Graph Neural Thompson Sampling

Shuang Wu
shuangwu222@ucla.edu
Department of Statistics and Data Science
University of California, Los Angeles

Arash A. Amini
aaamini@stat.ucla.edu
Department of Statistics and Data Science
University of California, Los Angeles

Abstract

We consider an online decision-making problem with a reward function defined over graph-structured data. We formally formulate the problem as an instance of graph action bandit. We then propose `GNN-TS`, a Graph Neural Network (GNN) powered Thompson Sampling (TS) algorithm which employs a GNN approximator for estimating the mean reward function and the graph neural tangent features for uncertainty estimation. We prove that, under certain boundness assumptions on the reward function, GNN-TS achieves a state-of-the-art regret bound which is (1) sub-linear of order $\tilde{\mathcal{O}}((\tilde{d}T)^{1/2})$ in the number of interaction rounds, T, and a notion of effective dimension $\tilde{d}$, and (2) independent of the number of graph nodes. Empirical results validate that our proposed `GNN-TS` exhibits competitive performance and scales well on graph action bandit problems.

1 Introduction

Thompson Sampling (Thompson, 1933) is a widely adopted and effective technique in sequential decision-making problems, known for its ease of implementation and practical success (Chapelle and Li, 2011; Kawale et al., 2015; Russo et al., 2018; Riquelme et al., 2018). The fundamental concept behind Thompson Sampling (TS) is to compute the posterior probability of each action being optimal for the present context, followed by the selection of an action from this distribution. Previous research has extended TS or developed variants of it to incorporate increasingly complex models of the reward function, such as Linear TS (Agrawal and Goyal, 2013; Abeille and Lazaric, 2017), Kernelized TS (Chowdhury and Gopalan, 2017), and Neural TS (Zhang et al., 2020). However, these efforts have mainly focused on conventional data types. In contrast, the application of sequential learning to graph-structured data, such as molecular or biological graph representations, introduces unique challenges that merit further investigation.

Recently, there has been a growing interest in studying bandit optimization over graphs. Several researchers have initiated this line of work by addressing the challenge of encoding graph structures in bandit problems (Gómez-Bombarelli et al., 2018; Jin et al., 2018; Griffiths and Hernández-Lobato, 2020; Korovina et al., 2020). More recently, Graph Neural Network (GNN) bandits have been proposed, which leverage expressive GNNs to approximate reward functions on graphs (Kassraie et al., 2022). Despite these advancements, the GNN bandits remain relatively unexplored compared to the extensive research on Neural bandits. Firstly, a formal formulation of this sequential graph selection problem is yet to be proposed. More importantly, there is a significant lack of comprehensive theoretical and empirical investigations regarding the use of TS in sequential graph selection.

Contribution. In this work, we address the online decision-making problem over graph-structured data by contributing a novel algorithm called `GNN-TS`. We begin by formulating the sequential graph selection as graph action bandit. We then propose Graph Neural Thompson Sampling, `GNN-TS`, to incorporate TS exploration with graph neural networks. We establish a regret bound for the proposed algorithm with sub-linear growth of order $\tilde{\mathcal{O}}((\tilde{d}T)^{1/2})$ with respect to the effective dimension $\tilde{d}$ and the number of interaction round T, and independent of the number of graph nodes. Finally, we corroborate the analysis with an empirical evaluation of the algorithm in simulations. Experiments

show that `GNN-TS` yields competitive performance and scalability, compared to the state-of-the-art baselines, underscoring its practical value in addition to its strong theoretical guarantees.

Notations. Let $[n] = \{1, 2, ..., n\}$. For a set or event $\mathcal{E}$, we denote its complement as $\bar{\mathcal{E}}$. $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ is the identity matrix. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}_{i*}$ and $\boldsymbol{A}_{*j}$ denote its i-th row and j-th column, respectively. $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ represents the maximum and minimum eigenvalues of the matrix $\boldsymbol{A}$. For any vector $\boldsymbol{x}$ and square matrix $\boldsymbol{A}$, $\|\boldsymbol{x}\|_{\boldsymbol{A}} = \sqrt{\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}$. We denote the history of randomness up to (but not including) round t as $\mathcal{F}_t$ and write $\mathbb{P}_t(\cdot) := \mathbb{P}(\cdot \mid \mathcal{F}_t)$ and $\mathbb{E}_t(\cdot) := \mathbb{E}[\cdot \mid \mathcal{F}_t]$ for the conditional probability and expectation given $\mathcal{F}_t$. We use $\lesssim$ and big-O, to denote "less than", up to a constant factor. We further use $\tilde{\mathcal{O}}(\cdot)$ for big-O up to logarithmic factor.

2 Related Works

Graph Bandit. Multiple works have studied graph bandit problems, which can be classified into two categories: graph as structure across arms and graph as data. Most research focuses on the former category, starting from spectral bandit (Kocák et al., 2014; 2020) to graphical bandit (Liu et al., 2018; Yu et al., 2020; Gou et al., 2023; Toni and Frossard, 2023). Within this field, bandit problems with graph feedback have garnered significant attention (Tossou et al., 2017; Dann et al., 2020; Chen et al., 2021; Kong et al., 2022), where learners observe rewards from selected nodes and their neighborhoods. The primary focus of these works have been improving sample efficiency (Bellemare et al., 2019; Waradpande et al., 2020; Idé et al., 2022), with some assuming that payoffs are shared according to the graph Laplacian (Esposito et al., 2022; Lee et al., 2020; Lykouris et al., 2020; Thaker et al., 2022; Yang et al., 2020). While the existing literature primarily aims to optimize over geometrical signal domains, our work focuses on optimization within graph domains. Specifically, we investigate the online graph selection problem, aligning with the second category of research that considers the entire graph as input data. A related recent work (Kassraie et al., 2022) proposed a GNN bandit approach with regret bound based on information gain and an elimination-based algorithm. In contrast, our work explores regret bound based on the effective dimension and builds upon the foundation of Thompson Sampling. This second category of research also encompasses empirical works (Upadhyay et al., 2020; Qi et al., 2022; 2023), particularly those centered around molecule optimization (Wang-Henderson et al., 2023a;b).

Neural Bandit. Our work contributes to the research on neural bandits, where deep neural networks are utilized to estimate the reward function. The work of Zahavy and Mannor (2019); Xu et al. (2020) investigated the Neural Linear bandit, while Zhou et al. (2020) developed Neural Upper Confidence Bound (UCB), an extension of Linear UCB. Zhang et al. (2020) adapted TS with deep neural networks, proposing Neural TS. Dai et al. (2022) makes improvements to neural bandit algorithms to overcome practical limitations. Nguyen-Tang et al. (2021) explores neural bandit in an offline contextual bandit setting and (Gu et al., 2024) examines batched learning for neural bandit. Our work can be seen as an extension of Neural TS (Zhang et al., 2020), incorporating significant improvements such as the utilization of graph neural tangent kernel and a distinct definition of effective dimension.

3 Problem Formulation and Methodology

3.1 Graph Action Bandit Problem

We consider an online decision-making problem in which the learner aims to optimize an unknown reward function by sequentially interacting with a stochastic environment. We identify the actions with graphs from an action space $\mathcal{G}$ and assume that the size of this action space, denoted as $|\mathcal{G}|$, is finite. At time $t \in [T]$, the learner selects a graph G_t from the action space $\mathcal{G}_t \subset \mathcal{G}$. The learner then observes a noisy reward $y_t = \mu(G_t) + \varepsilon_t$ where $\mu : \mathcal{G} \to \mathbb{R}$ is the true (unknown) reward function and $\{\varepsilon_t\}_{t\in[T]}$ are i.i.d zero-mean sub-gaussian noise with variance proxy σ_ε^2. The goal of the learner is to maximize the expected cumulative reward in T rounds, which equivalently entails minimizing the expected (pseudo-)regret denoted as $R_T = \sum_{t=1}^{T} \mathbb{E}[\mu(G_t^*) - \mu(G_t)]$ where $G_t^* = \text{argmax}_{G\in\mathcal{G}_t} \mu(G)$ represents the optimal graph at time t.

The graph space $\mathcal{G}$ is a finite set of undirected graphs with at most N nodes. Note that the graphs with less than N nodes can be treated by adding auxiliary isolated nodes with no features. We denote an undirected attributed graph with N nodes as $G = (\mathbf{X}, \mathbf{A})$, where $\mathbf{X} \in \mathbb{R}^{N \times d}$ represents the feature matrix with d features, and $\mathbf{A} \in \{0,1\}^{N \times N}$ is the unweighted adjacency matrix. The rows of $\mathbf{X}$ correspond to node features. The size of the node set of a graph G is denoted as $|\mathcal{V}(G)| \leq N$.

Graph action bandit has several applications such as chemical molecules optimization. Consider the graph structures representing the molecules (Weininger, 1988) and rewards are molecular properties. The goal is to sequentially recommend the optimal molecules for experimental testing.

3.2 Graph Neural Network Model

We propose to learn the unknown reward function $\mu(\cdot)$ by fitting a Graph Neural Network (GNN). We consider a relatively simple GNN architecture where the output of a single graph convolution layer is normalized (to unit ℓ_2 norm) and passed through a multilayer perceptron (MLP). A single-layer graph convolution can be compactly stated as $\boldsymbol{AX}$ using the adjacency matrix $\boldsymbol{A}$ of the network. Additionally, we normalize each row of the resulting matrix to have a unit ℓ_2 norm. Letting $u(\boldsymbol{x}) = \boldsymbol{x}/\|\boldsymbol{x}\|_2$ denote the normalization operator, the aggregated feature of node i in a graph G is $\boldsymbol{h}_i^G = u((\boldsymbol{AX})_{i*}) = u(\sum_{j \in \mathcal{N}_i} \boldsymbol{X}_{j*})$ where $\mathcal{N}_j$ is the neighborhood of node j. Our GNN also consists of an L-layer m-width MLP neural network f_{MLP} which is defined recursively as follows

$$\begin{aligned} f^{(1)}(\boldsymbol{h}_i^G) &= \boldsymbol{W}^{(1)}\boldsymbol{h}_i^G, \quad i \in [N], \\ f^{(l)}(\boldsymbol{h}_i^G) &= \frac{1}{\sqrt{m}}\boldsymbol{W}^{(l)}\text{ReLU}(f^{(l-1)}(\boldsymbol{h}_i^G)), \quad 2 \leq l \leq L, \\ f_{\text{MLP}}(\boldsymbol{h}_i^G; \boldsymbol{\theta}) &= f^{(L)}(\boldsymbol{h}_i^G). \end{aligned} \tag{1}$$

Here, $\text{ReLU}(\cdot) = \max(\cdot, 0)$, $\boldsymbol{W}^{(1)} \in \mathbb{R}^{m \times d}$, $\boldsymbol{W}^{(L)} \in \mathbb{R}^{1 \times m}$, $\boldsymbol{W}^{(l)} \in \mathbb{R}^{m \times m}$ for any $1 < l < L$ are weight matrices of the MLP and $\boldsymbol{\theta} := (\boldsymbol{W}^{(1)}, \ldots, \boldsymbol{W}^{(L)}) \in \mathbb{R}^p$ is the collection of parameters of the neural network where $p = dm + (L-2)m^2 + m$. Our GNN model to estimate the reward function is

$$f_{\text{GNN}}(G; \boldsymbol{\theta}) := \frac{1}{N}\sum_{i=1}^{N} f_{\text{MLP}}(\boldsymbol{h}_i^G; \boldsymbol{\theta}). \tag{2}$$

The gradient of $\boldsymbol{\theta} \mapsto f_{\text{GNN}}(G; \boldsymbol{\theta})$ denoted as $g(G; \boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} f_{\text{GNN}}(G; \boldsymbol{\theta})$ will play a key role in uncertainty quantification, which will be discussed in Section 3.3. The GNN model (2) is trained by minimizing the mean-squared loss with ℓ_2 penalty, described concretely in (6). A hyperparameter λ is used to tune the strength of ℓ_2 regularization. For the simplicity of exposition, in the theoretical analysis, we solve the optimization via gradient descent with learning rate η, total number of iterations J and initialize parameters $\boldsymbol{\theta}_0$ such that $f_{\text{GNN}}(G; \boldsymbol{\theta}_0) = 0$ for all $G \in \mathcal{G}$, which can be fulfilled based on the work of Zhou et al. (2020); Kassraie and Krause (2022).

3.3 Graph Neural Thompson Sampling

We adapt Thompson Sampling (TS) for graph exploration, due to its robust performance in balancing exploration against exploitation. Algorithm 1 outlines our proposed GNN Thompson sampling, following the idea of `NeuralTS` in Zhang et al. (2020). The key step is the sampling of an estimated reward mean $\widehat{r}_t(G)$ for each graph G in the action space at time t, from a normal distribution as in equation (4). The mean of the normal distribution in (4) is our current estimate, $f_{\text{GNN}}(G; \boldsymbol{\theta}_{t-1})$, of the true mean reward for graph G (i.e., $\mu(G)$). This estimate is obtained by fitting the GNN to all the past data as in (6). The variance of the normal distribution $\nu^2\sigma_t^2(G)$ is our current measure of uncertainty about the true reward of graph G. Note that

$$\sigma_t^2(G) = \frac{1}{m}\|g(G; \boldsymbol{\theta}_{t-1})\|^2_{\boldsymbol{U}_{t-1}^{-1}} \quad \text{where} \quad \boldsymbol{U}_{t-1} = \lambda \boldsymbol{I}_p + \frac{1}{m}\sum_{i=1}^{t-1} g(G_i; \boldsymbol{\theta}_{i-1}) g(G_i; \boldsymbol{\theta}_{i-1})^\top. \tag{3}$$

Algorithm 1 Graph Neural Thompson Sampling (`GNN-TS`)

1: **Input:** T, λ, ν
2: Initialization: $\boldsymbol{\theta}_0$, $\boldsymbol{U}_0 = \lambda \boldsymbol{I}_p$.
3: **for** $t = 1, ..., T$ **do**
4: Compute $\sigma_t^2(G) := \frac{1}{m}\|g(G;\boldsymbol{\theta}_{t-1})\|^2_{\boldsymbol{U}_{t-1}^{-1}}$ and sample

$$\widehat{r}_t(G) \sim \mathcal{N}\big(f_{\text{GNN}}(G;\boldsymbol{\theta}_{t-1}), \nu^2\sigma_t^2(G)\big), \quad \text{for all } G \in \mathcal{G}_t. \tag{4}$$

5: Select graph $G_t = \text{argmax}_{G\in\mathcal{G}_t} \widehat{r}_t(G)$, and collect reward $y_t := \mu(G_t) + \varepsilon_t$.
6: Update uncertainty estimate as

$$\boldsymbol{U}_t = \boldsymbol{U}_{t-1} + g(G_t;\boldsymbol{\theta}_{t-1})g(G_t;\boldsymbol{\theta}_{t-1})^\top/m. \tag{5}$$

7: Update the parameter estimate as

$$\boldsymbol{\theta}_t = \underset{\boldsymbol{\theta}}{\text{argmin}} \frac{1}{2t}\sum_{i=1}^{t}\big(f_{\text{GNN}}(G_i;\boldsymbol{\theta}) - y_i\big)^2 + \frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2. \tag{6}$$

8: **end for**

The rationale behind $\sigma_t^2(G)$ comes from a linear approximation to $f_{\text{GNN}}(G;\boldsymbol{\theta})$. In particular, the idea is that (6) approximately looks like a linear ridge regression problem, with features $\{g(G_i;\boldsymbol{\theta}_i)/\sqrt{m}\}_{i\in[t]}$. The expression (3) is then the familiar estimated covariance matrix from linear bandits after we make this identification. This approximation can be made rigorous via the neural tangent kernel idea, as discussed in Section 4.

The sampled reward mean $\widehat{r}_t(G)$ is the index for decision-making. The learner selects the graph with the highest index, i.e., $G_t = \text{argmax}_{G\in\mathcal{G}} \widehat{r}_t(G)$. The randomness in $\widehat{r}_t(G)$, due to the positive variance of the sampling distribution, is what allows TS to efficiently explore the action space. We want the uncertainty, as captured by $\sigma_t^2(G)$ not to be too small early on, to allow for effective exploration, but not too large either to miss out on the optimal choice too often. Lemma 5.2 in Section 5 captures the two sides of this trade-off in our theory.

It is worth noting that our proposed Algorithm 1 is not exact TS. In our approach, (4) serves as an *approximation to* a posterior for mean reward function, rather than a true posterior. The difference between our proposed method and an exact Bayesian method will be smaller if the GNN model is better approximated by a linear model.

Lastly, we note that $\widehat{r}_t(G)$ is also referred to as the perturbed mean reward, as it can be expressed as: $\widehat{r}_t(G) = f_{\text{GNN}}(G;\boldsymbol{\theta}_{t-1}) + \nu\sigma_t(G)z$ where $z \sim \mathcal{N}(0,1)$. This perturbed reward includes both the estimated part ($f_{\text{gnn}}(G;\boldsymbol{\theta}_{t-1})$) and the random perturbation part ($\nu\sigma_t(G)\cdot z$). The use of perturbations for exploration has been shown to be a strong strategy in previous works (Kim and Tewari, 2019; Kveton et al., 2019a). Algorithm 1 can be summarized as greedily selecting the graph with the highest *perturbed* mean reward.

4 Regret Bound for `GNN-TS`

Graph Neural Tangent Kernel. Let us briefly review the idea of graph neural tangent kernel (GNTK) (Kassraie et al., 2022) which is based on the neural tangent kernel (NTK) of (Jacot et al., 2018). The tangent kernel on graph space $\mathcal{G}$, induced by initialization $\boldsymbol{\theta}_0$, is defined as the inner product of the gradient at initialization, i.e $\tilde{k}(G,G') := g(G;\boldsymbol{\theta}_0)^\top g(G';\boldsymbol{\theta}_0)$ for any $G, G' \in \mathcal{G}$. The GNTK is the limiting kernel of $\tilde{k}(G,G')/m$. We define the finite-width (empirical) and infinite-width GNTK as

$$\hat{k}(G,G') := \frac{1}{m}\langle g(G;\boldsymbol{\theta}_0), g(G';\boldsymbol{\theta}_0)\rangle, \quad k(G,G') := \lim_{m\to\infty}\frac{1}{m}\langle g(G;\boldsymbol{\theta}_0), g(G';\boldsymbol{\theta}_0)\rangle. \tag{7}$$

We assume the reward function falls within the RKHS corresponding to the GNTK k defined in (7). Define $\boldsymbol{K} \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$ as the GNTK matrix with entries $k(G,G')$ for all $G, G' \in \mathcal{G}$ and

$\boldsymbol{\mu} = (\mu(G))_{G\in\mathcal{G}} \in \mathbb{R}^{|\mathcal{G}|}$ as the reward function vector. The kernel matrix $\boldsymbol{K}$ is positive definite with maximum eigenvalue $\rho_{\max} := \lambda_{max}(\boldsymbol{K})$ and minimum eigenvalue $\rho_{min} := \lambda_{min}(\boldsymbol{K})$. We also define the finite-width GNTK matrix $\hat{\boldsymbol{K}} \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$ with entries $\hat{k}(G, G')$ for all $G, G' \in \mathcal{G}$ and maximum eigenvalues $\hat{\rho}_{\max} = \lambda_{\max}(\hat{\boldsymbol{K}})$. Note that $\hat{\boldsymbol{K}} \to \boldsymbol{K}$ as $m \to \infty$.

Effective Dimension. We define the effective dimension $\tilde{d}$ of the GNTK matrix $\boldsymbol{K}$ with regularization λ as

$$\tilde{d} := \frac{\log\det(\boldsymbol{I}_{|\mathcal{G}|} + T\boldsymbol{K}/\lambda)}{\log(1 + T\rho_{\max}/\lambda)}. \tag{8}$$

This quantity, which appears in our regret bound, measures the actual underlying dimension of the reward function space as seen by the bandit problem (Valko et al., 2013; Bietti and Mairal, 2019). Our definition is adapted from (Yang and Wang, 2020). The key difference is that our $\tilde{d}$ does not directly depend on $|\mathcal{G}|$, which is replaced by $\rho_{\max}$, compared to the definition in (Zhang et al., 2020). Our definition is the ratio of the sum over the maximum of the sequence of log-eigenvalues of matrix $\boldsymbol{I}_{|\mathcal{G}|} + T\boldsymbol{K}/\lambda$. As such, it is a robust measure of matrix rank. In particular, we always have $\tilde{d} \le |\mathcal{G}|$. Moreover, previous work on GNN bandit (Kassraie et al., 2022) utilized the notion of information gain which we replace with the related, but different, notion of effective dimension $\tilde{d}$.

We will make the following assumptions:

Assumption 1 (Bounded RKHS norm for Reward)**.** *The reward function μ has R-bounded RKHS norm with respect to a positive definite kernel k: $\|\mu\|_k = \sqrt{\boldsymbol{\mu}^\top \boldsymbol{K}^{-1}\boldsymbol{\mu}} \le R$.*

Assumption 2 (Bounded Reward Differences)**.** *Reward differences between any graph in action space are bounded. Formally, $\forall G, G' \in \mathcal{G}$: $|\mu(G) - \mu(G')| \le B$, for some $B \ge 1$.*

Assumption 3 (Subgaussian Noise)**.** *Noise process $\{\epsilon_t\}_{t\in[T]}$ satisfies $\mathbb{E}_{t-1}[e^{\eta\epsilon_t}] \le e^{\sigma_\epsilon^2\eta^2/2}, \forall\eta > 0$.*

Assumption 1 aligns with the regularity assumption commonly found in the kernelized and neural bandit literature (Srinivas et al., 2009; Chowdhury and Gopalan, 2017; Kassraie and Krause, 2022). Assumption 2 implies that instantaneous regret is bounded: $|\mu(G_t^*) - \mu(G_t)| \le B$ for all $t \in [T]$ and Assumption 3 is the conditional subgaussian assumption for stochastic process $\{\epsilon_t\}_{t\in[T]}$.

We are now ready to state our main result. Recall that N is the maximum number of (graph) nodes and L the depth of MLP and m its width.

Theorem 4.1. *Suppose Assumption 1,2 and 3 hold. For a fixed horizon $T \in \mathbb{N}$, let*

$$m \ge \mathrm{poly}\big(T, L, |\mathcal{G}|, \lambda^{-1}, R, \sigma_\varepsilon, \rho_{\min}^{-1}, \log(TLN|\mathcal{G}|)\big)$$
$$\nu \gtrsim 1 + \sigma_\varepsilon\sqrt{\tilde{d}\log T} + \sqrt{\lambda}R, \quad \lambda \gtrsim (\sigma_\varepsilon^2 + R^2)^3 + \rho_{\max}$$

and learning rate $\eta \le (\tilde{C}mL + m\lambda)^{-1}$, for some constant $\tilde{C}$. Then, the regret of Algorithm 1 is bounded as

$$R_T \le C\,B\sqrt{\tilde{d}T\log(T|\mathcal{G}|)\cdot\log(2 + T\rho_{\max}/\lambda)}$$

for some universal constant $C > 0$.

The order of regret upper bound in Theorem 4.1, $\tilde{\mathcal{O}}((\tilde{d}T)^{1/2})$ matches the state-of-the-art regret bounds in the literature of Thompson Sampling (Agrawal and Goyal, 2013; Chowdhury and Gopalan, 2017; Kveton et al., 2020; Zhang et al., 2020). As in (Kassraie et al., 2022), our regret bound is independent of N, indicating that `GNN-TS` is valid for large graphs. Moreover, for low complexity reward functions of effective dimension $\tilde{d} = O(1)$, the regret scales as $\sqrt{\log|\mathcal{G}|}$ in the size of the action space, showing the robust scalability of `GNN-TS`.

5 Proof of the Regret Bound

Similar to the previous literature, the key is to to obtain probabilistic control on the 'discrepancy' of the policy in `GNN-TS` Consider the following events

$$\mathcal{E}_t^\mu := \Big\{ \big| f_{\text{GNN}}(G; \boldsymbol{\theta}_{t-1}) - \mu(G) \big| \ \leq \ c_t^\mu(G), \text{ for all } G \in \mathcal{G}_t \Big\}$$
$$\mathcal{E}_t^\sigma := \Big\{ \big| \widehat{r}_t(G) - f_{\text{GNN}}(G; \boldsymbol{\theta}_{t-1}) \big| \ \leq \ c_t^\sigma(G), \text{ for all } G \in \mathcal{G}_t \Big\}$$
$$\mathcal{E}_t^a := \Big\{ \widehat{r}_t(G_t^*) - f_{\text{GNN}}(G_t^*; \boldsymbol{\theta}_{t-1}) \ > \ \nu \sigma_t(G_t^*) \Big\}$$

where $c_t^\mu(G) := \nu\sigma_t(G) + \varepsilon(t,m)$ and $c_t^\sigma(G) := \nu\sigma_t(G)\sqrt{2\log(t^2|\mathcal{G}_t|)}$ as well as $\varepsilon(t,m) = (C_0 \nu L^{9/2})\, m^{-1/6}\sqrt{\log m} \cdot t$ and C_0 is some universal constant. Events $\mathcal{E}_t^\mu$ and $\mathcal{E}_t^\sigma$ control the discrepancies with constants $c_t^\mu(G)$ and $c_t^\sigma(G)$ respectively: $c_t^\mu(G)$ is bounding the estimation discrepancy while $c_t^\sigma(G)$ is bounding the exploration discrepancy. Note that event $\mathcal{E}_t^a$ is only for G_t^*, the optimal graph at round t.

5.1 Estimation Bound ($\mathcal{E}_t^\mu$)

The following lemma ensures that event $\mathcal{E}_t^\mu$ happens with high probability.

Lemma 5.1. *Fix $\delta \in (0,1)$. For $m \geq \text{poly}(R, \sigma_\varepsilon, L, |\mathcal{G}|, \lambda^{-1}, \rho_{\min}^{-1}, \log(TLN|\mathcal{G}|/\delta))$ and (ν, λ, η) satisfying conditions of Theorem 4.1, we have $\mathbb{P}(\mathcal{E}_t^\mu) \geq 1 - \delta/T$.*

In other words, given a large enough width of the GNN (m) and a small enough learning rate (η), there is a high probability upper bound for the estimation error $|f_{\text{GNN}}(G; \boldsymbol{\theta}_{t-1}) - \mu(G)|$. This Lemma 5.1 also gives an approximate upper confidence bound similar to `GNN-UCB` in (Kassraie et al., 2022): $\mu(G) \leq f_{\text{GNN}}(G; \boldsymbol{\theta}_{t-1}) + \nu\sigma_t(G) + \varepsilon(t,m)$. Since $\varepsilon(t,m)$ is negligible for large m, the approximate upper confidence bound, $f_{\text{GNN}}(G; \boldsymbol{\theta}_{t-1}) + \nu\sigma_t(G)$ is used as the index for `GNN-UCB`. Note that this lemma controls the estimation error produced by GNNs, hence applicable to all GNN bandit algorithms using model (2). Our $c_t^\mu(G) = \nu\sigma_t(G) + \varepsilon(t,m)$ is similar in form to that of Zhang et al. (2020) which is different from the earlier analysis of TS in Agrawal and Goyal (2013).

5.2 Exploration Bound ($\mathcal{E}_t^\sigma, \mathcal{E}_t^a$)

We also need event $\mathcal{E}_t^\sigma$ to quantify the level of exploration achieved by the algorithm. Intuitively, $\mathcal{E}_t^\sigma$ ensures our exploration is moderate. On the other hand, indicated by the regret analysis in (Kveton et al., 2019b), instead of controlling the exploration independently, the relation between two sources of explorations needs to be considered because this relation is critical for finding the optimal action. To meet such observation, we define an extra "good" event for anti-concentration on the optimal actions, which is $\mathcal{E}_t^a$. Under event $\mathcal{E}_t^a$, the policy index $\widehat{r}_t(G_t^*)$ of the optimal graph has the higher future positive exploration, which guides the learner to have higher chance to pick the optimal graph. A formal lemma for exploration discrepancy using TS is given as below:

Lemma 5.2. *For `GNN-TS`, for all $t \in [T]$, we have $\mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma) \leq t^{-2}$ and $\mathbb{P}(\mathcal{E}_t^a) \geq (4e\sqrt{\pi})^{-1}$.*

Lemma 5.2 shows that `GNN-TS` has a positive probability of moderate exploration of the optimal arm, which is beneficial to regret reduction.

5.3 Proof of Theorem 4.1

Let $\Delta_t := \mu(G_t^*) - \mu(G_t)$ be the instantaneous regret. We will need two additional lemmas:

Lemma 5.3 (One Step Regret Bound)**.** *Assume the same as Theorem 4.1. Suppose $\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma) > 0$. Then for any $t \in [T]$, almost surely,*

$$\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{E}_t^\mu}] \leq \mathbb{I}_{\mathcal{E}_t^\mu} \cdot \Big\{ \Big(\frac{2}{\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma)} + 1 \Big) \mathbb{E}_t[\gamma_t(G_t)] - \varepsilon(t,m) + B \cdot \mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma) \Big\}$$

where $\gamma_t(G) = c_t^\mu(G) + c_t^\sigma(G)$.

Lemma 5.4 (Cumulative Uncertainty Bound)**.** *Assume the same as Theorem 4.1. Then with probability at least* $1-\delta/T$*,*

$$\frac{1}{2}\sum_{t=1}^{T}\min\{1,\sigma_t^2(G_t)\} \le \tilde{d}\log(1+\lambda^{-1}T\rho_{\max}) + 3C_\psi|\mathcal{G}|^{3/2}\sqrt{T}\lambda^{-1/2}\varepsilon_m$$

where $\varepsilon_m = o(1)$ *as* $m\to\infty$ *and* C_ψ *is some constant. We always have* $\tilde{d}\le|\mathcal{G}|$*.*

Main Proof. The expected cumulative regret is

$$R_T = \sum_{t=1}^{T}\mathbb{E}[\Delta_t] = \sum_{t=1}^{T}\mathbb{E}[\Delta_t\mathbb{I}_{\mathcal{E}_t^\mu}] + \sum_{t=1}^{T}\mathbb{E}[\Delta_t\mathbb{I}_{\bar{\mathcal{E}}_t^\mu}].$$

By Lemma 5.1, letting $\mathbb{P}(\bar{\mathcal{E}}_t^\mu)\le\delta/T$ and $\Delta_t\le B$, we have the upper bound for the second term

$$\sum_{t=1}^{T}\mathbb{E}[\Delta_t\mathbb{I}_{\bar{\mathcal{E}}_t^\mu}] \le BT(\delta/T) = B\delta.$$

Now our focus is controlling the first summation term. By Lemma 5.3, almost surely, we have

$$\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{E}_t^\mu}] \le \mathbb{I}_{\mathcal{E}_t^\mu}\cdot\Big\{\Big(\frac{2}{\mathbb{P}_t(\mathcal{E}_t^a)-\mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma)}+1\Big)\mathbb{E}_t[\gamma_t(G_t)] - \varepsilon(t,m) + B\cdot\mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma)\Big\}$$

where $\gamma_t(G) = c_t^\mu(G)+c_t^\sigma(G)$. Assuming that $t\ge 5$, we have $t^2\ge 5e\sqrt{\pi}$. By Lemma 5.2, $\mathbb{P}_t(\mathcal{E}_t^a)-\mathbb{P}(\bar{\mathcal{E}}_t^\sigma)\ge\frac{1}{4e\sqrt{\pi}}-\frac{1}{t^2}\ge\frac{1}{20e\sqrt{\pi}}$. Then, for $t\ge 5$, dropping $\varepsilon(t,m)$ from the bound,

$$\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{E}_t^\mu}] \le 194\mathbb{E}_t[\gamma_t(G_t)] + Bt^{-2} \le \big(194\mathbb{E}_t[\min\{1,\gamma_t(G_t)\}] + t^{-2}\big)B$$

using $40e\sqrt{\pi}+1\le 194$, $\Delta_t\le B$ and $B\ge 1$. Therefore, we have

$$\sum_{t=1}^{T}\mathbb{E}[\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{E}_t^\mu}]] \le 194B\sum_{t=5}^{T}\mathbb{E}[\mathbb{E}_t[\min\{1,\gamma_t(G_t)\}]] + 4B + B(\pi^2/6) \tag{9}$$

using $\sum_{t=1}^{\infty}t^{-2}=\pi^2/6$. Note that $\gamma_t(G_t)\le\sigma_t(G_t)\sqrt{8\log(T^2|\mathcal{G}|)}+\varepsilon(T,m)$ for all $t\in[T]$. Then by Cauchy-Schwarz inequality,

$$\sum_{t=5}^{T}\min\{1,\gamma_t(G_t)\} \le \sqrt{8T\log(T^2|\mathcal{G}|)}\Big(\sum_{t=5}^{T}\min\{1,\sigma_t^2(G_t)\}\Big)^{1/2} + T\varepsilon(T,m).$$

By Lemma 5.4 and take m sufficiently large such that $3C_\psi|\mathcal{G}|^{3/2}\sqrt{T}\lambda^{-1/2}\varepsilon_m\le\tilde{d}\log(1+\lambda^{-1}T\rho_{\max})$, we have

$$\sum_{t=1}^{T}\mathbb{E}[\min\{1,\sigma_t^2(G_t)\}] \le 4\tilde{d}\log(1+T\rho_{\max}/\lambda) + T(\delta/T).$$

Recall that the $\varepsilon(T,m)=C_1\,T\,m^{-1/6}\sqrt{\log m}$. Take m large enough we have $T\varepsilon(T,m)\le\sqrt{T}$. Then put the above results back into (9), we have:

$$\sum_{t=1}^{T}\mathbb{E}[\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{E}_t^\mu}]] \le 194B\big(\sqrt{16T\log(T|\mathcal{G}|)}\cdot\sqrt{4\tilde{d}\log(1+T\rho_{\max}/\lambda)+\delta}+\sqrt{T}\big) + 4B + B(\pi^2/6)$$

by using $\log(T^2|\mathcal{G}|)\le 2\log(T|\mathcal{G}|)$. Therefore, we have our regret bound:

$$R_T \le CB\sqrt{\tilde{d}T\log(T|\mathcal{G}|)\cdot\big(1+\log(1+T\rho_{\max}/\lambda)\big)}$$

for some universal constant C. We have used $\tilde{d}\ge 1$ and $B\ge 1$, to simplify the bound. Finally, note that $1+\log(1+x)\le 2\log(2+x)$ for all $x\ge 0$. □

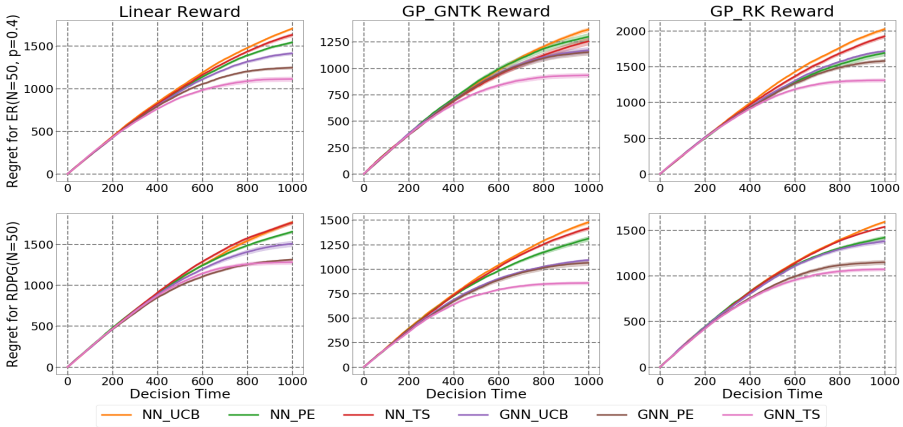

Figure 1: Regret over horizon $T = 1000$ for Erdös–Rényi random graphs with $p = 0.4$ and $N = 50$ in the first row and random dot product graphs with $N = 50$. Three columns are three types of reward function generation: linear model, Gaussian process with GNTK, Gaussian process with representation kernel. `GNN-TS` is competitive and robust to different environment settings.

6 Experiments

We create synthetic graph data and generate the rewards through three different mechanisms. For the graph structures, we use random graph models including Erdös–Rényi and random dot product graph models. The features are generated i.i.d. from the $\mathcal{N}(0, 1)$. The noisy reward is assumed to have $\sigma_\varepsilon = 0.01$. Our experiments investigate `GNN-UCB`, `GNN-PE`, `NN-UCB`, `NN-PE`, and `NN-TS` as baselines from Kassraie et al. (2022). All performance curves in our empirical studies show an average of over 10 repetitions with a standard deviation of the corresponding bandit algorithm with horizon $T = 1000$. We assume the graph domain is fully observable, $\mathcal{G}_t = \mathcal{G}$ for all $t \in [T]$. Below is a brief overview of the simulation elements. For more details, see Appendix D.

Random Graph. We use two types of random graphs including Erdös–Rényi (ER) random graphs and random dot product graphs (RDPG). ER graphs are generated with edge probability p and number of nodes N. RDPGs are generated by modeling the expected edge probabilities as the function of the inner product of features. In the first row of Figure 1, the graphs in $\mathcal{G}$ are from the ER model with $p = 0.4$ and in the second row from an RDPG, both of size $N = 50$.

Reward Function. To generate the rewards, we use models of three different types: linear model, Gaussian Process (GP) with GNTK, Gaussian process with the representation kernel. For the linear model, we have $\mu(G) = \langle \boldsymbol{\theta}^*, \bar{\boldsymbol{h}}^G \rangle$ with true parameter $\boldsymbol{\theta}^* \sim \mathcal{N}(0, \boldsymbol{I}_d)$ and $\bar{\boldsymbol{h}}^G = \sum_{i=1}^N \boldsymbol{h}_i^G / N$. For the GP with GNTK, we fit a GP regression model with empirical GNTK matrix $\hat{\boldsymbol{K}} \in \mathbb{R}^{|\mathcal{G}| \times |\mathcal{G}|}$ as the covariance matrix of the prior, trained on $\{(G, y_G)\}_{G \in \mathcal{G}}$ where $\{y_G\}_{G \in \mathcal{G}}$ are i.i.d. from $\mathcal{N}(0, 1)$. For the GP with the representation kernel, we trained a GNN for a graph property prediction task and used the mean pooling over all the nodes of the last layer representations as the graph representation, denoted as $\bar{\boldsymbol{h}}_{\text{rep}}^G$. We then define the representation kernel as $k_{\text{rep}}(G, G') := \langle \bar{\boldsymbol{h}}_{\text{rep}}^G, \bar{\boldsymbol{h}}_{\text{rep}}^{G'} \rangle$ and draw $\mu(\cdot)$ from a zero-mean GP with this covariance function (over $\mathcal{G}$).

Algorithms. We investigate two baselines `GNN-UCB` and `GNN-PE` along with our proposed `GNN-TS`. `GNN-PE` is the proposed state-of-the-art algorithm that selects the graph with the highest uncertainty and eliminates the graph candidates by the upper confidence bounds. All the algorithms in our work use the loss function (6) which is different from previous work. All gradients used for our experiments are $g(G; \boldsymbol{\theta}_t)$, not $g(G; \boldsymbol{\theta}_0)$, unless otherwise specified. In addition, in order to show the benefit of considering the graph structure, we include `NN-UCB`, `NN-TS`, and `NN-PE` as our baselines. For these NN-based algorithms, we ignore the adjacency matrix of a graph (setting $\boldsymbol{A} = \boldsymbol{I}_N$), and pass through

the model in (1) and (2) with $\boldsymbol{h}_i^G = \boldsymbol{X}_{i*}$. The MLPs in our experiments have $L = 2$ layers and width $m = 512$. We use SGD as the optimizer, with mini-batch size 5, and train for 30 epochs. For the tuning of the hyperparameters (η, λ) and other algorithmic setup, see Appendix D. The matrix inversion in the algorithms is approximated by diagonal inversion across all policy algorithms.

Regret Experiments. In Figure 1, we show the performance of all the algorithms for the six possible environments: ER or RDPG model coupled with either of the three reward models. We set the size of the graph domain to $|\mathcal{G}| = 100$ in Figure 1 and we experiment across different $|\mathcal{G}|$ in Appendix D. Figure 1 demonstrates that `GNN-TS` consistently outperforms the baseline algorithms and is robust to all types of random graph models and reward function generations in our experiment. In addition, GNN-based algorithms are clearly better than NN-based algorithms in graph action bandit settings.

References

Marc Abeille and Alessandro Lazaric. Linear thompson sampling revisited. In *Artificial Intelligence and Statistics*, pages 176–184. PMLR, 2017.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pages 127–135. PMLR, 2013.

Sanjeev Arora, Simon S Du, Wei Hu, Zhiyuan Li, Russ R Salakhutdinov, and Ruosong Wang. On exact computation with an infinitely wide neural net. *Advances in neural information processing systems*, 32, 2019.

Marc Bellemare, Will Dabney, Robert Dadashi, Adrien Ali Taiga, Pablo Samuel Castro, Nicolas Le Roux, Dale Schuurmans, Tor Lattimore, and Clare Lyle. A geometric perspective on optimal representations for reinforcement learning. *Advances in neural information processing systems*, 32, 2019.

Alberto Bietti and Julien Mairal. On the inductive bias of neural tangent kernels. *Advances in Neural Information Processing Systems*, 32, 2019.

Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, 24, 2011.

Houshuang Chen, Shuai Li, Chihao Zhang, et al. Understanding bandits with graph feedback. *Advances in Neural Information Processing Systems*, 34:24659–24669, 2021.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, pages 844–853. PMLR, 2017.

Zhongxiang Dai, Yao Shu, Bryan Kian Hsiang Low, and Patrick Jaillet. Sample-then-optimize batch neural thompson sampling. *Advances in Neural Information Processing Systems*, 35:23331–23344, 2022.

Christoph Dann, Yishay Mansour, Mehryar Mohri, Ayush Sekhari, and Karthik Sridharan. Reinforcement learning with feedback graphs. *Advances in Neural Information Processing Systems*, 33: 16868–16878, 2020.

Emmanuel Esposito, Federico Fusco, Dirk van der Hoeven, and Nicolò Cesa-Bianchi. Learning on the edge: Online learning with stochastic feedback graphs. *Advances in Neural Information Processing Systems*, 35:34776–34788, 2022.

Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

Yutian Gou, Jinfeng Yi, and Lijun Zhang. Stochastic graphical bandits with heavy-tailed rewards. In *Uncertainty in Artificial Intelligence*, pages 734–744. PMLR, 2023.

Ryan-Rhys Griffiths and José Miguel Hernández-Lobato. Constrained bayesian optimization for automatic chemical design using variational autoencoders. *Chemical science*, 11(2):577–586, 2020.

Quanquan Gu, Amin Karbasi, Khashayar Khosravi, Vahab Mirrokni, and Dongruo Zhou. Batched neural bandits. *ACM/IMS Journal of Data Science*, 1(1):1–18, 2024.

Tsuyoshi Idé, Keerthiram Murugesan, Djallel Bouneffouf, and Naoki Abe. Targeted advertising on social networks using online variational tensor regression. *arXiv preprint arXiv:2208.10627*, 2022.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, 31, 2018.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *International conference on machine learning*, pages 2323–2332. PMLR, 2018.

Parnian Kassraie and Andreas Krause. Neural contextual bandits without regret. In *International Conference on Artificial Intelligence and Statistics*, pages 240–278. PMLR, 2022.

Parnian Kassraie, Andreas Krause, and Ilija Bogunovic. Graph neural network bandits. *Advances in Neural Information Processing Systems*, 35:34519–34531, 2022.

Jaya Kawale, Hung H Bui, Branislav Kveton, Long Tran-Thanh, and Sanjay Chawla. Efficient thompson sampling for online matrix-factorization recommendation. *Advances in neural information processing systems*, 28, 2015.

Baekjin Kim and Ambuj Tewari. On the optimality of perturbations in stochastic and adversarial multi-armed bandit problems. *Advances in Neural Information Processing Systems*, 32, 2019.

Tomáš Kocák, Michal Valko, Rémi Munos, and Shipra Agrawal. Spectral thompson sampling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

Tomáš Kocák, Rémi Munos, Branislav Kveton, Shipra Agrawal, and Michal Valko. Spectral bandits. *The Journal of Machine Learning Research*, 21(1):9003–9046, 2020.

Fang Kong, Yichi Zhou, and Shuai Li. Simultaneously learning stochastic and adversarial bandits with general graph feedback. In *International Conference on Machine Learning*, pages 11473–11482. PMLR, 2022.

Ksenia Korovina, Sailun Xu, Kirthevasan Kandasamy, Willie Neiswanger, Barnabas Poczos, Jeff Schneider, and Eric Xing. Chembo: Bayesian optimization of small organic molecules with synthesizable recommendations. In *International Conference on Artificial Intelligence and Statistics*, pages 3393–3403. PMLR, 2020.

Branislav Kveton, Csaba Szepesvari, Mohammad Ghavamzadeh, and Craig Boutilier. Perturbed-history exploration in stochastic multi-armed bandits. *arXiv preprint arXiv:1902.10089*, 2019a.

Branislav Kveton, Csaba Szepesvari, Mohammad Ghavamzadeh, and Craig Boutilier. Perturbed-history exploration in stochastic linear bandits. *arXiv preprint arXiv:1903.09132*, 2019b.

Branislav Kveton, Manzil Zaheer, Csaba Szepesvari, Lihong Li, Mohammad Ghavamzadeh, and Craig Boutilier. Randomized exploration in generalized linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2066–2076. PMLR, 2020.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Chung-Wei Lee, Haipeng Luo, and Mengxiao Zhang. A closer look at small-loss bounds for bandits with graph feedback. In *Conference on Learning Theory*, pages 2516–2564. PMLR, 2020.

Fang Liu, Zizhan Zheng, and Ness Shroff. Analysis of thompson sampling for graphical bandits without the graphs. *arXiv preprint arXiv:1805.08930*, 2018.

Thodoris Lykouris, Eva Tardos, and Drishti Wali. Feedback graph regret bounds for thompson sampling and ucb. In *Algorithmic Learning Theory*, pages 592–614. PMLR, 2020.

Thanh Nguyen-Tang, Sunil Gupta, A Tuan Nguyen, and Svetha Venkatesh. Offline neural contextual bandits: Pessimism, optimization and generalization. *arXiv preprint arXiv:2111.13807*, 2021.

Yunzhe Qi, Yikun Ban, and Jingrui He. Neural bandit with arm group graph. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1379–1389, 2022.

Yunzhe Qi, Yikun Ban, and Jingrui He. Graph neural bandits. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1920–1931, 2023.

Carlos Riquelme, George Tucker, and Jasper Snoek. Deep bayesian bandits showdown: An empirical comparison of bayesian deep networks for thompson sampling. *arXiv preprint arXiv:1802.09127*, 2018.

Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen, et al. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, 11(1):1–96, 2018.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

Parth Thaker, Mohit Malu, Nikhil Rao, and Gautam Dasarathy. Maximizing and satisficing in multi-armed bandits with graph information. *Advances in Neural Information Processing Systems*, 35:2019–2032, 2022.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

Laura Toni and Pascal Frossard. Online network source optimization with graph-kernel mab. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 242–258. Springer, 2023.

Aristide Tossou, Christos Dimitrakakis, and Devdatt Dubhashi. Thompson sampling for stochastic bandits with graph feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Sohini Upadhyay, Mikhail Yurochkin, Mayank Agarwal, Yasaman Khazaeni, and Djallel Bouneffouf. Graph convolutional network upper confident bound. 2020.

Sattar Vakili, Nacime Bouziani, Sepehr Jalali, Alberto Bernacchia, and Da-shan Shiu. Optimal order simple regret for gaussian process bandits. *Advances in Neural Information Processing Systems*, 34:21202–21215, 2021.

Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nelo Cristianini. Finite-time analysis of kernelised contextual bandits. *arXiv preprint arXiv:1309.6869*, 2013.

Miles Wang-Henderson, Bartu Soyuer, Parnian Kassraie, Andreas Krause, and Ilija Bogunovic. Graph neural network powered bayesian optimization for large molecular spaces. In *ICML 2023 Workshop on Structured Probabilistic Inference {\&} Generative Modeling*, 2023a.

Miles Wang-Henderson, Bartu Soyuer, Parnian Kassraie, Andreas Krause, and Ilija Bogunovic. Graph neural bayesian optimization for virtual screening. In *NeurIPS 2023 Workshop on Adaptive Experimental Design and Active Learning in the Real World*, 2023b.

Vikram Waradpande, Daniel Kudenko, and Megha Khosla. Deep reinforcement learning with graph-based state representations. *arXiv preprint arXiv:2004.13965*, 2020.

David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

Shuang Wu, Chi-Hua Wang, Yuantong Li, and Guang Cheng. Residual bootstrap exploration for stochastic linear bandit. In *Uncertainty in Artificial Intelligence*, pages 2117–2127. PMLR, 2022.

Pan Xu, Zheng Wen, Handong Zhao, and Quanquan Gu. Neural contextual bandits with deep representation and shallow exploration. *arXiv preprint arXiv:2012.01780*, 2020.

Kaige Yang, Laura Toni, and Xiaowen Dong. Laplacian-regularized graph bandits: Algorithms and theoretical analysis. In *International Conference on Artificial Intelligence and Statistics*, pages 3133–3143. PMLR, 2020.

Lin Yang and Mengdi Wang. Reinforcement learning in feature space: Matrix bandit, kernels, and regret bound. In *International Conference on Machine Learning*, pages 10746–10756. PMLR, 2020.

Tong Yu, Branislav Kveton, Zheng Wen, Ruiyi Zhang, and Ole J Mengshoel. Graphical models meet bandits: A variational thompson sampling approach. In *International Conference on Machine Learning*, pages 10902–10912. PMLR, 2020.

Tom Zahavy and Shie Mannor. Deep neural linear bandits: Overcoming catastrophic forgetting through likelihood matching. *arXiv preprint arXiv:1901.08612*, 2019.

Weitong Zhang, Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural thompson sampling. *arXiv preprint arXiv:2010.00827*, 2020.

Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural contextual bandits with ucb-based exploration. In *International Conference on Machine Learning*, pages 11492–11502. PMLR, 2020.

A Proof for Lemmas in Regret Analysis

A.1 Notations

In the following parts, we further define the some notations to represent the linear and kernelized models:

$$\begin{aligned}
\boldsymbol{G}_t &= [g(G_1;\boldsymbol{\theta}_0),...,g(G_t;\boldsymbol{\theta}_{t-1}))] \in \mathbb{R}^{p\times t}\\
\bar{\boldsymbol{G}}_t &= [g(G_1;\boldsymbol{\theta}_0),...,g(G_t;\boldsymbol{\theta}_0)] \in \mathbb{R}^{p\times t}\\
\boldsymbol{\mu}_t &= [\mu(G_1),...,\mu(G_t)]^\top \in \mathbb{R}^{t\times 1}\\
\boldsymbol{y}_t &= [y_1,...,y_t]^\top \in \mathbb{R}^{t\times 1}\\
\boldsymbol{\epsilon}_t &= [\varepsilon_1,...,\varepsilon_t]^\top \in \mathbb{R}^{t\times 1}.
\end{aligned}$$

Then we define the uncertainty estimate with initial gradient $\boldsymbol{\theta}_0$:

$$\bar{\sigma}_t^2(G) = \frac{1}{m}\|g(G;\boldsymbol{\theta}_0)\|^2_{\bar{\boldsymbol{U}}_{t-1}^{-1}} \quad \text{and} \quad \bar{\boldsymbol{U}}_t = \lambda \boldsymbol{I}_p + \sum_{i=1}^{t} g(G_i;\boldsymbol{\theta}_0)g(G_i;\boldsymbol{\theta}_0)^\top/m.$$

A.2 Proof of Lemma 5.1

Let us write

$$\widetilde{\boldsymbol{\theta}}_{t-1} := \bar{\boldsymbol{U}}_{t-1}^{-1}\bar{\boldsymbol{G}}_{t-1}\boldsymbol{y}_{t-1}/m$$

for the ridge regression solution. We will need the following auxiliary lemmas:

Lemma A.1 (Taylor Approximation of a GNN)**.** *Suppose learning rate $\eta \le (\tilde{C}mL + m\lambda)^{-1}$ for some constant $\tilde{C}$, then for any fixed $t \in [T]$ and $G \in \mathcal{G}$, with probability at least $1-\delta$*

$$|f_{\text{GNN}}(G;\boldsymbol{\theta}_t^{(J)}) - f_{\text{GNN}}(G;\boldsymbol{\theta}_0) - \langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}_t^{(J)} - \boldsymbol{\theta}_0\rangle| \le CL^3\Big(\frac{R^2+\sigma_\varepsilon^2}{m\lambda}\Big)^{2/3}\sqrt{m\log(m)}$$

where C is some constant independent of m and t.

Lemma A.2. *Suppose $m \ge poly(R, \sigma_\varepsilon, L, \lambda^{-1}, |\mathcal{G}|, \rho_{min}^{-1}, \log(LN|\mathcal{G}|/\delta))$ given a fixed $\delta \in (0,1)$ and learning rate $\eta \le (\tilde{C}mL + m\lambda)^{-1}$ for some constant $\tilde{C}$. For $G \in \mathcal{G}_t$ and $t > 1$, with probability at lease $1-\delta$,*

$$|\langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_{t-1}\rangle| \le C\bar{\sigma}_t(G)$$

where $C = (C_1(2-\eta m\lambda)^J + C_2)\sqrt{\frac{\sigma_\varepsilon^2+R^2}{\lambda}(1+\frac{3\rho_{\max}}{2\lambda})}$ with $C_1 = \mathcal{O}(1)$ and $C_2 = \mathcal{O}(\lambda^{1/3})$.

Lemma A.3. *Fix $\delta \in (0,1)$ and let $m = \Omega(L^{10}T^4|\mathcal{G}|^6\rho_{min}^{-4}\log(LN^2|\mathcal{G}|^2/\delta))$. Then, there exists $\boldsymbol{\theta}^* \in \mathbb{R}^p$ with $\sqrt{m}\|\boldsymbol{\theta}^*\|_2 \le \sqrt{2}R$ such that with probability at least $1-\delta$,*

$$\begin{aligned}
\mu(G) &= \langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}^*\rangle, \quad \textit{for all } G \in \mathcal{G}\\
\log\det(\lambda^{-1}\bar{\boldsymbol{U}}_t) &\le \log\det(\boldsymbol{I}_{|\mathcal{G}|} + \lambda^{-1}t\boldsymbol{K}) + 1.
\end{aligned}$$

Lemma A.4. *With probability at least $1-\delta$, we have*

$$|\bar{\sigma}_t(G) - \sigma_t(G)| \le Ct\lambda^{-1/6}L^{9/2}(R^2+\sigma_\varepsilon^2)^{1/6}m^{-1/6}\sqrt{\log(m)}.$$

We choose an arbitrary small $\delta \in (0,1)$ and set $\delta_i = \delta/(5T)$ for $i = 1,\ldots,5$. For all $\forall G \in \mathcal{G}_t$, we have

$$|f_{\text{GNN}}(G;\boldsymbol{\theta}_{t-1}) - \mu(G)| \le \underbrace{|f_{\text{GNN}}(G;\boldsymbol{\theta}_{t-1}) - \langle g(G;\boldsymbol{\theta}_0), \widetilde{\boldsymbol{\theta}}_{t-1}\rangle|}_{:=I_1} + \underbrace{|\mu(G) - \langle g(G;\boldsymbol{\theta}_0), \widetilde{\boldsymbol{\theta}}_{t-1}\rangle|}_{:=I_2}.$$

We then turn to bounding I_1 and I_2. Throughout the proof, let

$$\gamma_m := m^{-1/6}\sqrt{\log m}$$

Bounding I_1: By Lemma A.1 and Lemma A.2, with probability at least $1-\delta_1-\delta_2$,

$$\begin{aligned} I_1 &= |f_{\text{GNN}}(G;\boldsymbol{\theta}_{t-1}) - \langle g(G;\boldsymbol{\theta}_0), \widetilde{\boldsymbol{\theta}}_{t-1}\rangle| \\ &\le |f_{\text{GNN}}(G;\boldsymbol{\theta}_{t-1}) - \langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_0\rangle| + |\langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_0-\widetilde{\boldsymbol{\theta}}_{t-1}\rangle| \\ &\le C_0 L^3\gamma_m + \tilde{C}_2\,\bar{\sigma}_t(G). \end{aligned}$$

where $C_0 := \tilde{C}_1\Big(\frac{R^2+\sigma_\varepsilon^2}{\lambda}\Big)^{2/3}$ and $\tilde{C}_2 := (\bar{C}_1(2-\eta m\lambda)^J + \bar{C}_2\lambda^{1/3})\sqrt{\frac{\sigma_\varepsilon^2+R^2}{\lambda}(1+\frac{3\rho_{\max}}{2\lambda})}$ for some constant $\bar{C}_1$, $\bar{C}_2$. For $\lambda \gtrsim (\sigma_\varepsilon^2+R^2)^3+\rho_{\max}$, we have $C_0, \tilde{C}_2 \lesssim 1$ subject to the constraint in η in Lemma A.2. Thus, we obtain

$$I_1 \lesssim L^3\gamma_m + \bar{\sigma}_t(G).$$

Bounding I_2: By Lemma B.5, with at least probability $1-\delta_3$, for all $G\in\mathcal{G}$, we have

$$I_2 = |\langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}^* - \widetilde{\boldsymbol{\theta}}_{t-1}\rangle|.$$

Recall that $\boldsymbol{y}_{t-1} = \boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_{t-1}$ and by Lemma A.3, we have $\boldsymbol{\mu}_{t-1} = \bar{\boldsymbol{G}}_{t-1}^\top\boldsymbol{\theta}^*$. Then,

$$\widetilde{\boldsymbol{\theta}}_{t-1} = \bar{\boldsymbol{U}}_{t-1}^{-1}\bar{\boldsymbol{G}}_{t-1}\bar{\boldsymbol{G}}_{t-1}^\top\boldsymbol{\theta}^*/m + \bar{\boldsymbol{U}}_{t-1}^{-1}\bar{\boldsymbol{G}}_{t-1}\boldsymbol{\epsilon}_{t-1}/m$$

We have $\bar{\boldsymbol{U}}_t = \lambda\boldsymbol{I}_p + \bar{\boldsymbol{G}}_t\bar{\boldsymbol{G}}_t^\top/m$. Hence, $\bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\bar{\boldsymbol{G}}_t^\top/m = \bar{\boldsymbol{U}}_t^{-1}(\bar{\boldsymbol{U}}_t - \lambda\boldsymbol{I}_p) = \boldsymbol{I}_p - \lambda\bar{\boldsymbol{U}}_t^{-1}$. This gives

$$\widetilde{\boldsymbol{\theta}}_{t-1} = \boldsymbol{\theta}^* - \lambda\bar{\boldsymbol{U}}_{t-1}^{-1}\boldsymbol{\theta}^* + \frac{1}{\sqrt{m}}\bar{\boldsymbol{U}}_{t-1}^{-1}\boldsymbol{S}_{t-1}$$

where we have defined $\boldsymbol{S}_{t-1} := \frac{1}{\sqrt{m}}\bar{\boldsymbol{G}}_{t-1}\boldsymbol{\epsilon}_{t-1}$. Thus, we have

$$I_2 \le \lambda|\langle g(G;\boldsymbol{\theta}_0), \bar{\boldsymbol{U}}_t^{-1}\boldsymbol{\theta}^*\rangle| + \frac{1}{\sqrt{m}}|\langle g(G;\boldsymbol{\theta}_0), \bar{\boldsymbol{U}}_{t-1}^{-1}\boldsymbol{S}_{t-1}\rangle| \tag{10}$$

Recall that $\sqrt{m}\bar{\sigma}_t(G) = \|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_{t-1}^{-1}}$. Since $\bar{\boldsymbol{U}}_{t-1}^{-1} \preccurlyeq \frac{1}{\lambda}\boldsymbol{I}_p$, for any vector $\boldsymbol{v}$, we have $\|\boldsymbol{v}\|_{\bar{\boldsymbol{U}}_{t-1}^{-1}} \le \frac{1}{\sqrt{\lambda}}\|\boldsymbol{v}\|$. Then, for the first term in (10), we have

$$\begin{aligned} \lambda|g(G;\boldsymbol{\theta}_0)^\top\bar{\boldsymbol{U}}_{t-1}^{-1}\boldsymbol{\theta}^*| &\le \lambda\|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_{t-1}^{-1}}\cdot\|\boldsymbol{\theta}^*\|_{\bar{\boldsymbol{U}}_{t-1}^{-1}} \\ &\le \sqrt{m}\bar{\sigma}_t(G)\cdot\sqrt{\lambda}\|\boldsymbol{\theta}^*\|_2 \le \bar{\sigma}_t(G)\sqrt{2\lambda}R \end{aligned}$$

where we have used Cauchy-Schwarz inequality for $\langle\cdot,\cdot\rangle_{\boldsymbol{U}_{t-1}^{-1}}$ and Lemma A.3. For the second term in (10), we have

$$\frac{1}{\sqrt{m}}|g(G;\boldsymbol{\theta}_0)^\top\bar{\boldsymbol{U}}_{t-1}^{-1}\boldsymbol{S}_{t-1}| \le \frac{1}{\sqrt{m}}\|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_{t-1}^{-1}}\|\boldsymbol{S}_{t-1}\|_{\bar{\boldsymbol{U}}_{t-1}^{-1}} = \bar{\sigma}_t(G)\cdot\|\boldsymbol{S}_{t-1}\|_{\bar{\boldsymbol{U}}_{t-1}^{-1}}$$

By Theorem 20.4 of Lattimore and Szepesvári (2020), with probability at least $1-\delta_4$, we have

$$\frac{1}{\sigma_\varepsilon^2}\|\boldsymbol{S}_t\|_{\bar{\boldsymbol{U}}_t^{-1}}^2 \le 2\log(1/\delta_4) + \log\deg(\lambda^{-1}\bar{\boldsymbol{U}}_t), \quad \text{for all } t\in\mathbb{N}.$$

By Lemma A.3, with high probability,

$$\log\det(\lambda^{-1}\bar{\boldsymbol{U}}_t) \le \log\det(\boldsymbol{I}_{|\mathcal{G}|} + T\boldsymbol{K}/\lambda) + 1 \le 2\tilde{d}\log(1+T\rho_{\max}/\lambda).$$

Using $\lambda \gtrsim \rho_{\max}$, we have $\log\det(\lambda^{-1}\bar{\boldsymbol{U}}_t) \lesssim \tilde{d}\log(T)+1 \lesssim \tilde{d}\log(T)$. We also have $\log(1/\delta_4) = \log(5T) \lesssim \log(T) \lesssim \tilde{d}\log(T)$.

Putting the pieces together, we have

$$\frac{1}{\sqrt{m}}|g(G;\boldsymbol{\theta}_0)^\top \bar{\boldsymbol{U}}_{t-1}^{-1}\boldsymbol{S}_{t-1}| \lesssim \sigma_\varepsilon\sqrt{\tilde{d}\log T}\cdot\bar{\sigma}_t(G).$$

Combining with the first term, we obtain

$$I_2 \lesssim \big(\sigma_\varepsilon\sqrt{\tilde{d}\log T}+\sqrt{\lambda}R\big)\bar{\sigma}_t(G).$$

Combining with the bound on I_1, we have

$$\begin{aligned}|f_{\mathrm{GNN}}(G;\boldsymbol{\theta}_{t-1})-\mu(G)| &\lesssim L^3\gamma_m + \big(1+\sigma_\varepsilon\sqrt{\tilde{d}\log T}+\sqrt{\lambda}R\big)\,\bar{\sigma}_t(G)\\ &=: L^3\gamma_m + \alpha\,\bar{\sigma}_t(G)\end{aligned}$$

where we have set $\alpha := 1+\sigma_\varepsilon\sqrt{\tilde{d}\log T}+\sqrt{\lambda}R$ for simplificty.

By Lemma A.4, with probability at least $1-\delta_5$,

$$\bar{\sigma}_t(G)-\sigma_t(G) \le CtL^{9/2}\Big(\frac{R^2+\sigma_\varepsilon^2}{\lambda}\Big)^{1/6}\gamma_m \lesssim t\cdot L^{9/2}\gamma_m$$

using the assumption $\lambda \gtrsim R^2+\sigma_\varepsilon^2$. We obtain

$$\begin{aligned}|f_{\mathrm{GNN}}(G;\boldsymbol{\theta}_{t-1})-\mu(G)| &\lesssim L^3\gamma_m + t\cdot\alpha L^{9/2}\gamma_m + \alpha\sigma_t(G)\\ &\le 2t\cdot\alpha L^{9/2}\gamma_m + \alpha\sigma_t(G)\end{aligned}$$

since $t\ge 1$ and $\alpha\ge 1$. Taking $\nu\ge\alpha$ finishes the proof.

A.3 Proof of Lemma 5.2

Proof of Lemma 5.2. Conditioned on $\mathcal{F}_t$, we have

$$\widehat{r}_t(G)\,|\,\mathcal{F}_t \sim \mathcal{N}(f_{\mathrm{GNN}}(G;\boldsymbol{\theta}_{t-1}),\nu^2\sigma_t^2(G)).$$

Using standard Gaussian tail bound, followed by a union bound gives

$$\mathbb{P}_t(|\widehat{r}_t(G)-f_{\mathrm{GNN}}(G;\boldsymbol{\theta}_{t-1})|\ge\nu\sigma_t(G)\cdot u)\le|\mathcal{G}_t|e^{-u^2/2}$$

which shows the first assertion by letting $u=\sqrt{2\log(t^2|\mathcal{G}_t|)}$.

For the second assertion, it is enough to note that $\mathbb{P}(Z\ge 1)\ge(4e\sqrt{\pi})^{-1}$ for $Z\sim\mathcal{N}(0,1)$. □

A.4 Proof of Lemma 5.3

Proof of Lemma 5.3. Our proof is inspired from the proof in Wu et al. (2022). Recall that $c_t^\mu(G)=\nu\sigma_t(G)+\varepsilon(t,m)$ and $c_t^\sigma(G):=\nu\sigma_t(G)\sqrt{2\log(t^2|\mathcal{G}_t|)}$ and

$$\begin{aligned}\mathcal{E}_t^\mu &= \{\forall G\in\mathcal{G}_t, |f_{\mathrm{GNN}}(G;\boldsymbol{\theta}_{t-1})-\mu(G)|\le c_t^\mu(G)\}\\ \mathcal{E}_t^\sigma &= \{\forall G\in\mathcal{G}_t, |\widehat{r}_t(G)-f_{\mathrm{GNN}}(G;\boldsymbol{\theta}_{t-1})|\le c_t^\sigma(G)\}\end{aligned}$$

Let $\gamma_t(G)=c_t^\mu(G)+c_t^\sigma(G)$ and $c_t(G)=\gamma_t(G)+\varepsilon(t,m)$. Then, on $\mathcal{E}_t^\mu\cap\mathcal{E}_t^\sigma$, by triangle inequality,

$$|\widehat{r}_t(G)-\mu(G)|\le\gamma_t(G). \tag{11}$$

We also recall that $\mathcal{E}_t^a := \{\widehat{r}_t(G_t^*)-f_{\mathrm{GNN}}(G_t^*;\boldsymbol{\theta}_{t-1})>\nu\sigma_t(G_t^*)\}$. Then, on $\mathcal{E}_t^\mu\cap\mathcal{E}_t^a$, we have

$$\begin{aligned}\widehat{r}_t(G_t^*) &> f_{\mathrm{GNN}}(G_t^*;\boldsymbol{\theta}_{t-1})+\nu\sigma_t(G_t^*)\\ &\ge \mu(G_t^*)-c_t^\mu(G_t^*)+\nu\sigma_t(G_t^*)\\ &= \mu(G_t^*)-\varepsilon(t,m)\end{aligned} \tag{12}$$

Recall that $\Delta_t := \mu(G_t^*) - \mu(G_t)$ for convenience. Consider the set of unsaturated actions

$$\mathcal{U}_t = \big\{G \in \mathcal{G}_t : \ \mu(G_t^*) < \mu(G) + c_t(G)\big\}$$

and let $\bar{G}_t$ be the least uncertain unsaturated action at time t:

$$\bar{G}_t := \operatorname*{argmin}_{G \in \mathcal{U}_t} c_t(G).$$

By $\bar{G}_t \in \mathcal{U}_t$, we have $\Delta_t \le c_t(\bar{G}_t) + \mu(\bar{G}_t) - \mu(G_t)$. Applying (11), twice, on $\mathcal{E}_t^\mu \cap \mathcal{E}_t^\sigma$, we have

$$\begin{aligned}\Delta_t &\le c_t(\bar{G}_t) + \gamma_t(\bar{G}_t) + \gamma_t(G_t) + \widehat{r}_t(\bar{G}_t) - \widehat{r}_t(G_t)\\ &\le \ c_t(\bar{G}) + \gamma_t(\bar{G}_t) + \gamma_t(G_t)\end{aligned}$$

for all $G \in \mathcal{G}_t$ where the second inequality follows since G_t maximizes $\widehat{r}_t(\cdot)$ over $\mathcal{G}_t$, by design.

Recall that $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \,|\, \mathcal{F}_t]$, where $\mathcal{F}_t$ is the history up to (but not including) time t. Given $\mathcal{F}_t$, the event $\mathcal{E}_t^\mu$ is deterministic while $\mathcal{E}_t^\sigma$ is only random due to the independent randomness in the sampling step (4). Next, we have

$$\begin{aligned}\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{E}_t^\mu}] &= \mathbb{I}_{\mathcal{E}_t^\mu} \cdot \mathbb{E}_t[\Delta_t]\\ &= \mathbb{I}_{\mathcal{E}_t^\mu} \cdot \big(\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{E}_t^\sigma}] + \mathbb{E}_t[\Delta_t \mathbb{I}_{\bar{\mathcal{E}}_t^\sigma}]\big)\\ &\le \mathbb{I}_{\mathcal{E}_t^\mu} \cdot \big(\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{E}_t^\sigma}] + B\,\mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma)\big)\end{aligned} \tag{13}$$

using the boundedness Assumption 2. Here, we are using the fact that $\mathcal{E}_t^\mu$ is measurable w.r.t. $\mathcal{F}_t$, hence it is deterministic conditioned on $\mathcal{F}_t$. Due to factor $\mathbb{I}_{\mathcal{E}_t^\mu}$ in the above, the bound is trivial when $\mathcal{E}_t^\mu$ fails, so for the rest of the proof we assume that $\mathcal{E}_t^\mu$ holds (conditioned on $\mathcal{F}_t$).

We have

$$\begin{aligned}\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{E}_t^\sigma}] &\le c_t(\bar{G}_t) + \gamma_t(\bar{G}_t) + \mathbb{E}_t[\gamma_t(G_t)\mathbb{I}_{\mathcal{E}_t^\sigma}]\\ &\le 2c_t(\bar{G}_t) - \varepsilon(t,m) + \mathbb{E}_t[\gamma_t(G_t)]\end{aligned}$$

where we have used the definition of $c_t(\cdot)$ and dropped the indicator $\mathbb{I}_{\mathcal{E}_t^\sigma}$ to get a further upper bound. It remains to bound $c_t(\bar{G}_t)$ in terms of $\gamma_t(G_t)$.

Since $\bar{G}_t$ is the least uncertain unsaturated action, we have

$$c_t(\bar{G}_t)\mathbb{I}\{G_t \in \mathcal{U}_t\} \le c_t(G_t).$$

Multiplying both sides by $\mathbb{I}_{\mathcal{E}_\sigma^t}$, taking $\mathbb{E}_t[\cdot]$, and rearranging

$$c_t(\bar{G}_t) \le \frac{\mathbb{E}_t[c_t(G_t)\mathbb{I}_{\mathcal{E}_\sigma^t}]}{\mathbb{P}_t(\{G_t \in \mathcal{U}_t\} \cap \mathcal{E}_t^\sigma)} \le \frac{\mathbb{E}_t[\gamma_t(G_t)]}{\mathbb{P}_t(\{G_t \in \mathcal{U}_t\} \cap \mathcal{E}_t^\sigma)}.$$

It remains to bound the denominator.

Recall that G_t maximizes $\widehat{r}_t(\cdot)$ over the entire $\mathcal{G}_t$. Thus, if

$$\widehat{r}_t(G_t^*) > \max_{G \in \bar{\mathcal{U}}_t} \widehat{r}_t(G) \tag{14}$$

then G_t cannot belong to $\bar{\mathcal{U}}_t$, hence $G_t \in \mathcal{U}_t$. On $\mathcal{E}_t^\mu \cap \mathcal{E}_t^\sigma$, for any $G \in \bar{\mathcal{U}}_t$, we have

$$\begin{aligned}\widehat{r}_t(G) \le \mu(G) + \gamma_t(G) &\le \mu(G_t^*) - c_t(G) + \gamma_t(G)\\ &\le \mu(G_t^*) - \varepsilon(t,m)\end{aligned}$$

where the second inequality is by the definition of $\bar{\mathcal{U}}_t$. Then for (14) to hold on $\mathcal{E}_t^\mu \cap \mathcal{E}_t^\sigma$, it is enough to have $\widehat{r}_t(G_t^*) > \mu(G_t^*) - \varepsilon(t,m)$. But this holds on $\mathcal{E}_t^\mu \cap \mathcal{E}_t^a$ by (12). That is,

$$\begin{aligned}\mathcal{E}_t^a \cap \mathcal{E}_t^\mu \cap \mathcal{E}_t^\sigma &\subset \{\widehat{r}_t(G_t^*) > \mu(G_t^*) - \varepsilon(t,m)\} \cap \mathcal{E}_t^\mu \cap \mathcal{E}_t^\sigma \\ &\subset \{\widehat{r}_t(G_t^*) > \max_{G \in \bar{\mathcal{U}}_t} \widehat{r}_t(G)\} \cap \mathcal{E}_t^\mu \cap \mathcal{E}_t^\sigma \\ &\subset \{G_t \in \mathcal{U}_t\} \cap \mathcal{E}_t^\mu \cap \mathcal{E}_t^\sigma.\end{aligned}$$

Assuming as before that $\mathcal{E}_t^\mu$ holds, we have

$$\mathbb{P}_t(\mathcal{E}_t^a \cap \mathcal{E}_t^\sigma) \le \mathbb{P}_t(\{G_t \in \mathcal{U}_t\} \cap \mathcal{E}_t^\sigma).$$

We have $\mathbb{P}_t(\mathcal{E}_t^a \cap \mathcal{E}_t^\sigma) \ge \mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma)$. Putting the pieces together

$$c_t(\bar{G}_t) \le \frac{\mathbb{E}_t[\gamma_t(G_t)]}{\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma)}$$

and we obtain

$$\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{E}_t^\sigma}] \le \Big(\frac{2}{\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^\sigma)} + 1\Big)\mathbb{E}_t[\gamma_t(G_t)] - \varepsilon(t,m)$$

Combining with (13) the result follows. □

A.5 Proof of Lemma 5.4

Proof of Lemma 5.4. For simplicity, we define

$$g_t := \frac{1}{\sqrt{m}} g(G_t; \boldsymbol{\theta}_{t-1}), \quad \bar{g}_t := \frac{1}{\sqrt{m}} g(G_t; \boldsymbol{\theta}_0).$$

Then, recall that

$$\sigma_t^2(G_t) = \|g_t\|^2_{\boldsymbol{U}_{t-1}^{-1}}, \quad \boldsymbol{U}_{t-1} = \lambda \boldsymbol{I}_p + \sum_{i=1}^{t-1} g_t g_t^\top.$$

Note that $\boldsymbol{U}_t = \boldsymbol{U}_{t-1} + g_t g_t^\top$.

Then we introduce following Lemmas:

Lemma A.5 (Elliptical Potential)**.** *Assume that $\boldsymbol{U}_t = \boldsymbol{U}_{t-1} + g_t g_t^\top$ for all $t \in [T]$. Then,*

$$\sum_{t=1}^T \min\{1, \|g_t\|^2_{\boldsymbol{U}_{t-1}^{-1}}\} \le 2 \log\Big(\frac{\det \boldsymbol{U}_T}{\det \boldsymbol{U}_0}\Big).$$

Lemma A.6. *Let $\boldsymbol{A} = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_n]$ and $\bar{\boldsymbol{A}} = [\bar{\boldsymbol{a}}_1\ \bar{\boldsymbol{a}}_2\ \cdots\ \bar{\boldsymbol{a}}_n]$ be $p \times n$ matrices, with columns $\{\boldsymbol{a}_i\}$ and $\{\bar{\boldsymbol{a}}_i\}$, respectively. Assume that for $\varepsilon \le C$, we have*

$$\|\boldsymbol{a}_i - \bar{\boldsymbol{a}}_i\| \le \varepsilon, \quad \|\boldsymbol{a}_i\| \le C$$

for all i. Then,

$$\begin{aligned}\log\det(\boldsymbol{I}_p + \boldsymbol{A}\boldsymbol{A}^\top) &\le \log\det(\boldsymbol{I}_p + \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^\top) + p\log(1 + 3Cn\varepsilon) \\ \log\det(\boldsymbol{I}_p + \boldsymbol{A}\boldsymbol{A}^\top) &\le \log\det(\boldsymbol{I}_n + \bar{\boldsymbol{A}}^\top\bar{\boldsymbol{A}}) + 3Cn^{3/2}\varepsilon.\end{aligned}$$

By Lemma A.5, we have

$$\frac{1}{2}\sum_{t=1}^T \min\{1, \sigma_t^2(G_t)\} \le \log\Big(\frac{\det \boldsymbol{U}_T}{\det \boldsymbol{U}_0}\Big) = \log\det(\lambda^{-1}\boldsymbol{U}_T) =: \log\det(\boldsymbol{V}_T)$$

using $\det(\boldsymbol{U}_0) = \det(\lambda \boldsymbol{I}_p) = \lambda^p$, and defining $\boldsymbol{V}_t := \lambda^{-1}\boldsymbol{U}_t$.

Let $\mathcal{G} = \{G^j : j \in [|\mathcal{G}|]\}$ be the collection of all the graphs and $n_j(t)$ be the number of graphs which are equal to $G^j \in \mathcal{G}$ in our selection of graphs up to and including time t, i.e $n_j(t) := \sum_{t=1}^{t} \mathbb{I}_{G_i = G^j}$. Let

$$\boldsymbol{\psi}_j := \frac{1}{\sqrt{m}} g(G^j; \boldsymbol{\theta}_{t-1}), \quad \bar{\boldsymbol{\psi}}_j := \frac{1}{\sqrt{m}} g(G^j; \boldsymbol{\theta}_0)$$

and let $\boldsymbol{\Psi}$ and $\bar{\boldsymbol{\Psi}}$ be the corresponding $p \times |\mathcal{G}|$ matrices with the above columns. Then, we have

$$\sum_{i=1}^{T} g_i g_i^\top = \sum_{j=1}^{|\mathcal{G}|} n_j(T) \boldsymbol{\psi}_j \boldsymbol{\psi}_j^\top = \boldsymbol{\Psi} \boldsymbol{D} \boldsymbol{\Psi}^\top \preceq T \cdot \boldsymbol{\Psi}\boldsymbol{\Psi}^\top$$

where $\boldsymbol{D} \in \mathbb{R}^{|\mathcal{G}| \times |\mathcal{G}|}$ is the diagonal matrix with diagonal elements $\{n_j(T)\}_{j=1}^{|\mathcal{G}|}$ and the last inequality due to $n_j(T) \leq T$ for all $j \in [|\mathcal{G}|]$.

Note that $\boldsymbol{V}_T = \boldsymbol{I}_p + \lambda^{-1} \sum_{i=1}^{T} g_i g_i^\top$, hence

$$\log\det(\boldsymbol{V}_T) \leq \log\det(\boldsymbol{I}_p + \lambda^{-1} T \cdot \boldsymbol{\Psi}\boldsymbol{\Psi}^\top).$$

By Lemma C.7, fix a $\delta_1 \in (0, 1)$, we have the following bound for $\|\boldsymbol{\psi}_j\|_2$ and $\|\boldsymbol{\psi}_j - \bar{\boldsymbol{\psi}}_j\|_2$, with probability at least $1 - \delta_1$,

$$\|\boldsymbol{\psi}_j\|_2 \leq \frac{1}{N} \sum_{i \in \mathcal{V}(G^j)} \|g_{\text{MLP}}(\boldsymbol{h}_i^{G^j}; \boldsymbol{\theta}_{t-1})/\sqrt{m}\|_2 \leq C_\psi$$

$$\|\boldsymbol{\psi}_j - \bar{\boldsymbol{\psi}}_j\|_2 \leq \frac{1}{N} \sum_{i \in \mathcal{V}(G^j)} \|g_{\text{MLP}}(\boldsymbol{h}_i^{G^j}; \boldsymbol{\theta}_{t-1})/\sqrt{m} - g_{\text{MLP}}(\boldsymbol{h}_i^{G^j}; \boldsymbol{\theta}_0)/\sqrt{m}\|_2 \leq \varepsilon_m$$

where $\varepsilon_m = o(1)$ as $m \to \infty$ and C_ψ is $C_7\sqrt{L}$ in Lemma C.7.

Then, applying Lemma A.6 with $n = |\mathcal{G}|$, $\boldsymbol{A} = \sqrt{\lambda^{-1}T}\boldsymbol{\Psi}$, $\bar{\boldsymbol{A}} = \sqrt{\lambda^{-1}T}\bar{\boldsymbol{\Psi}}$ and ε replaced with $\sqrt{\lambda^{-1}T}\varepsilon_m$, we obtain

$$\log\det(\boldsymbol{V}_T) \leq \log\det(\boldsymbol{I}_{|\mathcal{G}|} + \lambda^{-1} T \cdot \bar{\boldsymbol{\Psi}}^\top \bar{\boldsymbol{\Psi}}) + 3C_\psi |\mathcal{G}|^{3/2} \sqrt{T} \lambda^{-1/2} \varepsilon_m$$

Recall $\hat{\boldsymbol{K}} = \bar{\boldsymbol{\Psi}}^\top \bar{\boldsymbol{\Psi}}$ and $\hat{\rho}_{\max} = \lambda_{\max}(\hat{\boldsymbol{K}})$ and note that $\hat{\boldsymbol{K}}$ is the finite-width GNTK matrix. By Lemma B.6, with high probability, $\hat{\rho}_{\max} \leq \rho_{\max} + \varepsilon_{\rho,m}$ and note that $\varepsilon_{\rho,m} = \Omega(m^{-1/4})$. Dropping $\varepsilon_{\rho,m}$ by large enough m, we have

$$\log\det(\boldsymbol{I}_{|\mathcal{G}|} + \lambda^{-1} T \cdot \bar{\boldsymbol{\Psi}}^\top \bar{\boldsymbol{\Psi}}) \leq |\mathcal{G}| \log(1 + T\rho_{\max}/\lambda).$$

Putting the pieces together with the definition of effective dimension $\tilde{d}$ in (8) finishes the proof. □

A.6 Proof of Lemma A.5

Proof of Lemma A.5. Since $\min\{1, x\} \leq 2\log(1 + x)$ for $x \geq 0$, we have

$$\begin{aligned}\sum_{t=1}^{T} \min\{1, \|g_t\|^2_{\boldsymbol{U}_{t-1}^{-1}}\} &\leq 2 \sum_{t} \log(1 + \|g_t\|^2_{\boldsymbol{U}_{t-1}^{-1}}) \\ &= 2 \sum_{t=1}^{T} \log\Big(\frac{\det \boldsymbol{U}_t}{\det \boldsymbol{U}_{t-1}}\Big) = 2\log\Big(\frac{\det \boldsymbol{U}_T}{\det \boldsymbol{U}_0}\Big)\end{aligned}$$

where the first equality follows from $\det(\boldsymbol{A} + \boldsymbol{v}\boldsymbol{v}^\top) = \det(\boldsymbol{A})(1 + \boldsymbol{v}^\top \boldsymbol{A}^{-1}\boldsymbol{v})$, obtained by an application of Sylvester's determinant identity: $\det(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A})$. □

A.7 Proof of Lemma A.6

Proof of Lemma A.6. Note that

$$\begin{aligned}\|\boldsymbol{a}_i\boldsymbol{a}_i^\top - \bar{\boldsymbol{a}}_i\bar{\boldsymbol{a}}_i^\top\|_{\mathrm{op}} &= \|\boldsymbol{a}_i(\boldsymbol{a}_i - \bar{\boldsymbol{a}}_i)^\top - (\bar{\boldsymbol{a}}_i - \boldsymbol{a}_i)\bar{\boldsymbol{a}}_i^\top\|_{\mathrm{op}} \\ &\le (\|\boldsymbol{a}_i\| + \|\bar{\boldsymbol{a}}_i\|)\|\boldsymbol{a}_i - \boldsymbol{a}_i\| \le (2C+\varepsilon)\varepsilon \le 3C\varepsilon\end{aligned}$$

Let $\boldsymbol{V} = \boldsymbol{I}_p + \boldsymbol{A}\boldsymbol{A}^\top$ and $\bar{\boldsymbol{V}} = \boldsymbol{I}_p + \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^\top$. We have

$$\|\boldsymbol{V} - \bar{\boldsymbol{V}}\|_{\mathrm{op}} \le \sum_{i=1}^n \|\boldsymbol{a}_i\boldsymbol{a}_i^\top - \bar{\boldsymbol{a}}_i\bar{\boldsymbol{a}}_i^\top\|_{\mathrm{op}} \le n \cdot 3C\varepsilon$$

Write $\lambda_i(\boldsymbol{V})$ for the ith eigenvalue of matrix $\boldsymbol{V}$. By Weyl's inequality $|\lambda_i(\boldsymbol{V}) - \lambda_i(\bar{\boldsymbol{V}})| \le 3Cn\varepsilon$. Then,

$$\begin{aligned}\log\det(\boldsymbol{V}) = \sum_{i=1}^p \log\lambda_i(\boldsymbol{V}) &\le \sum_i \log\big(\lambda_i(\bar{\boldsymbol{V}}) + 3Cn\varepsilon\big) \\ &= \sum_i \log(\lambda_i(\bar{\boldsymbol{V}})) + \sum_i \log\Big(1 + \frac{3Cn\varepsilon}{\lambda_i(\bar{\boldsymbol{V}})}\Big) \\ &\le \log\det(\bar{\boldsymbol{V}}) + p\log(1 + 3Cn\varepsilon)\end{aligned}$$

using $\lambda_i(\bar{\boldsymbol{V}}) \ge 1$. This proves one of the bounds.

For the second bound, let $\boldsymbol{W} = \boldsymbol{I}_n + \boldsymbol{A}^\top\boldsymbol{A}$ and $\bar{\boldsymbol{W}} = \boldsymbol{I}_n + \bar{\boldsymbol{A}}^\top\bar{\boldsymbol{A}}$. Then, then by concavity of the $\boldsymbol{X} \mapsto \log\det(\boldsymbol{X})$ and the fact that its derivative is $\boldsymbol{X}^{-1}$ over symmetric matrices, we have

$$\log\det(\boldsymbol{X} + \boldsymbol{\Delta}) - \log\deg(\boldsymbol{X}) \le \mathrm{tr}(\boldsymbol{X}^{-1}\boldsymbol{\Delta}) \le \|\boldsymbol{X}^{-1}\|_F\|\boldsymbol{\Delta}\|_F.$$

Let $\boldsymbol{\Delta} = \boldsymbol{W} - \bar{\boldsymbol{W}}$. We have $|\boldsymbol{\Delta}_{ij}| = |\langle\boldsymbol{a}_i, \boldsymbol{a}_j\rangle - \langle\bar{\boldsymbol{a}}_i, \bar{\boldsymbol{a}}_j\rangle| \le 3C\varepsilon$, hence $\|\boldsymbol{\Delta}\|_F \le 3Cn\varepsilon$ Then,

$$\begin{aligned}\log\det(\boldsymbol{V}) - \log\det(\bar{\boldsymbol{W}}) &\overset{(a)}{=} \log\det(\boldsymbol{W}) - \log\det(\bar{\boldsymbol{W}}) \\ &\le \mathrm{tr}(\bar{\boldsymbol{W}}^{-1}\boldsymbol{\Delta}) \\ &\le \sqrt{n}\|\bar{\boldsymbol{W}}^{-1}\|_{\mathrm{op}}\|\boldsymbol{\Delta}\|_F \overset{(b)}{\le} \sqrt{n}\cdot 3Cn\varepsilon.\end{aligned}$$

where (a) is by Sylvester's identity and (b) uses the fact that $\bar{\boldsymbol{W}} \succeq \boldsymbol{I}_n$, hence $\bar{\boldsymbol{W}}^{-1} \preceq \boldsymbol{I}_n$ giving $\|\bar{\boldsymbol{W}}^{-1}\|_{\mathrm{op}} \le 1$. □

B Technical Lemmas

In this Section, we provides the Proof for Lemmas in Appendix A and other Technical Lemmas supporting the proofs. Most technical Lemmas are related to NTK and optimization in depp learning theory, mainly modified from the GNN helper Lemmas in (Kassraie et al., 2022) and technical Lemmas in Zhou et al. (2020); Vakili et al. (2021).

B.1 Notations for MLP

Recall our GNN with one layer of linear graph convolution and a MLP:

$$\begin{aligned}f^{(1)}(\boldsymbol{h}_i^G) &= \boldsymbol{W}^{(1)}\boldsymbol{h}_i^G, \quad i \in [N], \\ f^{(l)}(\boldsymbol{h}_i^G) &= \frac{1}{\sqrt{m}}\boldsymbol{W}^{(l)}\mathrm{ReLU}(f^{(l-1)}(\boldsymbol{h}_i^G)), \quad 2 \le l \le L, \\ f_{\mathrm{MLP}}(\boldsymbol{h}_i^G; \boldsymbol{\theta}) &= f^{(L)}(\boldsymbol{h}_i^G) \\ f_{\mathrm{GNN}}(G; \boldsymbol{\theta}) &= \frac{1}{N}\sum_{i=1}^N f_{\mathrm{MLP}}(\boldsymbol{h}_i^G; \boldsymbol{\theta}).\end{aligned}$$

We denote the gradients for GNN and associated MLP as

$$g(G;\boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} f_{\text{GNN}}(G;\boldsymbol{\theta})$$
$$g_{\text{MLP}}(\cdot;\boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} f_{\text{MLP}}(\cdot;\boldsymbol{\theta})$$

and the connection between gradients for the MLP and the gradient for the whole GNN is

$$g(G;\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} g_{\text{MLP}}(\boldsymbol{h}_i^G;\boldsymbol{\theta})$$

Similarly, we define a tangent kernel for the a MLP as

$$\tilde{k}_{MLP}(\boldsymbol{x},\boldsymbol{x}') := g_{MLP}(G;\boldsymbol{\theta}_0)^\top g_{MLP}(G';\boldsymbol{\theta}_0)$$

for any MLP inputs $\boldsymbol{x}$, $\boldsymbol{x}'$ and the associated neural tangent kernel $k_{MLP}(\boldsymbol{x},\boldsymbol{x}')$ is defined as limiting kernel of $\tilde{k}_{MLP}(\boldsymbol{x},\boldsymbol{x}')/m$:

$$k_{MLP}(\boldsymbol{x},\boldsymbol{x}') := \lim_{m\to\infty} \tilde{k}_{MLP}(\boldsymbol{x},\boldsymbol{x}')/m.$$

By the connection between f_{GNN} and f_{MLP}, we have

$$k(G,G') = \frac{1}{N^2}\sum_{i\in\mathcal{V}(G)}\sum_{j\in\mathcal{V}(G')} k_{\text{MLP}}(\boldsymbol{h}_i^G,\boldsymbol{h}_j^{G'}).$$

B.2 Proof for Lemmas in Appendix A

Proof of Lemma A.1. By Lemma C.7, with probability at least $1-\delta\in(0,1)$

$$\begin{aligned}
&|f_{\text{GNN}}(G;\boldsymbol{\theta}_t^{(J)}) - f_{\text{GNN}}(G;\boldsymbol{\theta}_0) - \langle g(G;\boldsymbol{\theta}_0),\boldsymbol{\theta}_t^{(J)}-\boldsymbol{\theta}_0\rangle| \\
&\quad\le \frac{1}{N}\sum_{j\in\mathcal{V}(G)} |f_{\text{MLP}}(\boldsymbol{h}_j^G;\boldsymbol{\theta}_t^{(J)}) - f_{\text{MLP}}(\boldsymbol{h}_j^G;\boldsymbol{\theta}_0) - \langle g_{\text{MLP}}(\boldsymbol{h}_j^G;\boldsymbol{\theta}_0),\boldsymbol{\theta}_t^{(J)}-\boldsymbol{\theta}_0\rangle| \\
&\quad\le C_1\tau^{4/3}L^3\sqrt{m\log(m)} \\
&\quad\le C_1(\tilde{C}\sqrt{(R^2+\sigma_\varepsilon^2)/m\lambda})^{4/3}L^3\sqrt{m\log(m)}
\end{aligned}$$

where the last inequality is from the choice of $\tau = \tilde{C}\sqrt{(R^2+\sigma_\varepsilon^2)/m\lambda}$ such that $\|\boldsymbol{\theta}_t^{(J)}-\boldsymbol{\theta}_0\|_2\le\tau$. Since $\tau\propto 1/\sqrt{m}$, it can be verified that technical condition (23) in Lemma C.7 is satisfied when m is large. Therefore, set $C_2 = C_1\tilde{C}^{4/3}$,

$$|f_{\text{GNN}}(G;\boldsymbol{\theta}_t^{(J)}) - f_{\text{GNN}}(G;\boldsymbol{\theta}_0) - \langle g(G;\boldsymbol{\theta}_0),\boldsymbol{\theta}_t^{(J)}-\boldsymbol{\theta}_0\rangle| \le C_2L^3(\frac{R^2+\sigma_\varepsilon^2}{m\lambda})^{2/3}\sqrt{m\log(m)}.$$

□

Proof of Lemma A.2. In this proof, set $\delta_1=\delta_2=\delta/2$ where $\delta\in(0,1)$ is an arbitrary small real value. We introduce $\{\tilde{\boldsymbol{\theta}}_t^{(j)}\}_{j=1}^J$ be the gradient descent update sequence of the following proximal optimization (Kassraie et al., 2022):

$$\min_{\boldsymbol{\theta}} \frac{1}{2t}\sum_{i=1}^{t}(\langle g(G_i;\boldsymbol{\theta}_0),\boldsymbol{\theta}-\boldsymbol{\theta}_0\rangle - y_i)^2 + \frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2$$

and $\{\boldsymbol{\theta}_t^{(j)}\}_{j=1}^J$ be the gradient descent update sequence of parameters of our primary optimization (6). In GNN training step in algorithms, we let $\boldsymbol{\theta}_t := \boldsymbol{\theta}_t^{(J)}$. Recall that $\bar{\boldsymbol{U}}_t = \lambda\boldsymbol{I} + \bar{\boldsymbol{G}}_t\bar{\boldsymbol{G}}_t^\top/m$. By

Lemma B.5, with probability at least $1-\delta_1 \in (0,1)$, $\bar{\boldsymbol{U}}_t \preccurlyeq (\lambda + \frac{3}{2}\rho_{\max})\boldsymbol{I}$. Therefore,

$$\begin{aligned}
|\langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}_t - \boldsymbol{\theta}_0 - \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m\rangle| &\leq \|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_t^{-1}}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_0 - \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m\|_{\bar{\boldsymbol{U}}_t} \\
&\leq \sqrt{\lambda + 3\rho_{\max}/2}\|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_t^{-1}}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_0 - \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m\|_2 \\
&\leq \sqrt{\lambda + 3\rho_{\max}/2}\|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_t^{-1}}(\|\tilde{\boldsymbol{\theta}}_t^{(J)} - \boldsymbol{\theta}_0 - \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m\|_2 \\
&\qquad + \|\tilde{\boldsymbol{\theta}}_t^{(J)} - \boldsymbol{\theta}_t\|_2
\end{aligned}$$

By Lemma B.3 and Lemma B.1, with probability at least $1-\delta_2 \in (0,1)$, for some constants C_1 and C_2, we have

$$\begin{aligned}
&|\langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}_t - \boldsymbol{\theta}_0 - \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m\rangle| \\
\leq &\sqrt{\lambda + 3\rho_{\max}/2}\|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_t^{-1}}\left(C_1(2-\eta m\lambda)^J\sqrt{\frac{\sigma_\varepsilon^2 + R^2}{m\lambda}} + \|\tilde{\boldsymbol{\theta}}_t^{(J)} - \boldsymbol{\theta}_t\|_2\right) \quad \text{(by Lemma B.3)} \\
\leq &\sqrt{\lambda + 3\rho_{\max}/2}\|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_t^{-1}} \times \left(C_1(2-\eta m\lambda)^J\sqrt{\frac{\sigma_\varepsilon^2 + R^2}{m\lambda}} + C_2\sqrt{\frac{\sigma_\varepsilon^2 + R^2}{m\lambda}}\right) \quad \text{(by Lemma B.1)} \\
= &\sqrt{m(1 + \frac{3\rho_{\max}}{2\lambda})}(C_1(2-\eta m\lambda)^J + C_2)\sqrt{\frac{\sigma_\varepsilon^2 + R^2}{m\lambda}}\bar{\sigma}_{t+1}(G)
\end{aligned}$$

The last equality is obtained from the definition of $\bar{\sigma}_{t+1}^2(G)$, which is $\bar{\sigma}_{t+1}^2(G) = \lambda g^\top(G;\boldsymbol{\theta}_0)\bar{\boldsymbol{U}}_t^{-1}g(G;\boldsymbol{\theta}_0)/m = \frac{\lambda}{m}\|g(G;\boldsymbol{\theta}_0)\|_{\bar{\boldsymbol{U}}_t^{-1}}^2$. Now we let $\tilde{C} = \sqrt{m(1 + \frac{3\rho_{\max}}{2\lambda})}(C_1(2-\eta m\lambda)^J + C_2)\sqrt{\frac{\sigma_\varepsilon^2+R^2}{m\lambda}}$. Note that this constant $\tilde{C} = \mathcal{O}(1)$ with respect to m since $\eta = \mathcal{O}(m^{-1})$. Then we have the desired result:

$$|\langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}_t - \boldsymbol{\theta}_0 - \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m\rangle| \leq \tilde{C}\bar{\sigma}_{t+1}(G)$$

where $\tilde{C} = (C_1(2-\eta m\lambda)^J + C_2)\sqrt{\frac{\sigma_\varepsilon^2+R^2}{\lambda}(1 + \frac{3\rho_{\max}}{2\lambda})}$ with $C_1 = \mathcal{O}(1)$ and $C_2 = \mathcal{O}(\lambda^{1/3})$. □

Proof of Lemma A.3. See Appendix B.4. □

Proof of Lemma A.4. Define function ψ_λ for vectors $\{\boldsymbol{v}, \boldsymbol{a}_1, ..., \boldsymbol{a}_{t-1}\}$ as followed:

$$\psi_\lambda(\boldsymbol{v}, \boldsymbol{a}_1, ..., \boldsymbol{a}_{t-1}) := \sqrt{\boldsymbol{v}^\top(\lambda\boldsymbol{I} + \sum_{i=1}^{t-1}\boldsymbol{a}_i\boldsymbol{a}_i^\top)^{-1}\boldsymbol{v}},$$

and denote the gradients for ψ_λ as

$$\begin{aligned}
\nabla_0\psi_\lambda &:= \nabla_{\boldsymbol{v}}\psi_\lambda(\boldsymbol{v}, \boldsymbol{a}_1, ..., \boldsymbol{a}_{t-1}) \\
\nabla_i\psi_\lambda &:= \nabla_{\boldsymbol{a}_i}\psi_\lambda(\boldsymbol{v}, \boldsymbol{a}_1, ..., \boldsymbol{a}_{t-1}), \forall i \in [t-1].
\end{aligned}$$

By setting $\boldsymbol{A} = (\lambda\boldsymbol{I} + \sum_{i=1}^{t-1}\boldsymbol{a}_i\boldsymbol{a}_i^\top)^{-1} \preccurlyeq \frac{1}{\lambda}\boldsymbol{I}$ with eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^\top$. The gradients are bounded as followed

$$\begin{aligned}
\|\nabla_0\psi_\lambda\|_2 &= \frac{\|\boldsymbol{A}\boldsymbol{v}\|_2}{\sqrt{\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}} = \sqrt{\frac{\boldsymbol{v}^\top\boldsymbol{A}^2\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}} \leq \sqrt{\lambda_{max}(\boldsymbol{A})} \leq 1/\sqrt{\lambda} \\
\|\nabla_i\psi_\lambda\|_2 &= \frac{\|\boldsymbol{A}\boldsymbol{v}\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{a}_i\|_2}{\sqrt{\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}} \leq \|\boldsymbol{a}_i\|_2\frac{\boldsymbol{v}^\top\boldsymbol{A}^2\boldsymbol{v}}{\sqrt{\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}} \leq \|\boldsymbol{a}_i\|_2\|\boldsymbol{v}\|_2/\lambda
\end{aligned} \tag{15}$$

We can express $\bar{\sigma}_t(G)$ and $\sigma_t(G)$ by ψ_λ:

$$\begin{aligned}
\bar{\sigma}_t(G) &= \psi_\lambda(\frac{g(G;\boldsymbol{\theta}_{t-1})}{\sqrt{m}}, \frac{g(G_1;\boldsymbol{\theta}_1)}{\sqrt{m}}, ..., \frac{g(G_{t-1};\boldsymbol{\theta}_{t-1})}{\sqrt{m}}) \\
\sigma_t(G) &= \psi_\lambda(\frac{g(G;\boldsymbol{\theta}_0)}{\sqrt{m}}, \frac{g(G_1;\boldsymbol{\theta}_0)}{\sqrt{m}}, ..., \frac{g(G_{t-1};\boldsymbol{\theta}_0)}{\sqrt{m}}).
\end{aligned}$$

From Lemma C.7, there exists positive constants such that the gradients and gradient differences are bounded with high probability, which indicates for some constant C_1 with probability greater than $1-\delta$,

$$\|g(G;\boldsymbol{\theta})\|_2 = \|\frac{1}{N}\sum_{j\in\mathcal{V}(G)} g_{\text{MLP}}(\boldsymbol{h}_j^G;\boldsymbol{\theta})\|_2 \leq C_1\sqrt{mL} \tag{16}$$

Note that ψ_λ is Lipschitz continuous, then with high probability, we have

$$\begin{aligned}
|\bar{\sigma}_t(G)-\sigma_t(G)| &= |\psi_\lambda(\frac{g(G;\boldsymbol{\theta}_{t-1})}{\sqrt{m}}, \frac{g(G_1;\boldsymbol{\theta}_1)}{\sqrt{m}}, ..., \frac{g(G_{t-1};\boldsymbol{\theta}_{t-1})}{\sqrt{m}}) - \psi_\lambda(\frac{g(G;\boldsymbol{\theta}_0)}{\sqrt{m}}, \frac{g(G_1;\boldsymbol{\theta}_0)}{\sqrt{m}}, ..., \frac{g(G_{t-1};\boldsymbol{\theta}_0)}{\sqrt{m}})| \\
&\leq \sup\{\|\nabla_0\psi_\lambda\|_2\}\|\frac{g(G;\boldsymbol{\theta}_{t-1})}{\sqrt{m}} - \frac{g(G;\boldsymbol{\theta}_0)}{\sqrt{m}}\|_2 + \sum_{i=1}^{t-1}\sup\{\|\nabla_i\psi_\lambda\|_2\}\|\frac{g(G_i;\boldsymbol{\theta}_i)}{\sqrt{m}} - \frac{g(G_i;\boldsymbol{\theta}_0)}{\sqrt{m}}\|_2 \\
&\leq \frac{1}{\sqrt{\lambda}}\|\frac{g(G;\boldsymbol{\theta}_{t-1})}{\sqrt{m}} - \frac{g(G;\boldsymbol{\theta}_0)}{\sqrt{m}}\|_2 + \frac{C_1^2 L}{\lambda}\sum_{i=1}^{t-1}\|\frac{g(G_i;\boldsymbol{\theta}_i)}{\sqrt{m}} - \frac{g(G_i;\boldsymbol{\theta}_0)}{\sqrt{m}}\|_2 (\text{ by (15) and (16)}) \\
&\leq C_2\sqrt{\log(m)}\tau^{1/3}L^3\|g(G;\boldsymbol{\theta}_0)\|_2/\sqrt{m}(\frac{1}{\sqrt{\lambda}} + \frac{C_1^2 Lt}{\lambda}) \quad (\text{by Lemma C.7}) \\
&\leq C_1C_2\sqrt{\log(m)}\tau^{1/3}L^{7/2}(\frac{1}{\sqrt{\lambda}} + \frac{C_1^2 Lt}{\lambda}) \quad (\text{by (16)})
\end{aligned}$$

Therefore, if $\lambda \leq C_1^4L^2t^2$ and let $\tau = \tilde{C}\sqrt{\frac{R^2+\sigma_\varepsilon^2}{m\lambda}}$, $C_3 = 2\tilde{C}C_2C_1^3$,

$$|\bar{\sigma}_t(G)-\sigma_t(G)| \leq C_3 t\lambda^{-7/6}L^{9/2}(R^2+\sigma_\varepsilon^2)^{1/6}m^{-1/6}\sqrt{\log(m)}$$

□

B.3 Lemmas for GNN training

Lemma B.1 (Parameter Bound for Primary Optimization)**.** *Let $\{\boldsymbol{\theta}_t^{(j)}\}_{j=1}^J$ be the gradient descent update sequence of parameters of the optimization (6) which is,*

$$\min_{\boldsymbol{\theta}} \frac{1}{2t}\sum_{i=1}^{t}(f_{\text{GNN}}(G_i;\boldsymbol{\theta}) - y_i)^2 + \frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2$$

then if $m \geq poly(R, \sigma_\varepsilon, L, \lambda^{-1}, \log(\frac{N}{\delta}))$ and learning rate $\eta \leq (\tilde{C}mL + m\lambda)^{-1}$ for some constant $\tilde{C}$. Then for a constant $C = \mathcal{O}(\lambda^{1/3})$ which is independent of m and t, with probability at least $1-\delta$

$$\|\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}\|_2 \leq C\sqrt{\frac{R^2+\sigma_\varepsilon^2}{m\lambda}}$$

where $\{\tilde{\boldsymbol{\theta}}_t^{(j)}\}_{j=1}^J$ be the gradient descent update sequence of parameters of the proximal optimization with loss function $\frac{1}{2t}\sum_{i=1}^t(\langle g(G_i;\boldsymbol{\theta}_0), \boldsymbol{\theta}-\boldsymbol{\theta}_0\rangle - y_i)^2 + \frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2$. Both optimization have the same initialization at $\tilde{\boldsymbol{\theta}}_t^{(0)} = \boldsymbol{\theta}_t^{(0)} = \boldsymbol{\theta}_0$ and same learning rate η.

Proof. In this proof, set $\delta_1 = \delta_2 = \delta/2$ where $\delta \in (0,1)$ is an arbitrary small real value. Define $\boldsymbol{G}_t^{(j)} := [g(G_1;\boldsymbol{\theta}_t^{(j)}), ..., g(G_t;\boldsymbol{\theta}_t^{(j)})] \in \mathbb{R}^{p\times t}$ as the j-th updates in our primary optimzation with loss (6) at round t. Also define $\boldsymbol{f}_{gnn,t}^{(j)} := [f_{\text{GNN}}(G_1;\boldsymbol{\theta}_t^{(j)}), ..., f_{\text{GNN}}(G_t;\boldsymbol{\theta}_t^{(j)})]^\top \in \mathbb{R}^{t\times 1}$. The gradient descent updates for sequences $\{\boldsymbol{\theta}_t^{(j)}\}_{j=1}^J$ and $\{\tilde{\boldsymbol{\theta}}_t^{(j)}\}_{j=1}^J$ are

$$\begin{aligned}
\boldsymbol{\theta}_t^{(j+1)} &= \boldsymbol{\theta}_t^{(j)} - \eta\left(\frac{1}{t}[\boldsymbol{G}_t^{(j)}]^\top(\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t) + m\lambda\boldsymbol{\theta}_t^{(j)}\right) \\
\tilde{\boldsymbol{\theta}}_t^{(j+1)} &= \tilde{\boldsymbol{\theta}}_t^{(j)} - \eta\left(\frac{1}{t}\bar{\boldsymbol{G}}_t^\top(\bar{\boldsymbol{G}}_t(\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_0) - \boldsymbol{y}_t) + m\lambda\tilde{\boldsymbol{\theta}}_t^{(j)}\right)
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\|\boldsymbol{\theta}_t^{(j+1)} - \tilde{\boldsymbol{\theta}}_t^{(j+1)}\|_2 \\
=&\|(1-\eta m\lambda)(\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}) - \frac{\eta}{t}[\boldsymbol{G}_t^{(j)}]^\top(\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t) + \frac{\eta}{t}\bar{\boldsymbol{G}}_t^\top(\bar{\boldsymbol{G}}_t(\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_0) - \boldsymbol{y}_t)\|_2 \\
=&\|(1-\eta m\lambda)(\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}) - \frac{\eta}{t}(\boldsymbol{G}_t^{(j)} - \bar{\boldsymbol{G}}_t)^\top(\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t) - \frac{\eta}{t}\bar{\boldsymbol{G}}_t^\top(\boldsymbol{f}_{gnn,t}^{(j)} - \bar{\boldsymbol{G}}_t(\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_0))\|_2 \\
=&\|(1-\eta m\lambda)(\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}) - \frac{\eta}{t}(\boldsymbol{G}_t^{(j)} - \bar{\boldsymbol{G}}_t)^\top(\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t) - \frac{\eta}{t}\bar{\boldsymbol{G}}_t^\top(\boldsymbol{f}_{gnn,t}^{(j)} - \bar{\boldsymbol{G}}_t(\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0) + \bar{\boldsymbol{G}}_t(\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}))\|_2 \\
=&\|(\boldsymbol{I} - \eta(m\lambda\boldsymbol{I} + \bar{\boldsymbol{G}}_t^\top\bar{\boldsymbol{G}}_t/t))(\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}) - \frac{\eta}{t}(\boldsymbol{G}_t^{(j)} - \bar{\boldsymbol{G}}_t)^\top(\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t) - \frac{\eta}{t}\bar{\boldsymbol{G}}_t^\top(\boldsymbol{f}_{gnn,t}^{(j)} - \bar{\boldsymbol{G}}_t(\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0))\|_2 \\
\leq&\underbrace{\|(\boldsymbol{I} - \eta(m\lambda\boldsymbol{I} + \bar{\boldsymbol{G}}_t^\top\bar{\boldsymbol{G}}_t/t))\|_2\|\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}\|_2}_{I_1} + \underbrace{\frac{\eta}{t}\|\bar{\boldsymbol{G}}_t\|_2\|\boldsymbol{f}_{gnn,t}^{(j)} - \bar{\boldsymbol{G}}_t(\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0)\|_2}_{I_2} + \underbrace{\frac{\eta}{t}\|\boldsymbol{G}_t^{(j)} - \bar{\boldsymbol{G}}_t\|_2\|\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t\|_2}_{I_3}
\end{aligned}
$$

For I_1, due to $\bar{\boldsymbol{G}}_t^\top\bar{\boldsymbol{G}}_t/t \succcurlyeq \boldsymbol{0}$, we have

$$
I_1 = \|(\boldsymbol{I} - \eta(m\lambda\boldsymbol{I} + \bar{\boldsymbol{G}}_t^\top\bar{\boldsymbol{G}}_t/t))\|_2\|\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}\|_2 \leq (1-\eta m\lambda)\|\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}\|_2
$$

For I_2, by Lemma B.4, set $\tau = \tilde{C}\sqrt{(R^2+\sigma_\varepsilon^2)/m\lambda}$. Since $\tau \propto 1/\sqrt{m}$, it can be verified that technical condition (23) in Lemma C.7 is satisfied when m is large. Then with probability at least $1-\delta_1 \in (0,1)$,

$$
I_2 = \frac{\eta}{t}\|\bar{\boldsymbol{G}}_t\|_2\|\boldsymbol{f}_{gnn,t}^{(j)} - \bar{\boldsymbol{G}}_t(\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0)\|_2 \leq \eta C_1(\tilde{C}\frac{R^2+\sigma_\varepsilon^2}{m\lambda})^{2/3}L^{7/2}m\sqrt{\log(m)}
$$

For I_3, by Lemma B.2 and Lemma B.4, and Lemma C.7, with probability at least $1-\delta_2 \in (0,1)$,

$$
I_3 = \frac{\eta}{t}\|\boldsymbol{G}_t^{(j)} - \bar{\boldsymbol{G}}_t\|_2\|\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t\|_2 \leq \eta C_2(\tilde{C}\frac{R^2+\sigma_\varepsilon^2}{m\lambda})^{1/6}L^{7/2}\sqrt{m\log(m)}\sqrt{R^2+\sigma_\varepsilon^2}
$$

Put the upper bound for I_1, I_2., I_3 together and set $C_3 = (\lambda^{1/3}C_1 + C_2)\tilde{C} = \mathcal{O}(\lambda^{1/3})$, then we get,

$$
\|\boldsymbol{\theta}_t^{(j+1)} - \tilde{\boldsymbol{\theta}}_t^{(j+1)}\|_2 \leq (1-\eta m\lambda)\|\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}\|_2 + C_3\eta(R^2+\sigma_\varepsilon^2)^{2/3}L^{7/2}m^{1/3}\lambda^{-1/6}\sqrt{\log(m)}
$$

Therefore, there exists $m = poly(R, \sigma_\varepsilon, \lambda, L)$ satisfies that $(R^2+\sigma_\varepsilon^2)^{1/6}L^{7/2}\lambda^{1/3}\sqrt{\log(m)} \leq m^{1/6}$, which indicates

$$
\|\boldsymbol{\theta}_t^{(j)} - \tilde{\boldsymbol{\theta}}_t^{(j)}\|_2 \leq C_3(R^2+\sigma_\varepsilon^2)^{2/3}L^{7/2}m^{-2/3}\lambda^{-1/6}\sqrt{\log(m)} \quad \leq C_3\sqrt{\frac{R^2+\sigma_\varepsilon^2}{m\lambda}}
$$

□

Lemma B.2 (Prediction Error Bound in Gradient Descent)**.** *Let $\{\boldsymbol{\theta}_t^{(j)}\}_{j=1}^J$ be the gradient descent update sequence of parameters of the optimization (6). Define $\boldsymbol{f}_{gnn,t}^{(j)} := [f_{\text{GNN}}(G_1;\boldsymbol{\theta}_t^{(j)}), ..., f_{\text{GNN}}(G_t;\boldsymbol{\theta}_t^{(j)})]^\top \in \mathbb{R}^{t\times 1}$. Assume τ is set such that $\|\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0\|_2 \leq \tau$ for all t and $\forall j \leq J$. Suppose $m \geq poly(L, \lambda^{-1}, \log(N/\delta))$ where $\delta \in (0,1)$ and learning rate $\eta \leq (\tilde{C}mL + m\lambda)^{-1}$ for some constant $\tilde{C}$, then with probability at least $1-\delta$,*

$$
\|\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t\|_2 \leq C\sqrt{t(R^2+\sigma_\varepsilon^2)}
$$

where C is some constant which does not depend on m and t.

Proof. Define $\boldsymbol{f}_t(\boldsymbol{\theta})$ and $\boldsymbol{G}_t(\boldsymbol{\theta})$ as follow

$$
\begin{aligned}
\boldsymbol{f}_t(\boldsymbol{\theta}) &= [f_{\text{GNN}}(G_1;\boldsymbol{\theta}), ..., f_{\text{GNN}}(G_t;\boldsymbol{\theta})]^\top \in \mathbb{R}^{t\times 1} \\
\boldsymbol{G}_t(\boldsymbol{\theta}) &= [g(G_1;\boldsymbol{\theta}), ..., g(G_t;\boldsymbol{\theta})] \in \mathbb{R}^{p\times t}
\end{aligned}
$$

Also define $\mathcal{L}_t(\boldsymbol{\theta}) := \frac{1}{2t}\sum_{i=1}^{t}(f_{\text{GNN}}(G_i;\boldsymbol{\theta})-y_i)^2+\frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2$ as the loss function in primary optimization. Note that $\mathcal{L}_t(\boldsymbol{\theta}) := \frac{1}{2t}\|\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t\|_2^2+\frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2$ First notice that loss function $\mathcal{L}_t(\boldsymbol{\theta})$ is convex due to the strongly convexity of $\|\cdot\|_2^2/2$. We are going to use the following two-sided bound from strongly convexity in this proof:

$$\|\boldsymbol{y}\|_2^2/2-\|\boldsymbol{x}\|_2^2/2=\boldsymbol{x}^\top(\boldsymbol{y}-\boldsymbol{x})+\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2$$

By 1-strongly convexity of $\|\cdot\|_2^2/2$, we have

$$\begin{aligned}\mathcal{L}_t(\boldsymbol{\theta}')-\mathcal{L}_t(\boldsymbol{\theta})=&\frac{1}{2t}\left(\|\boldsymbol{f}_t(\boldsymbol{\theta}')-\boldsymbol{y}_t\|_2^2-\|\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t\|_2^2\right)+\frac{m\lambda}{2}\left(\|\boldsymbol{\theta}'\|_2^2-\|\boldsymbol{\theta}\|_2^2\right)\\ \leq&\frac{1}{t}\left((\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)^\top(\boldsymbol{f}_t(\boldsymbol{\theta}')-\boldsymbol{f}_t(\boldsymbol{\theta}))+\frac{1}{2}\|\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{f}_t(\boldsymbol{\theta}')\|_2^2\right)+m\lambda\left(\boldsymbol{\theta}^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{1}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2\right).\end{aligned}$$

Define $\boldsymbol{e}_t := \boldsymbol{f}_t(\boldsymbol{\theta}')-\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{G}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}'-\boldsymbol{\theta})$. By Lemma B.4, with probability at least $1-\delta_1\in(0,1)$

$$\begin{aligned}&\mathcal{L}_t(\boldsymbol{\theta}')-\mathcal{L}_t(\boldsymbol{\theta})\\ \leq&\frac{1}{t}(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)^\top(\boldsymbol{G}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}'-\boldsymbol{\theta})+\boldsymbol{e}_t)+\frac{1}{2t}\|\boldsymbol{G}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}'-\boldsymbol{\theta})+\boldsymbol{e}_t\|_2^2+m\lambda\left(\boldsymbol{\theta}^\top(\boldsymbol{\theta}-\boldsymbol{\theta})+\frac{1}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2\right)\\ =&\frac{1}{t}[\boldsymbol{G}_t(\boldsymbol{\theta})(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)+m\lambda\boldsymbol{\theta}]^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{1}{t}(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)^\top\boldsymbol{e}_t+\frac{1}{2t}\|\boldsymbol{G}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}'-\boldsymbol{\theta})+\boldsymbol{e}_t\|_2^2+\frac{m\lambda}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2\\ =&\nabla\mathcal{L}_t(\boldsymbol{\theta})^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{1}{t}(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)^\top\boldsymbol{e}_t+\frac{1}{2t}\|\boldsymbol{G}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}'-\boldsymbol{\theta})+\boldsymbol{e}_t\|_2^2+\frac{m\lambda}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2\\ \leq&\nabla\mathcal{L}_t(\boldsymbol{\theta})^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2+\frac{1}{t}\|\boldsymbol{G}_t(\boldsymbol{\theta})\|_2^2\|\boldsymbol{\theta}'-\boldsymbol{\theta}\|_2^2+\frac{1}{t}\|\boldsymbol{e}_t\|_2^2+\frac{m\lambda}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2\\ \leq&\nabla\mathcal{L}_t(\boldsymbol{\theta})^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2+\frac{1}{t}\|\boldsymbol{e}_t\|_2^2+(C_1^2mL+m\lambda/2)\|\boldsymbol{\theta}'-\boldsymbol{\theta}\|_2^2\quad\text{(by Lemma B.4)}\end{aligned}\tag{17}$$

Similarly by 1-strongly convexity of $\|\cdot\|_2^2/2$, we also investigate the lower bound:

$$\mathcal{L}_t(\boldsymbol{\theta}')-\mathcal{L}_t(\boldsymbol{\theta})\geq\frac{1}{t}\left((\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)^\top(\boldsymbol{f}_t(\boldsymbol{\theta}')-\boldsymbol{f}_t(\boldsymbol{\theta}))+\frac{1}{2}\|\boldsymbol{f}_t(\boldsymbol{\theta}')-\boldsymbol{f}_t(\boldsymbol{\theta})\|_2^2\right)+m\lambda\left(\boldsymbol{\theta}^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{1}{2}\|\boldsymbol{\theta}'-\boldsymbol{\theta}\|_2^2\right)$$

Using $\boldsymbol{e}_t := \boldsymbol{f}_t(\boldsymbol{\theta}')-\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{G}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}'-\boldsymbol{\theta})$, we obtain

$$\begin{aligned}\mathcal{L}_t(\boldsymbol{\theta}')-\mathcal{L}_t(\boldsymbol{\theta})\geq&\frac{1}{t}(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)^\top(\boldsymbol{G}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}'-\boldsymbol{\theta})+\boldsymbol{e}_t)+m\lambda\boldsymbol{\theta}^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{m\lambda}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2\\ =&\frac{1}{t}[\boldsymbol{G}_t(\boldsymbol{\theta})(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)+m\lambda\boldsymbol{\theta}]^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{1}{t}(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)^\top\boldsymbol{e}_t+\frac{m\lambda}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2\end{aligned}$$

Then using $\nabla\mathcal{L}_t(\boldsymbol{\theta})=\boldsymbol{G}_t(\boldsymbol{\theta})(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)+m\lambda\boldsymbol{\theta}$, we have

$$\begin{aligned}\mathcal{L}_t(\boldsymbol{\theta}')-\mathcal{L}_t(\boldsymbol{\theta})\geq&\nabla\mathcal{L}_t(\boldsymbol{\theta})^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{1}{t}(\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t)^\top\boldsymbol{e}_t+\frac{m\lambda}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2\\ \geq&\nabla\mathcal{L}_t(\boldsymbol{\theta})^\top(\boldsymbol{\theta}'-\boldsymbol{\theta})+\frac{m\lambda}{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2-\frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2\\ \geq&-\frac{\|\nabla\mathcal{L}_t(\boldsymbol{\theta})\|_2^2}{2m\lambda}-\frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta})-\boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2\quad\text{(by Lemma C.1)}\end{aligned}\tag{18}$$

Now recall the update step $\boldsymbol{\theta}_t^{(j+1)} = \boldsymbol{\theta}_t^{(j)} - \eta\nabla\mathcal{L}_t(\boldsymbol{\theta}_t^{(j)})$ and combine the above upper and lower bounds,

$$
\begin{aligned}
&\mathcal{L}_t(\boldsymbol{\theta} - \eta\nabla\mathcal{L}_t(\boldsymbol{\theta})) - \mathcal{L}_t(\boldsymbol{\theta}) \\
\leq& -\eta\|\nabla\mathcal{L}_t(\boldsymbol{\theta})\|_2^2 + \frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2 + \frac{1}{t}\|\boldsymbol{e}_t\|_2^2 + \eta^2(C_1^2 mL + m\lambda/2)\|\nabla\mathcal{L}_t(\boldsymbol{\theta})\|_2^2 \quad \text{(by update step and (17))} \\
=& -\eta\left(1 - \frac{\eta}{2}(2C_1^2 mL + m\lambda)\right)\|\nabla\mathcal{L}_t(\boldsymbol{\theta})\|_2^2 + \frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2 + \frac{1}{t}\|\boldsymbol{e}_t\|_2^2 \\
\leq& -\frac{\eta}{2}\|\nabla\mathcal{L}_t(\boldsymbol{\theta})\|_2^2 + \frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2 + \frac{1}{t}\|\boldsymbol{e}_t\|_2^2 \quad \text{(by choice of } \eta) \\
\leq& \eta m\lambda\left(\mathcal{L}_t(\boldsymbol{\theta}') - \mathcal{L}_t(\boldsymbol{\theta}) + \frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2\right) + \frac{1}{t}\|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2\|\boldsymbol{e}_t\|_2 + \frac{1}{t}\|\boldsymbol{e}_t\|_2^2 \quad \text{(by (18))} \\
\leq& \eta m\lambda\left(\mathcal{L}_t(\boldsymbol{\theta}') - \mathcal{L}_t(\boldsymbol{\theta}) + \|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2^2/8t + 2\|\boldsymbol{e}_t\|_2^2/t\right) + \frac{1}{t}(\eta m\lambda\|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2^2/8 + 2\|\boldsymbol{e}_t\|_2^2/\eta m\lambda) + \frac{1}{t}\|\boldsymbol{e}_t\|_2^2 \\
=& \eta m\lambda(\mathcal{L}_t(\boldsymbol{\theta}') - \mathcal{L}_t(\boldsymbol{\theta})) + \frac{\eta m\lambda}{4t}\|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2^2 + (\frac{2\eta m\lambda}{t} + \frac{2}{\eta m\lambda t} + \frac{1}{t})\|\boldsymbol{e}_t\|_2^2 \\
\leq& \eta m\lambda(\mathcal{L}_t(\boldsymbol{\theta}') - \mathcal{L}_t(\boldsymbol{\theta})) + \eta m\lambda\mathcal{L}_t(\boldsymbol{\theta})/2 + (\frac{2\eta m\lambda}{t} + \frac{2}{\eta m\lambda t} + \frac{1}{t})\|\boldsymbol{e}_t\|_2^2 \quad \text{(by } \|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2^2 \leq 2t\mathcal{L}(\boldsymbol{\theta})) \\
=& \eta m\lambda(\mathcal{L}_t(\boldsymbol{\theta}') - \mathcal{L}_t(\boldsymbol{\theta})/2) + (\frac{2\eta m\lambda}{t} + \frac{2}{\eta m\lambda t} + \frac{1}{t})\|\boldsymbol{e}_t\|_2^2
\end{aligned} \tag{19}
$$

For $\|\boldsymbol{e}_t\|_2^2$, by Lemma C.7, with probability at least $1 - \delta_2 \in (0,1)$ for some constant C_2, we have

$$
\begin{aligned}
\|\boldsymbol{e}_t\|_2 &= \|\boldsymbol{f}_t(\boldsymbol{\theta}') - \boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{G}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}' - \boldsymbol{\theta})\|_2 \\
&\leq \sqrt{t}\max_{i\in[t]}|f_{\text{GNN}}(G_i;\boldsymbol{\theta}') - f_{\text{GNN}}(G_i;\boldsymbol{\theta}) + g^\top(G_i;\boldsymbol{\theta})(\boldsymbol{\theta}' - \boldsymbol{\theta})| \\
&\leq \frac{\sqrt{t}}{N}\max_{i\in[t]}\sum_{j\in\mathcal{V}(G_i)}|f_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta}') - f_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta}) + g_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta})^\top(\boldsymbol{\theta}' - \boldsymbol{\theta})| \\
&\leq C_2\tau^{4/3}L^3\sqrt{tm\log(m)}
\end{aligned} \tag{20}
$$

where $\mathcal{V}(G)$ as vertice set of a graph G. Moreove, by Lemma C.4, we have the high probability upper bound for $\frac{1}{t}\|\boldsymbol{y}_t\|_2^2$: with probability at least $1 - \delta_3 \in (0,1)$ and some constant C_3 depends on δ_3,

$$
\frac{1}{t}\|\boldsymbol{y}_t\|_2^2 \leq \frac{1}{t}(tR^2 + \|\boldsymbol{\epsilon}_t\|_2^2 + 2\sqrt{t}R\|\boldsymbol{\epsilon}_t\|_2) \leq C_3(\sigma_\varepsilon^2 + R^2) \tag{21}
$$

Then let $\boldsymbol{\theta}' = \boldsymbol{\theta}_0$ and plug in $\boldsymbol{\theta}_t^{(j+1)}$ and $\boldsymbol{\theta}_t^{(j)}$ in (19), by Lemma B.3, with probability at least $1 - \delta_4$,

$$
\begin{aligned}
\mathcal{L}_t(\boldsymbol{\theta}_t^{(j+1)}) - \mathcal{L}_t(\boldsymbol{\theta}_0) &\leq (1 - \eta m\lambda/2)(\mathcal{L}_t(\boldsymbol{\theta}_t^{(j)}) - \mathcal{L}_t(\boldsymbol{\theta}_0)) + \frac{\eta m\lambda}{2}\mathcal{L}_t(\boldsymbol{\theta}_0) + (\frac{2\eta m\lambda}{t} + \frac{2}{\eta m\lambda t} + \frac{1}{t})\|\boldsymbol{e}_t\|_2^2 \\
&\leq (1 - \eta m\lambda/2)(\mathcal{L}_t(\boldsymbol{\theta}_t^{(j)}) - \mathcal{L}_t(\boldsymbol{\theta}_0)) + \frac{\eta m\lambda}{2}(\frac{1}{t}\|\boldsymbol{y}_t\|_2^2 + m\lambda\|\boldsymbol{\theta}_0\|_2^2) \\
&\quad + (2\eta m\lambda + 2/\eta m\lambda + 1)C_2^2\tau^{8/3}L^6 m\log(m) \quad \text{(by (21))} \\
&\leq (1 - \eta m\lambda/2)(\mathcal{L}_t(\boldsymbol{\theta}_t^{(j)}) - \mathcal{L}_t(\boldsymbol{\theta}_0)) + \frac{\eta m\lambda}{2}(C_3(\sigma_\varepsilon^2 + R^2) + m\lambda\|\boldsymbol{\theta}_0\|_2^2) \\
&\quad + \frac{5}{\eta m\lambda}C_2^2\tau^{8/3}L^6 m\log(m) \quad \text{(by (20) and } \eta m\lambda \leq 1) \\
&\leq (1 - \eta m\lambda/2)(\mathcal{L}_t(\boldsymbol{\theta}_t^{(j)}) - \mathcal{L}_t(\boldsymbol{\theta}_0)) + C_4\eta m\lambda(\sigma_\varepsilon^2 + R^2) + \frac{5}{\eta m\lambda}C_2^2\tau^{8/3}L^6 m\log(m) \\
&\quad \text{(by Lemma B.3)}
\end{aligned}
$$

Now we further set $\tau = \tilde{C}\sqrt{\frac{\sigma_\varepsilon^2+R^2}{m\lambda}}$ and the upper bound for $\mathcal{L}_t(\boldsymbol{\theta}_t^{(j+1)}) - \mathcal{L}_t(\boldsymbol{\theta}_0)$ is

$$\begin{aligned}
\mathcal{L}_t(\boldsymbol{\theta}_t^{(j+1)}) - \mathcal{L}_t(\boldsymbol{\theta}_0) &\leq (1-\eta m\lambda/2)(\mathcal{L}_t(\boldsymbol{\theta}_t^{(j)}) - \mathcal{L}_t(\boldsymbol{\theta}_0)) + C_4\eta m\lambda(\sigma_\varepsilon^2+R^2) + \frac{5}{\eta m\lambda}\tilde{C}^2C_2^2(\sigma_\varepsilon^2+R^2)\tau^{2/3}\lambda^{-1}L^6\log(m) \\
&\quad (\text{by } \tau = \tilde{C}\sqrt{\frac{\sigma_\varepsilon^2+R^2}{m\lambda}}) \\
&\leq (1-\eta m\lambda/2)(\mathcal{L}_t(\boldsymbol{\theta}_t^{(j)}) - \mathcal{L}_t(\boldsymbol{\theta}_0)) + C_4\eta m\lambda(\sigma_\varepsilon^2+R^2) + C_5\eta m\lambda(\sigma_\varepsilon^2+R^2) \\
&\quad (\text{by choice of } \tau \text{ in Lemma C.7})
\end{aligned}$$

where C_4 is a constant depends on δ_3 and δ_4 and C_5 depends on δ_2, δ_3 and δ_4. Then by recursion,

$$\mathcal{L}_t(\boldsymbol{\theta}_t^{(j+1)}) - \mathcal{L}_t(\boldsymbol{\theta}_0) \leq \frac{C_6\eta m\lambda(\sigma_\varepsilon^2+R^2)}{\eta m\lambda/2} = \tilde{C}_6(\sigma_\varepsilon^2+R^2)$$

where $C_6 = C_4 + C_5$ and $\tilde{C}_6 = 2C_6$. Recall that $\|\boldsymbol{f}_t(\boldsymbol{\theta}) - \boldsymbol{y}_t\|_2^2 = 2t\mathcal{L}_t(\boldsymbol{\theta}) - \frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2 \leq 2t\mathcal{L}_t(\boldsymbol{\theta})$, with some constant C_7 derived from C_6 and C_4, then we have

$$\begin{aligned}
\|\boldsymbol{f}_{gnn,t}^{(j)} - \boldsymbol{y}_t\|_2^2 \leq 2t\mathcal{L}_t(\boldsymbol{\theta}_t^{(j)}) &\leq 2t\tilde{C}_6(\sigma_\varepsilon^2+R^2) + 2t\mathcal{L}_t(\boldsymbol{\theta}_0) \\
&= 2t\tilde{C}_6(\sigma_\varepsilon^2+R^2) + 2t(\frac{1}{t}\|\boldsymbol{y}_t\|_2^2 + \frac{m\lambda}{2}\|\boldsymbol{\theta}_0\|_2^2) \\
&\leq C_7t(\sigma_\varepsilon^2+R^2) \quad (\text{by Lemma B.3})
\end{aligned}$$

which implies our result by setting $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta/4$ where $\delta \in (0,1)$ is arbitrary small.

□

Lemma B.3 (Parameter Bound for Proximal Optimization)**.** *Let $\{\tilde{\boldsymbol{\theta}}_t^{(j)}\}_{j=1}^J$ be the gradient descent update sequence of parameters of the following optimization,*

$$\min_{\boldsymbol{\theta}} \frac{1}{2t}\sum_{i=1}^t(\langle g(G_i;\boldsymbol{\theta}_0), \boldsymbol{\theta}-\boldsymbol{\theta}_0\rangle - y_i)^2 + \frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2$$

Then if $m \geq poly(L, \lambda^{-1}, \log(N/\delta))$ and learning rate $\eta \leq (\tilde{C}mL + m\lambda)^{-1}$ for some constant $\tilde{C}$. Then for some constant C and for any $\forall t \in [T]$ and $\forall j \in [J]$, with probability at least $1-\delta \in (0,1)$,

$$\begin{aligned}
\|\tilde{\boldsymbol{\theta}}_t^{(j)}\|_2 &\leq C\sqrt{\frac{\sigma_\varepsilon^2+R^2}{m\lambda}} \\
\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_0\|_2 &\leq C\sqrt{\frac{\sigma_\varepsilon^2+R^2}{m\lambda}} \\
\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_0 - \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m\|_2 &\leq C(2-\eta m\lambda)^j\sqrt{\frac{\sigma_\varepsilon^2+R^2}{m\lambda}}
\end{aligned}$$

for some constant C which is independent of m and t.

Proof. Denote $\mathcal{L}_t(\boldsymbol{\theta}) := \frac{1}{2t}\sum_{i=1}^t(\langle g(G_i;\boldsymbol{\theta}_0), \boldsymbol{\theta}-\boldsymbol{\theta}_0\rangle - y_i)^2 + \frac{m\lambda}{2}\|\boldsymbol{\theta}\|_2^2$ as the loss function in our proximal optimization. By Lemma B.4, with probability at least $1-\delta_1 \in (0,1)$ the Hessian of $\mathcal{L}_t(\boldsymbol{\theta})$ satisfies:

$$\boldsymbol{0} \prec \nabla^2\mathcal{L}_t = \bar{\boldsymbol{G}}_t\bar{\boldsymbol{G}}_t^\top/t + m\lambda\boldsymbol{I} \preccurlyeq (\|\bar{\boldsymbol{G}}_t\|_F^2/t + m\lambda)\boldsymbol{I} \preccurlyeq (C_1^2mL + m\lambda)\boldsymbol{I}$$

which reveals that $\mathcal{L}_t$ is strongly convex and $(C_1^2mL+m\lambda)$-smooth. Thus if $\eta \leq (C_1^2mL+m\lambda)^{-1}$, $\mathcal{L}_t$ is a monotonically decreasing function:

$$\frac{1}{2t}\|\bar{\boldsymbol{G}}_t^\top(\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_0) - \boldsymbol{y}_t\|_2^2 + \frac{m\lambda}{2}\|\tilde{\boldsymbol{\theta}}_t^{(j)}\|_2^2 \leq \frac{1}{2t}\|\boldsymbol{y}_t\|_2^2 + \frac{m\lambda}{2}\|\boldsymbol{\theta}_0\|_2^2$$

which indicates

$$\begin{aligned}\|\tilde{\boldsymbol{\theta}}_t^{(j)}\|_2^2 \leq & \frac{1}{tm\lambda}\|\boldsymbol{y}_t\|_2^2 + \|\boldsymbol{\theta}_0\|_2^2 \\ \leq & \frac{1}{tm\lambda}(\|\boldsymbol{\mu}_t\|_2^2 + \|\boldsymbol{\epsilon}_t\|_2^2 + 2\|\boldsymbol{\mu}_t\|_2\|\boldsymbol{\epsilon}_t\|_2) + \|\boldsymbol{\theta}_0\|_2^2\end{aligned}$$

Note that the proximal optimization is optimization for ridge regression which has the closed form solution:

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}_0 + \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m$$

and $\tilde{\boldsymbol{\theta}}_t^{(j)}$ converges to $\boldsymbol{\theta}^*$ with the following rate:

$$\begin{aligned}\|\tilde{\boldsymbol{\theta}}_t^{(j+1)} - \boldsymbol{\theta}^*\|_2^2 &= \|\tilde{\boldsymbol{\theta}}_t^{(j)} - \eta\nabla\mathcal{L}(\tilde{\boldsymbol{\theta}}_t^{(j)}) - \boldsymbol{\theta}^*\|_2^2 \\ &= \|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2 + \eta^2\|\nabla\mathcal{L}(\tilde{\boldsymbol{\theta}}_t^{(j)})\|_2^2 - 2\eta(\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*)^\top\nabla\mathcal{L}(\tilde{\boldsymbol{\theta}}_t^{(j)}) \\ &\leq \|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2 + \eta^2(C_1^2mL + m\lambda)^2\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2 - 2\eta(\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*)^\top\nabla\mathcal{L}(\tilde{\boldsymbol{\theta}}_t^{(j)}) \quad \text{(by smoothness)} \\ &\leq \|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2 + \eta^2(C_1^2mL + m\lambda)^2\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2 + 2\eta(\mathcal{L}(\boldsymbol{\theta}^*) - \mathcal{L}(\tilde{\boldsymbol{\theta}}_t^{(j)})) \quad \text{(by convexity)} \\ &\leq 2\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2 + 2\eta(\mathcal{L}(\boldsymbol{\theta}^*) - \mathcal{L}(\tilde{\boldsymbol{\theta}}_t^{(j)})) \quad \text{(by } \eta \leq (C_1^2mL + m\lambda)^{-1}\text{)} \\ &\leq 2\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2 - \eta m\lambda\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2 \quad \text{(by } m\lambda\text{-strongly convexity)} \\ &= (2 - \eta m\lambda)\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}^*\|_2^2\end{aligned}$$

Therefore,

$$\begin{aligned}\|\tilde{\boldsymbol{\theta}}_t^{(j+1)} - \boldsymbol{\theta}^*\|_2^2 &\leq (2 - \eta m\lambda)^j\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*\|_2^2 \\ &\leq (2 - \eta m\lambda)^j\frac{2}{m\lambda}(\mathcal{L}(\boldsymbol{\theta}_0)) - \mathcal{L}(\boldsymbol{\theta}^*)) \quad \text{(by } m\lambda\text{-strongly convexity)} \\ &\leq (2 - \eta m\lambda)^j\frac{2}{m\lambda}\mathcal{L}(\boldsymbol{\theta}_0) \\ &= (2 - \eta m\lambda)^j\left(\frac{1}{tm\lambda}\|\boldsymbol{y}_t\|_2^2 + \|\boldsymbol{\theta}_0\|_2^2\right)\end{aligned}$$

Then combine with Lemma C.4 and $\|\boldsymbol{\mu}_t\|_2 \leq \sqrt{t}\|\mu\|_{\mathcal{H}} \leq \sqrt{t}R$, we have that with probability at least $1 - \delta_2 \in (0, 1)$,

$$\frac{1}{tm\lambda}\|\boldsymbol{y}_t\|_2^2 \leq \frac{1}{tm\lambda}(tR^2 + \|\boldsymbol{\epsilon}_t\|_2^2 + 2\sqrt{t}R\|\boldsymbol{\epsilon}_t\|_2) \leq \tilde{C}_1(\sigma_\varepsilon^2 + R^2)/m\lambda$$

where $\tilde{C}_1$ is some constant depends on δ_2. Therefore, for any $\delta \in (0, 1)$, set $\delta_1 = \delta_2 = \delta/2$, with probability at least $1 - \delta_2$,

$$\|\tilde{\boldsymbol{\theta}}_t^{(j)}\|_2 \leq \tilde{C}_2\sqrt{\frac{\sigma_\varepsilon^2 + R^2}{m\lambda}}$$

$$\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_0\|_2 \leq \tilde{C}_2\sqrt{\frac{\sigma_\varepsilon^2 + R^2}{m\lambda}}$$

and

$$\|\tilde{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_0 - \bar{\boldsymbol{U}}_t^{-1}\bar{\boldsymbol{G}}_t\boldsymbol{y}_t/m\|_2 \leq (2 - \eta m\lambda)^j\tilde{C}_2\sqrt{\frac{\sigma_\varepsilon^2 + R^2}{m\lambda}}$$

where $\tilde{C}_2$ is some constant depends on δ_2 and $\|\boldsymbol{\theta}_0\|_2$.

□

Lemma B.4 (Gradient Descent Norm Bound)**.** *Define* $\boldsymbol{G}_t^{(j)} := [g(G_1; \boldsymbol{\theta}_t^{(j)}), ..., g(G_t; \boldsymbol{\theta}_t^{(j)})] \in \mathbb{R}^{p\times t}$ *for the gradients in the* j*-th updates in GNN training (optimization of* (6)*) at round* t*. Also define* $\boldsymbol{f}_{gnn,t}^{(j)} := [f_{\text{GNN}}(G_1; \boldsymbol{\theta}_t^{(j)}), ..., f_{\text{GNN}}(G_t; \boldsymbol{\theta}_t^{(j)})]^\top \in \mathbb{R}^{t\times 1}$*. Assume* τ *is set such that* $\|\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0\|_2 \leq \tau$

for all t and $\forall j \leq J$. Suppose $m \geq poly(L, \lambda^{-1}, \log(N/\delta))$ where $\delta \in (0,1)$, then with probability at least $1-\delta$,

$$\|\bar{\boldsymbol{G}}_t\|_F \leq C_1\sqrt{tmL}$$
$$\|\boldsymbol{G}_t^{(j)}\|_F \leq C_1\sqrt{tmL}$$
$$\|\bar{\boldsymbol{G}}_t - \boldsymbol{G}_t^{(j)}\|_F \leq C_2\tau^{1/3}L^{7/2}\sqrt{tm\log(m)}$$
$$\|\boldsymbol{f}_{gnn,t}^{(j)} - (\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0)^\top \bar{\boldsymbol{G}}_t\|_2 \leq C_3\tau^{4/3}L^3\sqrt{tm\log(m)}$$

for some constant C_1, C_2, C_3 which does not depend on m and t.

Proof. From Lemma C.7, we can bounding the $\|g(G;\boldsymbol{\theta}_0)\|_2$ with probability at least $1-\delta \in (0,1)$, which provides the high probability upper bound for the Frobenius norm of $\bar{\boldsymbol{G}}_t$:

$$\|\bar{\boldsymbol{G}}_t\|_F \leq \sqrt{t}\max_{i\in[t]}\|g(G_i;\boldsymbol{\theta}_0)\|_2 \leq \frac{\sqrt{t}}{N}\max_{i\in[t]}\sum_{j\in\mathcal{V}(G_i)}\|g_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta}_0)\|_2 \leq C_1\sqrt{tmL}$$

and the high probability upper bound for the Frobenius norm of $\boldsymbol{G}_t^{(j)}$:

$$\|\boldsymbol{G}_t^{(j)}\|_F \leq \sqrt{t}\max_{i\in[t]}\|g(G_i;\boldsymbol{\theta}_t^{(j)})\|_2 \leq \frac{\sqrt{t}}{N}\max_{i\in[t]}\sum_{j\in\mathcal{V}(G_i)}\|g_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta}_t^{(j)})\|_2 \leq C_1\sqrt{tmL}$$

For the gradients difference, by Lemma C.7, with probability at least $1-\delta$,

$$\begin{aligned}
\|\bar{\boldsymbol{G}}_t - \boldsymbol{G}_t^{(j)}\|_F &\leq \sqrt{t}\max_{i\in[t]}\|g(G_i;\boldsymbol{\theta}_0) - g(G_i;\boldsymbol{\theta}_t^{(j)})\|_2 \\
&\leq \frac{\sqrt{t}}{N}\max_{i\in[t]}\sum_{j\in\mathcal{V}(G_i)}\|g_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta}_0) - g_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta}_t^{(j)})\|_2 \\
&\leq C_2\tau^{1/3}L^{7/2}\sqrt{tm\log(m)}
\end{aligned}$$

The last norm for difference between the GNN prediction and linearized prediction is bounded due to Lemma C.7, with probability at least $1-\delta$,

$$\begin{aligned}
\|\boldsymbol{f}_{gnn,t}^{(j)} - (\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0)^\top \boldsymbol{G}_t^{(j)}\|_2 &\leq \sqrt{t}\max_{i\in[t]}|f_{\text{GNN}}(G_i;\boldsymbol{\theta}_t^{(j)}) - (\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0)^\top g(G_i;\boldsymbol{\theta}_0)| \\
&\leq \frac{\sqrt{t}}{N}\max_{i\in[t]}\sum_{j\in\mathcal{V}(G_i)}|f_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta}_t^{(j)}) - (\boldsymbol{\theta}_t^{(j)} - \boldsymbol{\theta}_0)^\top g_{\text{MLP}}(\boldsymbol{h}_j;\boldsymbol{\theta}_0)| \\
&\leq C_3\tau^{4/3}L^3\sqrt{tm\log(m)}
\end{aligned}$$

□

B.4 Lemmas for GNTK

Lemma B.5 (Approximation from GNTK)**.** *Set $\delta \in (0,1)$ and*

$$m = \Omega(L^{10}T^4|\mathcal{G}|^6\rho_{min}^{-4}\log(LN^2|\mathcal{G}|^2/\delta)).$$

Then with probability at least $1-\delta$,
(i) (Approximate Linearized Nerual Network) $\exists\boldsymbol{\theta}^*$ such that, for $\forall G \in \mathcal{G}$

$$\mu(G) = \langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}^*\rangle$$
$$\sqrt{m}\|\boldsymbol{\theta}^*\|_2 \leq \sqrt{2}R$$

(ii) (Spectral Bound for Uncertainty Matrix $\bar{\boldsymbol{U}}_t$ by GNTK)

$$\lambda_{max}(\bar{\boldsymbol{U}}_t) \leq \lambda + \frac{3}{2}\rho_{\max}$$
$$\log\det(\lambda^{-1}\bar{\boldsymbol{U}}_t) \leq \log\det(\boldsymbol{I}_{|\mathcal{G}|} + \lambda^{-1}t\boldsymbol{K}) + 1$$

Proof. In this proof, set $\delta_1 = \delta_2 = \delta/2$ where $\delta \in (0,1)$ is an arbitrary real value. Recall the definition of the true reward function $\mu : \mathcal{G} \to \mathbb{R}$ and the GNTK matrix $\boldsymbol{K} \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$. We further define the vector of function values $\boldsymbol{\mu} \in \mathbb{R}^{|\mathcal{G}|\times 1}$ as well as the gradient matrix $\bar{\boldsymbol{G}} \in \mathbb{R}^{p\times|\mathcal{G}|}$ on initialization $\boldsymbol{\theta}_0$.

$$
\begin{aligned}
[\boldsymbol{K}]_{ij} &= k(G^i, G^j) \quad \forall G^i, G^j \in \mathcal{G} \\
[\boldsymbol{\mu}]_i &= \mu(G^i) \quad \forall G^i \in \mathcal{G} \\
\bar{\boldsymbol{G}}_{*i} &= g(G^i; \boldsymbol{\theta}_0)
\end{aligned}
$$

Proof for (i): By the connection between GNTK and NTK,

$$
\begin{aligned}
\|\boldsymbol{K} - \bar{\boldsymbol{G}}^\top \bar{\boldsymbol{G}}/m\|_F &= \sqrt{\sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|}(k(G^i, G^j) - g^\top(G^i;\boldsymbol{\theta}_0)g(G^j;\boldsymbol{\theta}_0)/m)^2} \\
&= \sqrt{\sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|}\left(\frac{1}{N^2}\sum_{u\in\mathcal{V}(G^i)}\sum_{v\in\mathcal{V}_{G^j}}(k_{\text{MLP}}(\boldsymbol{h}_u^{G^i}, \boldsymbol{h}_v^{G^j}) - g_{\text{MLP}}^\top(\boldsymbol{h}_u^{G^i};\boldsymbol{\theta}_0)g_{\text{MLP}}(\boldsymbol{h}_v^{G^j};\boldsymbol{\theta}_0)/m)\right)^2} \\
&\le \sqrt{\sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|}\sum_{u\in\mathcal{V}(G^i)}\sum_{v\in\mathcal{V}(G^j)}(k_{\text{MLP}}(\boldsymbol{h}_u^{G^i}, \boldsymbol{h}_v^{G^j}) - g_{\text{MLP}}^\top(\boldsymbol{h}_u^{G^i};\boldsymbol{\theta}_0)g_{\text{MLP}}(\boldsymbol{h}_v^{G^j};\boldsymbol{\theta}_0)/m)^2}
\end{aligned}
$$

where $\mathcal{V}_G$ denotes the vertice set of a graph G. By Lemma C.6, when $m = \Omega(L^{10}N^4|\mathcal{G}|^4\rho_{min}^{-4}\log(LN^2|\mathcal{G}|^2/\delta_1)$, then with probability at least $1-\delta_1/(N^2|\mathcal{G}|^2)$, $|k_{\text{MLP}}(\boldsymbol{h}_u^{G^i}, \boldsymbol{h}_v^{G^j}) - g_{\text{MLP}}^\top(\boldsymbol{h}_u^{G^i};\boldsymbol{\theta}_0)g_{\text{MLP}}(\boldsymbol{h}_v^{G^j};\boldsymbol{\theta}_0)/m| \le \frac{\rho_{min}}{2N|\mathcal{G}|}$. Then apply union bound over all pairs $(\boldsymbol{h}_u^{G^i}, \boldsymbol{h}_v^{G^j})$, the following holds with probability at least $1-\delta_1$,

$$
\|\boldsymbol{K} - \bar{\boldsymbol{G}}^\top \bar{\boldsymbol{G}}/m\|_F \le \rho_{min}/2
$$

which shows that

$$
\begin{aligned}
\bar{\boldsymbol{G}}^\top \bar{\boldsymbol{G}}/m &\succcurlyeq \boldsymbol{K} - \|\boldsymbol{K} - \bar{\boldsymbol{G}}^\top \bar{\boldsymbol{G}}/m\|_2 \boldsymbol{I}_{|\mathcal{G}|} \\
&\succcurlyeq \boldsymbol{K} - \|\boldsymbol{K} - \bar{\boldsymbol{G}}^\top \bar{\boldsymbol{G}}/m\|_F \boldsymbol{I}_{|\mathcal{G}|} \\
&\succcurlyeq \boldsymbol{K} - \frac{\rho_{min}}{2}\boldsymbol{I}_{|\mathcal{G}|} \\
&\succcurlyeq \boldsymbol{K}/2 \succ \boldsymbol{0}
\end{aligned} \tag{22}
$$

Suppose $\bar{\boldsymbol{G}} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{Q}^\top$ is the decomposition of $\bar{\boldsymbol{G}}$ where $\boldsymbol{P} \in \mathbb{R}^{p\times|\mathcal{G}|}$, $\boldsymbol{Q} \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$ are unitary and $\boldsymbol{\Lambda} \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$. By (22), we know $\boldsymbol{\Lambda} \succ \boldsymbol{0}$ with probability at least $1-\delta_1$. Now denote $\boldsymbol{\theta}^* = \boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^\top\boldsymbol{\mu}$ and it satisfies

$$
\begin{aligned}
\bar{\boldsymbol{G}}^\top\boldsymbol{\theta}^* &= \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^\top\boldsymbol{\mu} = \boldsymbol{\mu} \\
\Rightarrow \mu(G) &= \langle g(G;\boldsymbol{\theta}_0), \boldsymbol{\theta}^*\rangle \quad \forall G \in \mathcal{G}
\end{aligned}
$$

Moreover, the norm of $\boldsymbol{\theta}^*$ is also bounded:

$$
\|\boldsymbol{\theta}^*\|_2^2 = \boldsymbol{\mu}^\top\boldsymbol{Q}\boldsymbol{\Lambda}^{-2}\boldsymbol{Q}^\top\boldsymbol{\mu} = \boldsymbol{\mu}^\top(\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}})^{-1}\boldsymbol{\mu} \le \frac{2}{m}\boldsymbol{\mu}^\top\boldsymbol{K}^{-1}\boldsymbol{\mu} \le \frac{2R^2}{m}
$$

which completes our proof for (i).

Proof for (ii): From the definition of $\bar{\boldsymbol{G}}_t$, we have

$$\begin{aligned}
\log\det(\boldsymbol{I}_{|\mathcal{G}|} + \lambda^{-1}\bar{\boldsymbol{G}}_t^\top \bar{\boldsymbol{G}}_t/m) &= \log\det\left(\boldsymbol{I}_{|\mathcal{G}|} + \sum_{i=1}^{t} g(G_i;\boldsymbol{\theta}_0)g^\top(G_i;\boldsymbol{\theta}_0)/(m\lambda)\right) \\
&\leq \log\det\left(\boldsymbol{I}_{|\mathcal{G}|} + t\sum_{G\in\cup_{i=1}^{t}\mathcal{G}_i} g(G;\boldsymbol{\theta}_0)g^\top(G;\boldsymbol{\theta}_0)/(m\lambda)\right) \\
&\leq \log\det\left(\boldsymbol{I}_{|\mathcal{G}|} + t\sum_{G\in\mathcal{G}} g(G;\boldsymbol{\theta}_0)g^\top(G;\boldsymbol{\theta}_0)/(m\lambda)\right) \quad (\text{by } \mathcal{G}_t \in \mathcal{G} \text{ for } \forall t \in [T]) \\
&= \log\det(\boldsymbol{I}_{|\mathcal{G}|} + t\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/(m\lambda)) \\
&= \log\det(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda + t(\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m - \boldsymbol{K})/\lambda) \\
(\text{by concavity of } \log\det(\cdot)) \quad &\leq \log\det(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda) + \langle(\boldsymbol{I} + t\boldsymbol{K}/\lambda)^{-1}, t(\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m - \boldsymbol{K})/\lambda\rangle_F \\
&\leq \log\det(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda) + \|(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda)^{-1}\|_F \|t(\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m - \boldsymbol{K})/\lambda\|_F \\
&\leq \log\det(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda) + t\sqrt{|\mathcal{G}|}\|(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda)^{-1}\|_2 \|\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m - \boldsymbol{K}\|_F/\lambda \\
&= \log\det(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda) + \sqrt{|\mathcal{G}|}(\lambda/t + \rho_{min})^{-1}\|\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m - \boldsymbol{K}\|_F
\end{aligned}$$

By Lemma C.6, when $m = \Omega(L^{10}N^4|\mathcal{G}|^6\rho_{min}^{-4}\log(LN^2|\mathcal{G}|^2/\delta_2)$, then with probability at least $1 - \delta_2/(N^2|\mathcal{G}|^2)$, $|k_{\text{MLP}}(\boldsymbol{h}_u^{G^i}, \boldsymbol{h}_v^{G^j}) - g_{\text{MLP}}^\top(\boldsymbol{h}_u^{G^i};\boldsymbol{\theta}_0)g_{\text{MLP}}(\boldsymbol{h}_v^{G^j};\boldsymbol{\theta}_0)/m| \leq \frac{\rho_{min}}{N|\mathcal{G}|^{3/2}}$. Then apply union bound over all pairs $(\boldsymbol{h}_u^{G^i}, \boldsymbol{h}_v^{G^j})$, with probability at least $1 - \delta_2$, $\|\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m - \boldsymbol{K}\|_F \leq \frac{\rho_{min}}{\sqrt{|\mathcal{G}|}}$, which indicates that

$$\begin{aligned}
\log\det(\boldsymbol{I}_{|\mathcal{G}|} + \lambda^{-1}\bar{\boldsymbol{G}}_t^\top\bar{\boldsymbol{G}}_t/m) &\leq \log\det(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda) + \sqrt{|\mathcal{G}|}(\lambda/t + \rho_{min})^{-1}\|\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m - \boldsymbol{K}\|_F \\
&\leq \log\det(\boldsymbol{I}_{|\mathcal{G}|} + t\boldsymbol{K}/\lambda) + 1
\end{aligned}$$

Finally, with probability at least $1 - \delta_1$,

$$\bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m \preccurlyeq \boldsymbol{K} + \|\boldsymbol{K} - \bar{\boldsymbol{G}}^\top\bar{\boldsymbol{G}}/m\|_2\boldsymbol{I}_{|\mathcal{G}|} \preccurlyeq \boldsymbol{K} + \frac{\rho_{\max}}{2}\boldsymbol{I}_{|\mathcal{G}|} \preccurlyeq \frac{3}{2}\rho_{\max}\boldsymbol{I}_{|\mathcal{G}|}$$

which indicates that $\lambda_{max}(\bar{\boldsymbol{U}}_t) \leq \lambda + \frac{3}{2}\rho_{\max}$. □

Lemma B.6. *Fix $\delta \in (0,1)$. Then, for $m = \Omega(L^{10}|\mathcal{G}|^4\varepsilon^{-4}\log(L/\delta))$, with probability at least $1 - \delta$,*

$$|\rho_{\max} - \hat{\rho}_{\max}| \leq \varepsilon.$$

Proof. Let m be as in Lemma C.6. Recall that $\|\boldsymbol{h}_u^G\| = 1$ for all $u \in \mathcal{V}(G)$ and $G \in \mathcal{G}$, by construction. Let $N_i := |\mathcal{V}(G^i)|$. Then, we have, with probability at least $1 - \delta$,

$$\begin{aligned}
&|k(G^i, G^j) - \hat{k}(G^i, G^j)| \\
&\quad\leq \frac{1}{N_iN_j}\sum_{\substack{u\in\mathcal{V}(G^i)\\ v\in\mathcal{V}(G^j)}} \left|k_{\text{MLP}}(\boldsymbol{h}_u^{G^i}, \boldsymbol{h}_v^{G^j}) - g_{\text{MLP}}(\boldsymbol{h}_u^{G^i};\boldsymbol{\theta}_0)^\top g_{\text{MLP}}(\boldsymbol{h}_v^{G^j};\boldsymbol{\theta}_0)/m\right| \leq \varepsilon
\end{aligned}$$

by Lemma C.6. Then

$$\|\boldsymbol{K} - \hat{\boldsymbol{K}}\|_{\text{op}} \leq \|\boldsymbol{K} - \hat{\boldsymbol{K}}\|_F \leq |\mathcal{G}|\varepsilon.$$

Then, from Weyl's inequality, $|\rho_{\max} - \hat{\rho}_{\max}| \leq |\mathcal{G}|\varepsilon$. Replacing ε with $\varepsilon/|\mathcal{G}|$ the result follows. □

C Supporting Lemmas

Lemma C.1. *Suppose $\boldsymbol{a}$, $\boldsymbol{b}$ are vectors and $\boldsymbol{A}$ is a matrix. c is assumed to be positive scalar. Then we have the following results: (i) $|\boldsymbol{a}^\top\boldsymbol{A}\boldsymbol{b}| \leq \sqrt{\boldsymbol{a}^\top\boldsymbol{A}\boldsymbol{a}}\sqrt{\boldsymbol{b}^\top\boldsymbol{A}\boldsymbol{b}}$. (ii) $\boldsymbol{a}^\top\boldsymbol{b} + c\|\boldsymbol{a}\|_2^2 \geq -\|\boldsymbol{b}\|_2^2/4c$.*

Lemma C.2. *Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ and $\beta > 0$, then*

$$\mathbb{P}(|X - \mu| \leq \beta\sigma) \geq 1 - e^{-\beta^2/2}$$

Lemma C.3. *Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ and $\beta > 0$, then*

$$\mathbb{P}(X - \mu > \beta\sigma) \geq \frac{e^{-\beta^2}}{4\beta\sqrt{\pi}}$$

Lemma C.4. *Suppose $\boldsymbol{\epsilon} \in \mathbb{R}^t$ is a subgaussian random vector with subgaussian constant σ^2, then*

$$\mathbb{E}[\|\boldsymbol{\epsilon}\|_2] \leq 4\sigma\sqrt{t}$$

and with probability at least $1 - \delta$ for $\delta \in (0, 1)$,

$$\|\boldsymbol{\epsilon}\|_2 \leq C\sigma\sqrt{t}.$$

where C is some constant depending on δ.

Lemma C.5. *(Theorem 1 (Chowdhury and Gopalan, 2017)) Let $\{\boldsymbol{x}_t\}_{t=1}^{\infty}$ be an $\mathbb{R}^d$-valued discrete time stochastic process that is predictable with respect to the filtration $\{\mathcal{F}_t\}_{t=1}^{\infty}$. Let $\{\varepsilon_t\}_{t=1}^{\infty}$ be a real-valued stochastic process and for any $\forall t$, ε_t is $\mathcal{F}_t$-measurable and subgaussian with constant R conditionally on $\mathcal{F}_{t-1}$. Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric positive-definite kernel. Then for any $\eta > 0$, $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\|\boldsymbol{\epsilon}_t\|^2_{((\boldsymbol{K}_t + \eta \boldsymbol{I}_t)^{-1} + \boldsymbol{I}_t)^{-1}} \leq R^2 \log\det((1 + \eta)\boldsymbol{I}_t + \boldsymbol{K}_t) + 2R^2 \log(1/\delta)$$

where $\boldsymbol{\epsilon}_t := (\varepsilon_1, ..., \varepsilon_t)^\top \in \mathbb{R}^t$ and $\boldsymbol{K}_t \in \mathbb{R}^{t \times t}$ is a matrix with $[\boldsymbol{K}_t]_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $1 \leq i, j \leq t$.

Lemma C.6 (Theorem 3.1 (Arora et al., 2019))**.** *Fix $\varepsilon > 0$ and $\delta \in (0, 1)$. Suppose a MLP $f_{\text{MLP}}(\cdot; \boldsymbol{\theta})$ with ReLU activation has L layers and width $m = \Omega(L^{10}\varepsilon^{-4}\log(L/\delta))$. Then for any input $\boldsymbol{x}$, $\boldsymbol{x}'$ such that $\|\boldsymbol{x}\|_2 \leq 1$, $\|\boldsymbol{x}'\|_2 \leq 1$, with probability at least $1 - \delta$,*

$$|k_{\text{MLP}}(\boldsymbol{x}, \boldsymbol{x}') - g_{\text{MLP}}(\boldsymbol{x}; \boldsymbol{\theta}_0)^\top g_{\text{MLP}}(\boldsymbol{x}'; \boldsymbol{\theta}_0)/m| \leq \varepsilon$$

where k_{MLP} is the neural tangent kernel associated with f_{MLP} and $g_{\text{MLP}}(\,\cdot\,; \boldsymbol{\theta}_0) = \nabla f_{\text{MLP}}(\,\cdot\,; \boldsymbol{\theta}_0)$.

Lemma C.7 (Lemma B.4/Lemma B.5/Lemma B.6 (Zhou et al., 2020) / Lemma C.4 (Zhang et al., 2020))**.** *Suppose $\boldsymbol{\theta}$ is parameters for a MLP $f_{\text{MLP}}(\cdot; \boldsymbol{\theta})$ with L layers and width m and this neural network $f_{\text{MLP}}(\cdot; \boldsymbol{\theta})$ is trained via gradient descent with initialization $\boldsymbol{\theta}_0$, learning rate η and ℓ_2 regularization constant λ in a mean squared loss. The input feature set is denoted as $\mathcal{X} = \{\boldsymbol{x}_i\}_{i \in [T]}$. Then there are positive constants $\{C_i\}_{i=1}^{7}$ such that for $\forall \delta \in (0, 1)$, if τ satisfies*

$$\begin{aligned} &\tau \geq C_1 m^{-3/2} L^{-3/2} max((\log(TL^2/\delta))^{3/2}, (\log(m))^{-3/2}) \\ &\tau \leq \min(C_2 L^{-6}(\log(m))^{-3/2}, C_3 L^{-9/2}(\log(m))^{-3}, C_4 m^3 \lambda^{9/2} \eta^3 L^{-9}(\log(m))^{-3/2}) \end{aligned} \tag{23}$$

then with probability at least $1 - \delta$, for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \leq \tau$ and $\|\boldsymbol{\theta}' - \boldsymbol{\theta}_0\|_2 \leq \tau$, for $\forall \boldsymbol{x} \in \mathcal{X}$, we have

$$\|g_{\text{MLP}}(\boldsymbol{x}; \boldsymbol{\theta}) - g_{\text{MLP}}(\boldsymbol{x}; \boldsymbol{\theta}_0)\|_2 \leq C_5 \sqrt{\log(m)} \tau^{1/3} L^3 \|g_{\text{MLP}}(\boldsymbol{x}; \boldsymbol{\theta}_0)\|_2$$

and

$$|f_{\text{MLP}}(\boldsymbol{x}; \boldsymbol{\theta}) - f_{\text{MLP}}(\boldsymbol{x}; \boldsymbol{\theta}') - \langle g_{\text{MLP}}(\boldsymbol{x}, \boldsymbol{\theta}'), \boldsymbol{\theta} - \boldsymbol{\theta}' \rangle| \leq C_6 \tau^{4/3} L^3 \sqrt{m \log(m)}$$

and

$$\|g_{\text{MLP}}(\boldsymbol{x}; \boldsymbol{\theta})\|_2 \leq C_7 \sqrt{mL}.$$

D Supplement to Experiments

D.1 Data Generation

We use synthetic data environments for our experiments. The datasets are generated from two different random graph models and three different reward function generating models. The random graph models are Erdös–Rényi random graph model and random dot product graph model. We use a linear model, Gaussian process with GNTK model, Gaussian process with representation kernel to generate our reward function. In all data environments, the feature dimension is set as $d = 10$. For any synthetic graph, all entries of the associated feature matrix $\{\boldsymbol{X}_{ji}\}_{j\in[N],i\in[d]}$ are i.i.d from a standard Gaussian distribution. The noisy reward is assumed to have standard deviation $\sigma_\varepsilon = 0.01$. All performance curves in our empirical studies show an average of over 10 repetitions with a standard deviation of the corresponding bandit problem with horizon $T = 1000$. Our experiment assumes the graph domain is fully observable, $\mathcal{G}_t = \mathcal{G}$ for all $t \in [T]$. We experiment four graph size $|\mathcal{G}| \in \{10, 50, 100, 200\}$ in the random dot product graphs with $N = 100$ and representation kernel.

D.1.1 Random Graph

Erdös–Rényi Random Graphs. Erdös–Rényi random graphs are generated by edge probability p and number of nodes N. Set the graph has N nodes and for any node pair $(i, j) \in [N]^2$, there is an edge linking i and j with probability p. We investigate $p \in \{0.2, 0.4, 0.6, 0.8\}$ and $N \in \{10, 50, 100, 500\}$ in our experiment. Including 3 types of reward function generating and 4 sizes of graph space $\mathcal{G}$, there are 192 combinations of datasets of Erdös–Rényi random graph environments.

Random Dot Product Graphs. Random dot product graphs are generated by modeling the expected edge probabilities as the function of the inner product of features. In our experiment, we set the latent embeddings observed as features, i.e. X_{i*} is the latent embedding of node i. Formally, the edge probability for node i and j is generated by $p_{ij} = \text{sigmoid}(\boldsymbol{X}_{i*}^\top \boldsymbol{X}_{j*})$. We also investigate $N \in \{10, 50, 100, 500\}$. Including 3 types of reward function generating and 4 sizes of graph space $\mathcal{G}$, there are 48 combinations of datasets of random dot product graph environments.

D.1.2 Reward Function Generation

Linear Model. We generate a true parameter $\boldsymbol{\theta}^* \in \mathbb{R}^d$ whose elements are i.i.d standard Gaussian. Then the true reward mean is

$$\mu(G) = \langle \boldsymbol{\theta}^*, \bar{\boldsymbol{h}}^G \rangle$$

where $\bar{\boldsymbol{h}}^G = \sum_{i=1}^N \boldsymbol{h}_i^G / N$.

Gaussian Process with GNTK. We also use Gaussian process and Graph Neural Tangent Kernel(GNTK) as introduced from experiment in (Kassraie et al., 2022). We approximately construct the GNTK matrix $\boldsymbol{K}$ by the empirical GNTK matrix $\hat{\boldsymbol{K}} \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$ whose entries are $\hat{\boldsymbol{K}}_{ij} = \frac{1}{m}\langle g(G^i; \boldsymbol{\theta}_0), g(G^j; \boldsymbol{\theta}_0)\rangle$ for any $G^i, G^j \in \mathcal{G}$. We use this empirical GNTK matrix $\hat{\boldsymbol{K}}$ as the covariance matrix of prior, i.e, $\mathcal{N}(0, \boldsymbol{K}^{gntk})$ and use $\{(G, y_G)\}_{G\in\mathcal{G}}$ where $\{y_G\}_{G\in\mathcal{G}}$ are i.i.d from $\mathcal{N}(0, 1)$ as our training data. To train this Gaussian process model, we use negative log-likelihood loss with Adam optimizer with learning rate 0.01 and 30 epochs. The true reward means are sampled from the posterior in this Gaussian process.

Gaussian Process with Representation Kernel. For the Gaussian process with representation kernel, we trained a GNN for a graph property prediction task and used the mean pooling over all nodes of the last layer representations as the graph representation. In our experiment, we utilize the average degree prediction as our task. That is, suppose outcome is $d_G = \frac{1}{N}\sum_{j=1}^N \deg(j)$ and train GNN in (2) to predict this outcome. Then denote the last layer representation as $\bar{\boldsymbol{h}}_{\text{rep}}^G = \frac{1}{N}\sum_{j=1}^N f^{(L-1)}(\boldsymbol{h}_j^G)$. Then we define the representation kernel as the inner product of the graph representations

$$k_{\text{rep}}(G, G') := \langle \bar{\boldsymbol{h}}_{\text{rep}}^G, \bar{\boldsymbol{h}}_{\text{rep}}^{G'} \rangle.$$

The associated kernel matrix is denoted as $\boldsymbol{K}^{rep} \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$ with entries $\{k^{rep}(G, G')\}_{G,G'\in mG}$. In this Gaussian process, we sample the true reward means by $\{\mu(G)\}_{G\in\mathcal{G}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}^{rep})$. To train this

Gaussian process model, we use MSE loss with Adam optimizer with learning rate 0.01 mini-batch size 2 and 30 epochs.

D.2 Algorithms Set Up

We provide the practical details and set up on our proposed algorithms and baseline algorithms.

Algorithms. We investigate 3 GNN-based bandit algorithms (`GNN-TS`, `GNN-UCB` and `GNN-PE`) and 3 corresponding NN-based bandit algorithms (`NN-TS`, `NN-UCB` and `NN-PE`). All algorithms in our work use the loss function (6) which is different from previous work. All gradients used for in our experiments are $g(G;\boldsymbol{\theta}_t)$ not $g(G;\boldsymbol{\theta}_0)$ unless special stated. In addition, in order to show the benefit of considering the graph structure, we include `NN-UCB`, `NN-TS`, `NN-PE` as our baselines. For this NN-based algorithm, we ignore the adjacency matrix for a graph (assume $\boldsymbol{A} = \boldsymbol{I}$), and pass through the model in (1) and (2) by $\boldsymbol{h}_i^G = \boldsymbol{X}_{i*}$. For `GNN-TS`, we tuned the exploration scale with grid search on $\nu \in \{0.01, 0.1, 1.0, 10.0\}$ and `NN-TS` follows the same value. For `GNN-UCB`, we tuned the hyperparameter with grid search on $\beta \in \{0.01, 0.1, 1.0, 10.0\}$ and `NN-UCB` follows the same value. For `GNN-PE`, we tuned the hyperparameter with grid search on $\beta \in \{0.01, 0.1, 1.0, 10.0\}$ and `NN-PE` follows the same value. All the hyperparameter tuning is performed in Erdös–Rényi random graphs with $p = 0.4$, $N = 50$, $|\mathcal{G}| = 100$ and Gaussian process with GNTK for all the Erdös–Rényi random graphs settings and random dot product graphs with 50 nodes and $|\mathcal{G}| = 100$ and Gaussian process with GNTK for all the random dot product graphs settings.

Neural Networks. The MLPs in our experiments have 2 layers ($L = 2$) and width $m = 512$. We use SGD optimizer with mini-batch size 5 and 30 epochs. Learning rates (η) we tuned from and the regularization hyperparameters λ we tuned from $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. Initialization for the trainable GNN parameter $\boldsymbol{\theta}$ satisfies the condition $f_{\text{GNN}}(G;\boldsymbol{\theta}_0) = 0$ for all $G \in \mathcal{G}$, which is handle by the treatment in Kassraie and Krause (2022). Suppose the initialization is $\boldsymbol{\theta}_0$. The matrix inversion in the algorithms is approximated by diagonal inversion across all policy algorithms.

D.3 Experiments on Scalability ($|\mathcal{G}|$)

We set the size of the graph domain to $|\mathcal{G}| = 100$ in Figure 1 and we experiment across different sizes $|\mathcal{G}| \in \{10, 50, 100, 200\}$ to check the scalability of the algorithms. Figure 2 shows that given a fixed horizon length, larger $|\mathcal{G}|$ leads to a harder bandit problem. It also shows that `GNN-TS` can achieve top performance across all algorithms in all scales of the graph space. This empirical observation shows that `GNN-TS` is robust to the scalability of the action space, supporting our theoretical justification in Section 4.

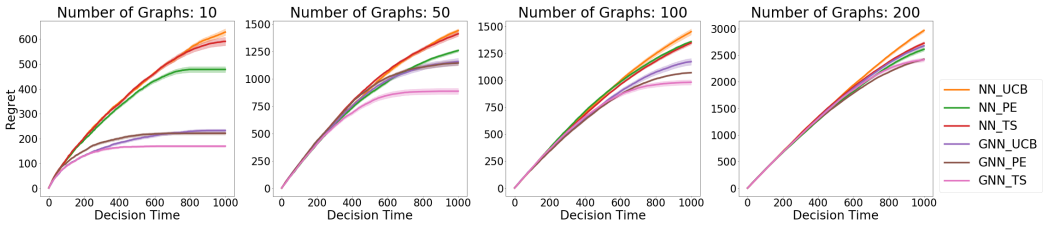

Figure 2: Competitive performance of `GNN-TS` is consistent across different sizes of graph space.

D.4 Effect of m and Initial Gradients

Our regret analysis depends on the assumption that the width of the neural network m must be large enough. We conduct an experiment to observe the effect from the width which is chosen from $\{32, 128, 512, 2048\}$. As some previous works on Neural bandit use the gradients at initialization ($g(G_t;\boldsymbol{\theta}_0)$) for uncertainty calculation (Zhou et al., 2020; Kassraie et al., 2022) while some works use $g(G_t;\boldsymbol{\theta}_{t-1})$ which aligns with ours (Zhang et al., 2020). Formally, instead of the update of

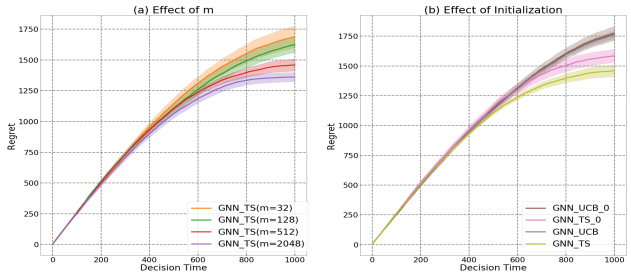

Figure 3: Increasing m can improve the performance of `GNN-TS` and no improvement of using $g(G_t; \boldsymbol{\theta}_0)$.

uncertainty estimate in (5), using initial gradient means performing the following

$$\bar{\sigma}_t^2(G) = \frac{1}{m}\|g(G;\boldsymbol{\theta}_0)\|^2_{\bar{\boldsymbol{U}}_t^{-1}}, \qquad \bar{\boldsymbol{U}}_t = \bar{\boldsymbol{U}}_{t-1} + g(G_t;\boldsymbol{\theta}_0)g(G_t;\boldsymbol{\theta}_0)^\top/m.$$

Part (a) of Figure 3 reflects that the wider MLP has better performance which matches our expectation. Moreover, part (b) of Figure 3 reflects that there are no benefits from setting gradients used in algorithms to be the initial gradients for all $t \in [T]$. One small final observation is that the effects of m and initialization are not strong.

D.5 Additional Figures and Tables

D.5.1 Results for Erdös–Rényi Random Graphs.

For better visualization of the 192 synthetic data environments using Erdös–Rényi random graphs, we summarised the result in Table 1. The metrics are relative regret and top rate, which are defined based on regret as follow. The relative regret of one algorithm in one data environment is defined as

$$\text{Relative Regret:}\tilde{R}^{\text{alg, env}} = \frac{R_T^{\text{alg, env}}}{\max_{\text{alg}} R_T^{\text{alg, env}}}$$

where $R_T^{\text{alg, env}}$ is the cumulative regret of algorithm alg, and data environment env.

We define the top rate for the policy in algorithm as the number of times such that the policy achieve the least two cumulative regret R_T. The denomnator is the number of total trails, which is the 1920, the 10 repetition and 192 combinations of ER environments. The top rate of one algorithm is defined as

$$\text{Top Rate:}\alpha_{\text{alg}} = \frac{\#\text{ times alg achieves "Top 2"}}{\#\text{ trails}}.$$

	`NN-UCB`	`NN-PE`	`NN-TS`	`GNN-UCB`	`GNN-PE`	`GNN-TS`
Top Rate (α_{alg})	0.0%	1.6%	0.0%	9.4%	90.6%	**98.4 %**
Relative Regret ($\tilde{R}^{\text{alg, env}}$)	0.994(0.02)	0.891(0.06)	0.943(0.05)	0.762(0.15)	0.690(0.14)	**0.595(0.16)**

Table 1: Results on Erdös–Rényi random graphs. 192 data environments with 10 repetitions.

D.5.2 Results for Random Dot Product Graphs

We provide the experiment results for regret on all random dot product graph settings. In thee plots, different rows represents different sizes of the graph space ($|\mathcal{G}|$) and columns represents the choices of the number of nodes in the graph (N).

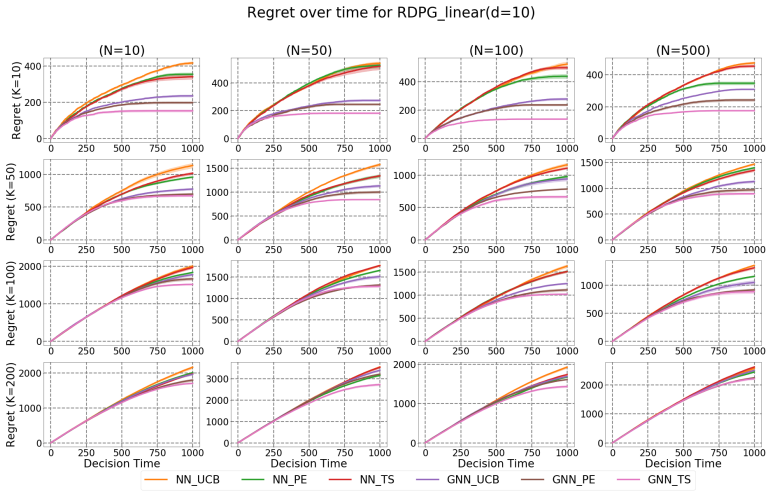

Figure 4: Random Dot Product Graphs with linear reward.

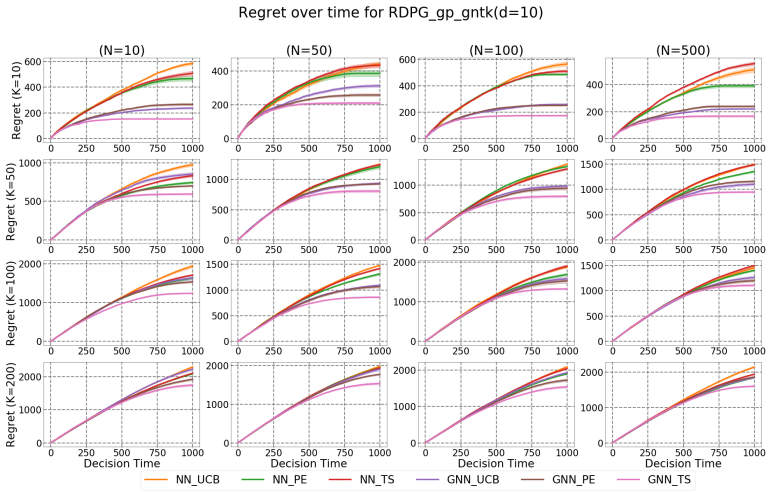

Figure 5: Random Dot Product Graphs with GP and GNTK for reward.

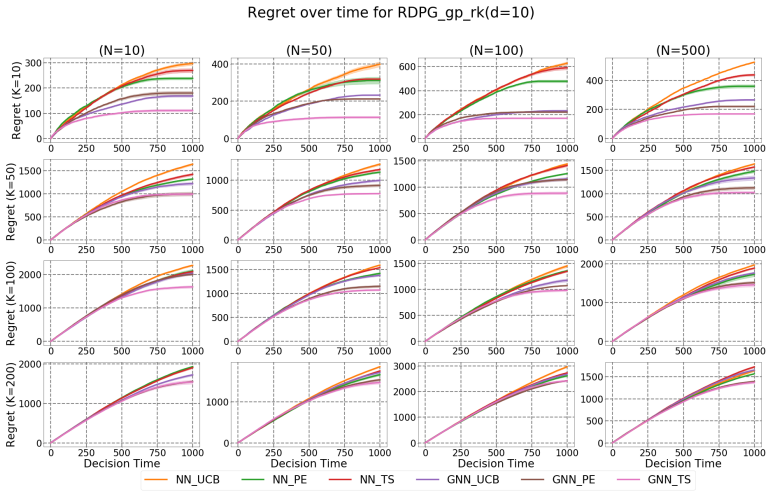

Figure 6: Random Dot Product Graphs with GP and representation kernel for reward.