Appendix

1 PROOF OF THEOREM 1

Theorem 1. *For a series of graphs $\mathcal{D} = \{(\mathcal{G}_1, y_1), \dots, (\mathcal{G}_n, y_n)\}$, a pre-trained GNN model $f_{W_0} = (f_1, \cdots, f_n)$ with each $f_i(X_i) = \mathbb{1} \cdot S_i X_i W_0$, and an arbitrary pattern graph $\mathcal{G}_p$, under the non-degeneracy condition [1], a linear projection head θ and a GNN $f_p \in \mathcal{F}$ for prompting, there exists $\mathcal{Y}' = \{y'_1, \dots, y'_n\}$ for which we have*

$$\begin{aligned} l_p &= \min_{f_p \in \mathcal{F}, \theta} \sum_{i=1}^{n} (f_i(X_i + \mathbb{1}^T \cdot f_p(X_p)) \cdot \theta - y_i)^2 \\ &< l_{ft} = \min_{\tilde{f} \in \tilde{\mathcal{F}}, \theta} \sum_{i=1}^{n} (\tilde{f}_i(X_i) \cdot \theta - y_i)^2 \end{aligned} \tag{1}$$

when $y_1 = y'_1, \dots, y_n = y'_n$. Here $\mathbb{1}$ is a vector with all items equals to 1 and with a proper size that changes from line to line, $\tilde{\mathcal{F}}$ is defined in (4), and $\mathcal{F}$ is defined as

$$\mathcal{F} = \left\{ f_p : f_p = \mathbb{1} \cdot S_p X_p W_p, \text{with } \forall\ (W_p - X_p^T X_p) \right\}.$$

Proof. Here S_i, $1 \le i \le n$, and S_p is the diffusion matrix [3] of the graph $(\mathcal{G}_i, y_i)$, $1 \le i \le n$ and the graph $\mathcal{G}_p$ respectively, i.e.,

$$S_i = \sum_{k=0}^{\infty} \theta_{i,k} T_i^k,\ 1 \le i \le n, \quad \text{and} \quad S_p = \sum_{k=0}^{\infty} \theta_{p,k} T_p^k, \tag{2}$$

for some properly chosen weighting coefficients $\{\theta_{i,k}\}_{1 \le i \le n, k \ge 1}$ and $\{\theta_{p,k}\}_{k \ge 1}$, and the properly chosen generalized transition matrix $\{T_i\}_{1 \le i \le n}$ and T_p that are chosen based on the corresponding graph's structure.

Now, recall Lemma 4.3 [2] states that

Lemma 4.3 (Prompt effectiveness guarantee [2]). *For a series of graphs $\mathcal{D} = \{(\mathcal{G}_1, y_1), \dots, (\mathcal{G}_n, y_n)\}$, a pre-trained GNN model $f_{W_0} = (f_1, \cdots, f_n)$ with each $f_i(X_i) = \mathbb{1} \cdot S_i X_i W_0$, under the non-degeneracy condition [1], a linear projection head θ, and a learnable vector $\mathbf{v}$, there exists $\mathcal{Y}' = \{y'_1, \dots, y'_n\}$ for which we have*

$$l_p = \min_{\mathbf{v}, \theta} \sum_{i=1}^{n} (f_i(X_i + \mathbb{1}^T \cdot \mathbf{v}) \cdot \theta - y_i)^2 < l_{ft} = \min_{\tilde{f} \in \tilde{\mathcal{F}}, \theta} \sum_{i=1}^{n} (\tilde{f}_i(X_i) \cdot \theta - y_i)^2 \tag{3}$$

when $y_1 = y'_1, \dots, y_n = y'_n$. Here $\mathbb{1}$ is a vector with all items equals to 1 and with a proper size that changes from line to line, and $\tilde{\mathcal{F}}$ is defined as

$$\tilde{\mathcal{F}} = \left\{ \tilde{f} = (\tilde{f}_1, \cdots, \tilde{f}_n) : \tilde{f}_i = \mathbb{1} \cdot S_i X_i W, 1 \le i \le n, \text{with } \forall\ (W - W_0) \right\}. \tag{4}$$

Therefore, to conclude Theorem 1, we only need to show that for arbitrary given $\mathcal{Y} = \{y_1, \dots, y_n\}$, we have

$$\begin{aligned} &\min_{f_p \in \mathcal{F}, \theta} \sum_{i=1}^{n} (f_i(X_i + \mathbb{1}^T \cdot f_p(X_p)) \cdot \theta - y_i)^2 \\ &= \min_{\mathbf{v}, \theta} \sum_{i=1}^{n} (f_i(X_i + \mathbb{1}^T \cdot \mathbf{v}) \cdot \theta - y_i)^2, \end{aligned} \tag{5}$$

or equivalently, if we assume each X_i has p_0 columns, then, for arbitrary $\mathbf{v} \in \mathbb{R}^{1 \times p_0}$, there exists a matrix $W_p \in \mathbb{R}^{p_0 \times p_0}$ such that

$$\mathbf{v} = \mathbb{1} \cdot S_p X_p W_p \triangleq b_p \cdot W_p. \tag{6}$$

To proof (6), notice that $b_p \in \mathbb{R}^{1 \times p_0}$ and according to Gram-Schmidt orthogonalization process, there exists a base $\{b_1, \cdots, b_{p_0}\}$ such that $b_i = (b_{i1}, \cdots, b_{ip_0}) \in \mathbb{R}^{1 \times p_0}$ for $1 \le i \le n$, $b_1 = b_p / \|b_p\|_2$ and

$$\tilde{W} = (b_1^T, \cdots, b_{p_0}^T) \tag{7}$$

is a orthonormal matrix. Now, denote $e_1 = (1, 0, \cdots, 0) \in \mathbb{R}^{1 \times p_0}$. Apparently,

$$W_p = \frac{1}{\|b_p\|_2} \cdot \left[b_1^T \cdot (\mathbf{v} - e_1) + \tilde{W} \right] \tag{8}$$

is one of the solutions of equation (6). Hence, for arbitrary $\mathbf{v} \in \mathbb{R}^{1 \times p_0}$, equation (6) has a solution $W_p \in \mathbb{R}^{p_0 \times p_0}$, and by applying lemma 4.3, further leads to Theorem 1.

□

REFERENCES

[1] Christopher M Bishop and Nasser M Nasrabadi. 2006. *Pattern recognition and machine learning*. Vol. 4. Springer.

[2] Taoran Fang, Yunchao Zhang, Yang Yang, Chunping Wang, and Lei Chen. 2024. Universal prompt tuning for graph neural networks. *Advances in Neural Information Processing Systems* 36 (2024).

[3] Johannes Gasteiger, Stefan Weißenberger, and Stephan Günnemann. 2019. Diffusion improves graph learning. *Advances in neural information processing systems* 32 (2019).