of a directed graph, we could simply sample the next walk not on neighbors but on successors of the graph, thus respecting the direction of the edges.

(b). To adapt to weighted graphs, instead of sampling uniformly among neighbors for determining the next walk, we could weight the probability of going to each neighboring node with the edge's weight. If we take u, the current node and $\{v_1, v_2, ..., v_k\}$ its neighbors, the transition formula could be :

$$P(u, v_i) = \frac{weight(u, v_i)}{\sum_{j=0}^{k} weight(u, v_j)}$$

Question 2

The two embeddings are closely related, indeed in the two matrices, each node has the same value in the first dimension and the opposite value in the second dimension, so we have the following relationship between X_1 and X_2 :

$$X_1 = X_2 * \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

We can notice that both embeddings have a symmetry between the first 3 and the last 3 nodes. In both cases, the embeddings of the first 3 nodes are the exact opposite of those of the last 3 nodes, reflecting the inherent symmetry of the graph. Thus, even if the two matrices can look different at first sight, they ultimately convey the same structural meaning.

Question 3

In this GCN architecture, the size of the receptive field directly depends on the number of message passing layer in the model. Indeed at each message passing layer, we multiply the embedding matrix X with $\hat{A}$ which corresponds to the normalized adjacency matrix giving for each node the information of its neighboring nodes. Hence, if we only multiply these matrices once, each node will only access to the information of its neighbors for the calculation of $\hat{Y}$.
In our case we use two message passing layers which means that we do two times this matrix multiplication increasing the receptive field to 2. If we have k message passing layer, the receptive field of the nodes is thus equal to k except if k exceeds, the maximum shortest distance between two nodes, it would then equal this distance.

Question 4

Computations for complete graph K_4

1. Adjacency matrix A :

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

2. Matrix $\bar{A} = A + I$: $\bar{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

3. Node degree : $D = \text{diag}(4, 4, 4, 4)$. Matrix $D^{-\frac{1}{2}}$:

$$D^{-\frac{1}{2}} = \text{diag}(0.5, 0.5, 0.5, 0.5) = 1/2 * I$$

4. Normaliazed adjacency matrix $\hat{A}$:

$$\hat{A} = D^{-\frac{1}{2}}\hat{A}D^{-\frac{1}{2}} = \frac{1}{4}\begin{bmatrix}1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\1 & 1 & 1 & 1\end{bmatrix}$$

5. Let's now compute Z^0 :

$$Z^0 = \text{ReLU}(\hat{A}XW^0)$$

We have :

$$XW^0 = \begin{bmatrix}-0.8 & 0.5\\-0.8 & 0.5\\-0.8 & 0.5\\-0.8 & 0.5\end{bmatrix}$$

Thus :

$$\hat{A}XW^0 = \frac{1}{4}\begin{bmatrix}1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\1 & 1 & 1 & 1\end{bmatrix}\begin{bmatrix}-0.8 & 0.5\\-0.8 & 0.5\\-0.8 & 0.5\\-0.8 & 0.5\end{bmatrix} = \begin{bmatrix}-0.8 & 0.5\\-0.8 & 0.5\\-0.8 & 0.5\\-0.8 & 0.5\end{bmatrix}$$

By applying ReLu function :

$$Z^0 = \begin{bmatrix}0 & 0.5\\0 & 0.5\\0 & 0.5\\0 & 0.5\end{bmatrix}.$$

6. Let's compute Z^1 :

$$Z^1 = \text{ReLU}(\hat{A}Z^0W^1).$$

We have :

$$Z^0W^1 = \begin{bmatrix}-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\end{bmatrix}.$$

Then :

$$\hat{A}Z^0W^1 = \frac{1}{4}\begin{bmatrix}1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\1 & 1 & 1 & 1\end{bmatrix}\begin{bmatrix}-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\end{bmatrix} = \begin{bmatrix}-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\\-0.2 & 0.3 & 0.25\end{bmatrix}$$

Applying ReLu function, we get the final result :

$$Z^1 = \begin{bmatrix}0 & 0.3 & 0.25\\0 & 0.3 & 0.25\\0 & 0.3 & 0.25\\0 & 0.3 & 0.25\end{bmatrix}$$

Computations for star graph S_4

1. Adjacency matrix A :

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

2. Matrix $\bar{A} = A + I$:

$$\bar{A} = \begin{bmatrix}1 & 1 & 1 & 1\\1 & 1 & 0 & 0\\1 & 0 & 1 & 0\\1 & 0 & 0 & 1\end{bmatrix}.$$

3. Node degree : $D = \text{diag}(4, 2, 2, 2)$. Matrix $D^{-\frac{1}{2}}$:

$$D^{-\frac{1}{2}} = \text{diag}(0.5, 1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2})$$

4. Normaliazed adjacency matrix $\hat{A}$:

$$\hat{A} = D^{-\frac{1}{2}}\hat{A}D^{-\frac{1}{2}} = \frac{1}{2}\begin{bmatrix} 0.5 & 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1 & 0 & 0 \\ 1/\sqrt{2} & 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 0 & 1 \end{bmatrix}$$

5. Let's now compute Z^0 :

$$Z^0 = \text{ReLU}(\hat{A}XW^0)$$

We have :

$$XW^0 = \begin{bmatrix} -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \\ -0.8 & 0.5 \end{bmatrix}$$

Thus :

$$\hat{A}XW^0 \approx \begin{bmatrix} -1.049 & 0.656 \\ -0.683 & 0.427 \\ -0.683 & 0.427 \\ -0.683 & 0.427 \end{bmatrix}$$

By applying ReLu function :

$$Z^0 = \begin{bmatrix} 0 & 0.656 \\ 0 & 0.427 \\ 0 & 0.427 \\ 0 & 0.427 \end{bmatrix}$$

6. Let's compute Z^1 :

$$Z^1 = \text{ReLU}(\hat{A}Z^0W^1).$$

We have :

$$Z^0W^1 \approx \begin{bmatrix} -0.262 & 0.393 & 0.328 \\ -0.171 & 0.256 & 0.213 \\ -0.171 & 0.256 & 0.213 \\ -0.171 & 0.256 & 0.213 \end{bmatrix}$$

Then :

$$\hat{A}Z^0W^1 = \frac{1}{2}\begin{bmatrix} 0.5 & 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1 & 0 & 0 \\ 1/\sqrt{2} & 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -0.262 & 0.393 & 0.328 \\ -0.171 & 0.256 & 0.213 \\ -0.171 & 0.256 & 0.213 \\ -0.171 & 0.256 & 0.213 \end{bmatrix} \approx \begin{bmatrix} -0.247 & 0.370 & 0.308 \\ -0.178 & 0.267 & 0.223 \\ -0.178 & 0.267 & 0.223 \\ -0.178 & 0.267 & 0.223 \end{bmatrix}$$

Applying ReLu function, we get the final result :

$$Z^1 = \begin{bmatrix} 0 & 0.370 & 0.308 \\ 0 & 0.267 & 0.223 \\ 0 & 0.267 & 0.223 \\ 0 & 0.267 & 0.223 \end{bmatrix}$$

The structure of the computed embeddings

The main thing that we observe from the matrices Z^1 is that each similar node has the same embeddings. For example for the first complete graph, all nodes had the same embeddings because of the graph's symmetry and the fact that every node has the same degree and connectivity pattern.
For the second "star" graph, we notice that the peripheral nodes all have the same embeddings and the central node has different ones. This is due to the different structure of the central node which has a degree of 3 and the peripheral nodes that have the same degree of one. The fact that each peripheral node has the same embeddings is due the the graph's symmetry.

If the nodes' features X was randomly sampled, the representation Z^1 would not only depend on the graph structure. Even for highly symmetrical graph like K_4, each node would have a different embedding as their initial feature is different.