3 Assignment 3 - Task 2

3.1 Theory Question 2 - Permutation Invariance of Aggregation Functions

For N vertices and M edges, let G be a graph $G = (V, E)$ with $|V| = N$ and $|E| = M$.

Let $x_i \in \mathbb{R}^k$ be the features of vertex i. We can stack them into a vertex feature matrix $X \in \mathbb{R}^{N \times k}$:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

We can represent these edges with an adjacency matrix $A \in \mathbb{R}^{N \times N}$, such that:

$$a_{ij} = \begin{cases} 1, & \text{if } (i, j) \in E \\ 0, & \text{otherwise} \end{cases}$$

It is worth noting that $\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} = 2M$.

For a vertex i, its (1-hop) neighborhood is commonly defined as follows:

$$\mathcal{N}_i = \{j : (i, j) \in E \vee (j, i) \in E\}$$

Accordingly, we can extract the multiset of features in the neighborhood:

$$X_{\mathcal{N}_i} = \{\{x_j : j \in \mathcal{N}_i\}\}$$

To prove permutation invariance, we need to show that the aggregation functions applied to the adjacency matrix A are invariant under the arbitrary ordering of the rows or columns.

$$f(PX, PAP^T) = f(X, A)$$

Now, we apply the aggregate function g over all neighborhoods:

$$f(X, A) = \begin{bmatrix} g(x_1, X_{\mathcal{N}_1}) \\ g(x_2, X_{\mathcal{N}_2}) \\ \vdots \\ g(x_N, X_{\mathcal{N}_N}) \end{bmatrix}$$

i) Let g be the sum over each neighborhood, then f becomes:

$$f(X, A) = \begin{bmatrix} \sum_{x \in X_{\mathcal{N}_1}} x \\ \sum_{x \in X_{\mathcal{N}_2}} x \\ \vdots \\ \sum_{x \in X_{\mathcal{N}_N}} x \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{|\mathcal{N}_1|} x_i \\ \sum_{i=0}^{|\mathcal{N}_2|} x_i \\ \vdots \\ \sum_{i=0}^{|\mathcal{N}_N|} x_i \end{bmatrix} \stackrel{*}{=} f(PX, PAP^T)$$

Explanation: The sum over each node's neighbors applied to the vertex matrix X and the adjacency matrix A yields the same result as the sum over each node's neighbors applied to the permutated vertex matrix PX and the permutated adjacency matrix PAP^T due to commutativity of addition. Thus, the function f is permutation invariant.

ii) Let g be the max over each neighborhood, then f becomes:

$$f(X, A) = \begin{bmatrix} \max_{x \in X_{\mathcal{N}_1}} x \\ \max_{x \in X_{\mathcal{N}_2}} x \\ \vdots \\ \max_{x \in X_{\mathcal{N}_N}} x \end{bmatrix} \stackrel{*}{=} f(PX, PAP^T)$$

Explanation: The maximum value over each node's neighbors applied to the vertex matrix X and the adjacency matrix A is equal to the maximum value over each node's neighbors applied to the permutated vertex matrix PX and the permutated adjacency matrix PAP^T. This holds true because the maximum operation is commutative, and permuting the order of the nodes does not affect the maximum value. Hence, the function f is permutation invariant.

iii) Let g be the average over each neighborhood, then f becomes:

$$f(X, A) = \begin{bmatrix} \frac{1}{|\mathcal{N}_1|} \sum_{x \in X_{\mathcal{N}_1}} x \\ \frac{1}{|\mathcal{N}_2|} \sum_{x \in X_{\mathcal{N}_2}} x \\ \vdots \\ \frac{1}{|\mathcal{N}_N|} \sum_{x \in X_{\mathcal{N}_N}} x \end{bmatrix} \stackrel{*}{=} f(PX, PAP^T)$$

Explanation: The average value over each node's neighbors applied to the vertex matrix X and the adjacency matrix A is equal to the average value over each node's neighbors applied to the permutated vertex matrix PX and the permutated adjacency matrix PAP^T. This holds true because averaging is a linear operation and is therefore invariant to permutations. Thus, the function f is permutation invariant.