WITH GREAT POWER COMES GREAT ADAPTATION: MESSAGE TUNING OUTSHINES PROMPT TUNING FOR GRAPH FOUNDATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph foundation models (GFMs), built upon the "Pre-training and Adaptation" paradigm, have emerged as a promising path toward artificial general intelligence on graphs. Despite the remarkable potential of large language models, most existing GFMs still adopt Graph Neural Networks as their backbone. For such GNN-based GFMs, prompt tuning has become the prevailing adaptation method for downstream tasks. However, while recent theoretical research has revealed why graph prompt tuning works, how to measure its adaptation capacity remains an open problem. In this paper, we propose **P**rismatic **S**pace Theory (PS-Theory) to quantify the capacity of adaptation approaches and establish the upper bound for the adaptation capacity of prompt tuning. Inspired by prefix-tuning, we introduce **M**essage **T**uning for **G**FMs (MTG), a lightweight approach that injects a small set of learnable message prototypes into each layer of the GNN backbone to adaptively guide message fusion without updating the frozen pre-trained weights. Through our PS-Theory, we rigorously prove that MTG has greater adaptation capacity than prompt tuning. Extensive experiments demonstrate that MTG consistently outperforms prompt tuning baselines across diverse benchmarks, validating our theoretical findings. Our code is available at https://anonymous.4open.science/r/MTG.

1 INTRODUCTION

Graph foundation models (GFMs) (Liu et al., 2025; Wang et al., 2025b), built upon the "Pre-training and Adaptation" paradigm, are expected to benefit from the pre-training of broad graph data and can be adapted to a wide range of downstream graph tasks. Since they are designed to natively learn graph structures, a capability fundamentally different from that of sequence-based Large Language Models, GNN-based GFMs represent a promising direction by leveraging self-supervised pre-training to acquire transferable knowledge through their GNN backbone architectures (Wang et al., 2024; Chen et al., 2025), including Message Passing Neural Networks (MPNNs) (Gilmer et al., 2017) and Graph Transformers (GTs) (Ying et al., 2021). For such pre-trained GFMs, fine-tuning (Hu et al., 2020b; Qiu et al., 2020; Rong et al., 2020) is the most intuitive and widely adopted method for downstream task adaptation. However, fine-tuning generally involves updating all model parameters, requiring a full model copy per task while demanding substantial computational resources and task-specific data for full customization. Furthermore, the pretext-downstream graph task gap poses a significant challenge for fine-tuning, potentially causing negative transfer (Wang et al., 2021), particularly in few-shot scenarios (Zhang et al., 2022).

Prompt tuning (Lester et al., 2021; Liu et al., 2022) as a popular finetune paradigm, has emerged as an efficient alternative to full-parameter fine-tuning by freezing the pre-trained model's parameters and adapting downstream tasks through input data transformations. Prompt tuning on graphs, known as Graph Prompt Learning (Sun et al., 2023b), enhances GNN-based models' performance and adaptability through input-space adaptations (e.g., inserting lightweight learnable tokens or subgraphs) to reformulate downstream tasks as pre-training tasks without modifying the pre-trained GNN backbone. Recent advances in graph prompt models (Fang et al., 2023; Sun et al., 2023a; Niu et al., 2024; Yu et al., 2025) have shown promising results in graph learning, highlighting their potential for broader graph intelligence applications spanning from molecular property prediction (Diao et al., 2023) to recommendation systems (Yang et al., 2023). Concurrently, several studies

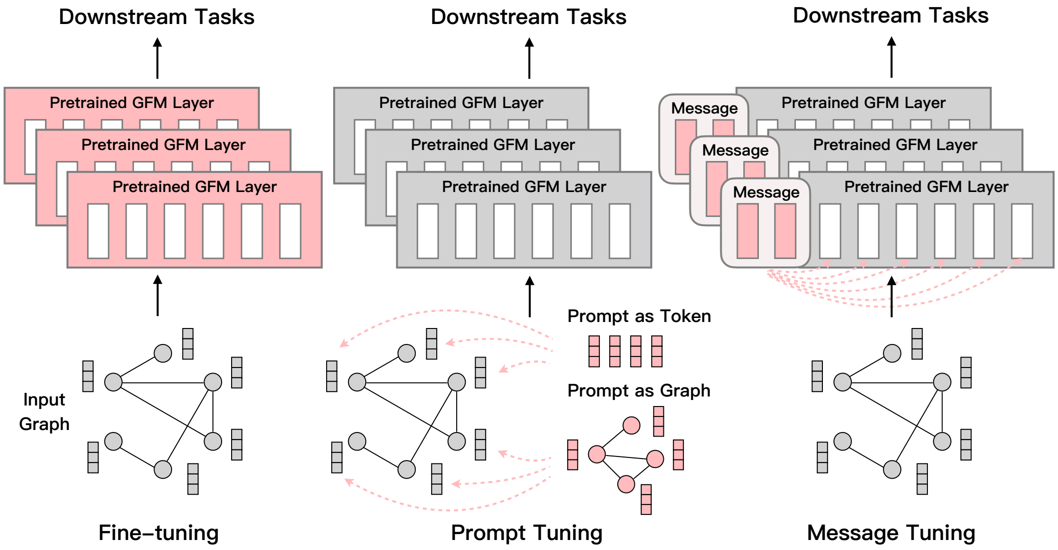

Figure 1: Fine-tuning (left) updates all GFM parameters (red GFM Layer boxes), while Prompt Tuning (middle) typically updates prompt tokens or the prompt graph (red prompt vectors) to transform the input graph, keeping GFM parameters frozen. We propose Message Tuning (right), which also freezes GFM parameters but optimizes the messages in each GFM Layer (red message blocks) to regulate message fusion. The red dashed lines indicate the inserting patterns of new parameters, which are also learnable.

(Fang et al., 2023; Wang et al., 2025a) have begun analyzing graph prompts from a data operation perspective, suggesting their effectiveness stems from simulating fundamental graph transformations such as node/edge addition/deletion, feature modification, and subgraph removal.

However, while recent theoretical research has revealed why graph prompt tuning works from a data operation perspective (Wang et al., 2025a), how to measure its adaptation capacity on a specific GFM remains an open problem. A more precise understanding of the capability bound of prompt tuning and the underlying reasons will facilitate the design of more powerful and efficient adaptation methods. To address this issue, we model each layer of GFMs as a piecewise linear refractive transformation and leverage ideas from geometric measure theory to quantify the “refractive” power of each layer, establishing fundamental bounds on the expressive power of prompt tuning methods and motivating the design of our novel message tuning paradigm.

Specifically, we propose **P**rismatic **S**pace Theory (PS-Theory), providing a rigorous mathematical framework to quantify adaptation capacity and establish the upper bound for the adaptation capacity of prompt tuning. Prefix-tuning (Li & Liang, 2021), widely used in language models, is specifically designed for transformer architectures and generative tasks on sequential data, making it not directly applicable to graph-structured data. Inspired by this technique, we introduce **M**essage **T**uning for **G**FMs (MTG), a novel adaptation approach that injects learnable message prototypes into each layer and dynamically fuses them with the model’s native messages, which is compatible with GFMs using either MPNN or GT backbones, as illustrated in Figure 1. Through our PS-Theory, we rigorously prove that MTG has greater adaptation capacity than prompt tuning. Extensive and fair evaluations on the latest Graph Prompt Learning benchmark ProG (Zi et al., 2024) further validate MTG’s superiority across diverse downstream tasks. The contributions of this paper are summarized as follows:

- **Theoretical Foundation.** Different from the prior theories focusing on data operations, we propose PS-Theory, providing a novel and rigorous mathematical framework to quantify adaptation capacity and establish the upper bound for the adaptation capacity of prompt tuning.
- **Adaptation Method.** We introduce MTG, a novel lightweight adaptation approach that dynamically guides message fusion by injecting learnable message prototypes across all layers without updating pre-trained weights, significantly enhancing adaptation capacity. Through our PS-Theory, we rigorously prove that MTG has greater adaptation capacity than prompt tuning.
- **Extensive Experiments.** Through comprehensive evaluations across diverse few-shot downstream tasks, we demonstrate MTG’s consistent superiority over state-of-the-art prompt tuning baselines, validating our theoretical claims on its enhanced adaptation capacity.

2 Related Work

The adaptation of GNN-based GFMs involves tailoring models or adjusting input data to align with specific downstream tasks or domains through techniques such as fine-tuning and prompt tuning. To the best of our knowledge, prefix-tuning (Li & Liang, 2021), despite its prevalence in language models, remains unexplored in GNN-based GFMs.

Fine-tuning. Specifically, fine-tuning can be further divided into full-parameter fine-tuning (FPFT) and parameter-efficient fine-tuning (PEFT). FPFT (Hu et al., 2020b; Qiu et al., 2020; Rong et al., 2020; Sun et al., 2024) entails training the entire pre-trained model on task-specific data, offering high customization at the cost of substantial computational resources. In contrast, PEFT methods optimize only a subset of parameters, balancing adaptation efficiency with performance. For instance, AdapterGNN (Li et al., 2024) modifies input graphs via parallel adapters around message passing, G-Adapter (Gui et al., 2024) integrates graph structure into transformer fine-tuning through graph message passing, and GraphLoRA (Yang et al., 2025) enhances efficiency by injecting a low-rank trainable GNN alongside the pre-trained model to address structural distribution gaps while mitigating catastrophic forgetting. In this paper, fine-tuning generally refers to FPFT unless otherwise specified.

Prompt Tuning. As a lightweight tuning method, prompt tuning typically freezes pre-trained model parameters while introducing additional learnable components in the input space. Following Liu et al. (2025), prompt tuning methods can be divided into two distinct groups: pre-prompt and post-prompt methods, depending on whether task-specific prompts operate before or after the backbone module. Pre-prompt methods either modify graph topology or node features before message passing to enhance task performance, or construct prompt graphs to boost adaptability. For instance, GPF (Fang et al., 2023) introduces an optimizable uniform feature vector for all nodes to adapt pre-trained GNNs across strategies, while All-in-one (Sun et al., 2023a) reformulates node-level and edge-level tasks to graph-level tasks and treats an additional subgraph as a prompt that merges with the node subgraph. Post-prompt methods apply task-specific prompts on representations after message passing for downstream adaptation. For instance, GPPT (Sun et al., 2022) transforms node classification into link prediction via class-specific token pairs, while GraphPrompt (Liu et al., 2023b) unifies tasks through subgraph similarity and learns task-specific prompt vectors to adapt the Readout operation, bridging link prediction and downstream tasks.

3 Prismatic Space Theory

In this section, we introduce Prismatic Space Theory (PS-Theory), providing a novel perspective and rigorous mathematical framework to quantify the capacity of adaptation approaches and establish the upper bound for the adaptation capacity of prompt tuning. Due to space constraints, the proofs of all theorems and additional theoretical details are provided in Appendix B.

3.1 A Unified Formulation for GNN-based GFMs

To facilitate theoretical analysis, we present a unified formal framework that generalizes both MPNNs and GTs architectures. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with $N = |\mathcal{V}|$ nodes. The node feature matrix is denoted $\boldsymbol{X} \in \mathbb{R}^{N \times d_0}$ and the adjacency matrix is denoted $\boldsymbol{A} \in \{0,1\}^{N \times N}$. The ℓ-th layer of a general GNN-based GFM is defined by the following formulation.

Definition 1 (Unified GFM Layer). *For any layer $\ell \in \{1, \dots, L\}$, the node representation matrix $\boldsymbol{H}^{(\ell)} \in \mathbb{R}^{N \times d_\ell}$ is computed as:*

$$\boldsymbol{H}^{(\ell)} = \mathfrak{U}^{(\ell)}\left(\mathfrak{M}^{(\ell)}\left(\mathfrak{A}^{(\ell)}\left(\boldsymbol{A}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_a^{(\ell)}\right), \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_m^{(\ell)}\right), \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_u^{(\ell)}\right), \tag{1}$$

where $\boldsymbol{H}^{(0)} = \boldsymbol{X}$. $\mathfrak{A}^{(\ell)}$, $\mathfrak{M}^{(\ell)}$, and $\mathfrak{U}^{(\ell)}$ denote the attention, message fusion, and update operators respectively: $\mathfrak{A}^{(\ell)}$ computes attention weights (encompassing both learnable dynamic attention and static structural attention), $\mathfrak{M}^{(\ell)}$ performs weighted aggregation of node messages using these attention scores, and $\mathfrak{U}^{(\ell)}$ combines previous node representations with the fused messages to produce the updated representation.

This formulation serves as a unified formalization of the core structure rather than encompassing all architecture details. The detailed correspondence between this formulation and classical backbone architectures is presented in Appendix B.2.

3.2 Prompt Tuning for Graphs

Following Wang et al. (2025a), we provide a mathematical formalization of prompt tuning for graphs, aiming to offer an intuitive perspective for theoretical analysis. Let f_{GFM} denote a pre-trained GNN-based GFM with frozen parameters, and let g_θ denote a graph prompt function with parameters θ that transforms the input graph $\mathcal{G}$ into a prompted graph $g_\theta(\mathcal{G})$. Given a downstream dataset $\mathbb{G} = \{\mathcal{G}\}$, the goal of prompt tuning is to optimize θ to maximize the likelihood of the optimal representation for a graph $\mathcal{G}$ from $\mathbb{G}$. This objective can be formulated as:

$$\max_\theta P_{f_{\text{GFM}}}(v_{\mathcal{G}} | g_\theta(\mathcal{G})) \tag{2}$$

The theory in Wang et al. (2025a) rests on the assumption that a GNN model acts as a surjective mapping operator from the graph set $\mathbb{G}$ to $\mathbb{R}^F$, where F is the dimensionality of the representation. However, since real-world graph data is inherently bounded, the model's output is unlikely to span the entire $\mathbb{R}^F$ space. Analyzing the properties of this actual output space will provide deeper insight into the adaptation capacity of prompt tuning.

3.3 A Geometric Measure Theoretic Formulation

Prism Metaphor. Prompt tuning typically operates by injecting a low-dimensional prompt into the high-dimensional input space of a frozen GFM. To quantify its efficacy, we need to understand how the GFM's architecture transforms this input space. We posit that each layer of a GFM, particularly those employing piecewise linear activations like ReLU (Nair & Hinton, 2010) or LeakyReLU (Maas et al., 2013), acts not merely as a contraction but as a "prism". The non-isometric, piecewise linear action of a "prism" refracts the input space, collapsing some dimensions into oblivion, and progressively folding the input manifold. We model the GFMs as a sequence of measurable maps that transform the input space into a sequence of increasingly complex, lower-dimensional prismatic space. We quantify the "refractive" power of each layer by leveraging ideas from geometric measure theory, focusing on the singular values of the layer's Jacobian and their effect on the intrinsic dimension and measure of the data manifold.

We adopt the unified GFM layer from Definition 1 and first introduce the piecewise linear map, which is a key abstraction for understanding the mechanisms of the model architecture.

Definition 2 (Piecewise Linear Function for Matrix Maps)**.** *A function $F : \mathbb{R}^{N\times d_{in}} \to \mathbb{R}^{N\times d_{out}}$ is called piecewise linear if there exists a finite collection of polyhedral regions $\{R_i\}_{i=1}^K$ in $\mathbb{R}^{N\times d_{in}}$ such that $\mathbb{R}^{N\times d_{in}} = \bigcup_{i=1}^K R_i$ and for each region R_i, the function F is affine, meaning there exists a matrix $\boldsymbol{A}_i \in \mathbb{R}^{(Nd_{out})\times(Nd_{in})}$ and a vector $\boldsymbol{b}_i \in \mathbb{R}^{Nd_{out}}$ such that:*

$$vec(F(\boldsymbol{H})) = \boldsymbol{A}_i \cdot vec(\boldsymbol{H}) + \boldsymbol{b}_i \quad \textit{for all } \boldsymbol{H} \in R_i. \tag{3}$$

Equivalently, in matrix form, $F(\boldsymbol{H}) = unvec(\boldsymbol{A}_i \cdot vec(\boldsymbol{H}) + \boldsymbol{b}_i)$, where unvec is the operation that reshapes the vector into an $N \times d_{out}$ matrix. (See Appendix B.3 for details.)

Proposition 1. *The attention, message fusion, and update operators $\mathfrak{A}^{(\ell)}, \mathfrak{M}^{(\ell)}, \mathfrak{U}^{(\ell)}$ are generally continuous, piecewise linear functions and differentiable almost everywhere (a.e.).*

Proposition 2. *The layer map $F^{(\ell)} : \mathbb{H}^{(\ell-1)}(\subset \mathbb{R}^{N\times d_{\ell-1}}) \to \mathbb{H}^{(\ell)}(\subset \mathbb{R}^{N\times d_\ell})$ is a piecewise linear function. For any point $\boldsymbol{H}$ where $F^{(\ell)}$ is differentiable, its Jacobian $\boldsymbol{J}^{(\ell)}(\boldsymbol{H}) \in \mathbb{R}^{Nd_\ell \times Nd_{\ell-1}}$ exists.*

The proofs of Propositions 1 and 2 are provided in Appendices B.4 and B.5, respectively. Having abstracted the model architecture, we now turn to mathematically modeling the input data and output space, with particular emphasis on capturing the bounded nature of the input data.

Definition 3 (Input Manifold and Representation Space)**.** *The input space is modeled as a compact, smooth input manifold $\mathcal{M}_0 \subset \mathcal{X} \subset \mathbb{R}^{N\times d_0}$, with intrinsic dimension $d_{int}(\mathcal{M}_0) = D_0$. $\mathcal{X}$ denotes the entire set of possible input data forms for the model. $\mathcal{M}_0$ represents a low-dimensional subset of $\mathcal{X}$ endowed with specific semantic and geometric structures (see Appendix B.6 for details). The representation at layer ℓ is the image of the input manifold under the composite map $\Phi^{(\ell)} = F^{(\ell)} \circ \cdots \circ F^{(1)}$:*

$$\mathcal{M}^{(\ell)} = \Phi^{(\ell)}(\mathcal{M}_0) \subset \mathbb{R}^{N\times d_\ell}. \tag{4}$$

Definition 4 (Prismatic Space)**.** *A set $\mathcal{M} \subset \mathbb{R}^n$ is called prismatic space if there exists a smooth manifold $\mathcal{N} \subset \mathbb{R}^m$ and a piecewise linear map $f : \mathbb{R}^m \to \mathbb{R}^n$ such that $\mathcal{M} = f(\mathcal{N})$.*

Real-world graph data can be viewed as points sampled from the input manifold. And the output representations of graph data lie in the prismatic space, defined as the image of the input manifold under the GFM's mapping. Hence, such a geometric modeling approach is of practical significance.

Proposition 3. *$\Phi^{(\ell)} = F^{(\ell)} \circ \cdots \circ F^{(1)}$ is piecewise linear. Assume that $\Phi^{(\ell)}$ is injective on each polyhedral region, then $\mathcal{M}^{(\ell)} = \Phi^{(\ell)}(\mathcal{M}_0)$ is a prismatic space and may have singularities.*

The proof of Proposition 3 is provided in Appendix B.7. As in many geometric theories, an intuitive strategy for analyzing complex geometric space is to begin with a local perspective, particularly since the formation process of prismatic space is already well understood. The core of the prismatic effect lies in the singular value decomposition (SVD) of the layer Jacobians.

Definition 5 (Spectral Prism of a Layer)**.** *For a point $\boldsymbol{H} \in \mathbb{H}^{(\ell-1)}$ where $F^{(\ell)}$ is differentiable, let $\boldsymbol{J}^{(\ell)}(\boldsymbol{H}) = \boldsymbol{U}^{(\ell)}\boldsymbol{\Sigma}^{(\ell)}(\boldsymbol{V}^{(\ell)})^\top$ be its SVD (see Appendix B.8 for details). The diagonal matrix $\boldsymbol{\Sigma}^{(\ell)} = diag(\sigma_1^{(\ell)}, \sigma_2^{(\ell)}, ..., \sigma_{r_\ell}^{(\ell)}, 0, ..., 0)$ contains the singular values, where r_ℓ is the rank.*

Theorem 1 (Local Measure Contraction Factor)**.** *Let $\mathbb{S} \subset \mathbb{H}^{(\ell-1)}$ be a sufficiently small measurable set contained in an s-dimensional subspace $\mathbb{V}$ on which $F^{(\ell)}$ is linear and injective, with constant Jacobian $\boldsymbol{J}^{(\ell)}$ of rank r_ℓ ($s \leq r_\ell$). Assume $\mathbb{V}$ is the subspace spanned by the first s right singular vectors of $\boldsymbol{J}^{(\ell)}$, corresponding to the s largest singular values $\sigma_1^{(\ell)} \geq \sigma_2^{(\ell)} \geq \cdots \geq \sigma_s^{(\ell)} > 0$. Then, for the s-dimensional Hausdorff measure $\mathcal{H}^s$:*

$$\mathcal{H}^s(F^{(\ell)}(\mathbb{S})) = \Big(\prod_{i=1}^{s} \sigma_i^{(\ell)}\Big)\mathcal{H}^s(\mathbb{S}). \tag{5}$$

In particular, if $s = r_\ell$, the volume contraction factor is $\prod_{i=1}^{r_\ell} \sigma_i^{(\ell)}$.

Corollary 1 (Local ReLU Prism Effect)**.** *Consider the ReLU activation function used within $F^{(\ell)}$. At points where ReLU is differentiable, its Jacobian $\boldsymbol{J}_{ReLU}$ is a diagonal matrix with diagonal entries either 0 or 1, and hence idempotent ($\boldsymbol{J}_{ReLU}^2 = \boldsymbol{J}_{ReLU}$). This implies that ReLU acts as a local projection, nullifying some dimensions (setting outputs to zero) and preserving others. The ReLU component contributes to the prismatic effect by introducing sparsity and reducing the effective rank of the Jacobian in local regions.*

The proof of Theorem 1 is provided in Appendix B.9. Corollary 1 provides a detailed explanation of the ReLU activation function. Having characterized the local properties via the singular values of the layer Jacobians, we now need to synthesize the global and local perspectives. This requires an abstract mathematical technique: constructing a global partition from local pieces is a common approach, even foundational to calculus. By Proposition 3, the piecewise linearity of the GFM network implies that the input manifold $\mathcal{M}_0$ is partitioned into multiple linear regions.

Definition 6 (Linear Region Partition)**.** *For each layer $\ell \in 1, \ldots, L$, let $\Omega^{(\ell)}$ be the set of polytopic regions in $\mathbb{H}^{(\ell-1)}$ on which the function $F^{(\ell)}$ is linear. The GFM network $\Phi = F^{(L)} \circ \cdots \circ F^{(1)}$ defines a recursive partition of the input manifold $\mathcal{M}_0$ into cells $\{C_k\}$, where each cell C_k is a connected subset of $\mathcal{M}_0$ such that there exists a sequence of regions $R_1 \in \Omega^{(1)}, R_2 \in \Omega^{(2)}, \ldots, R_L \in \Omega^{(L)}$ satisfying:*

$$C_k \subseteq R_1, F^{(1)}(C_k) \subseteq R_2, F^{(2)}(F^{(1)}(C_k)) \subseteq R_3, \ldots, F^{(L-1)} \circ \cdots \circ F^{(1)}(C_k) \subseteq R_L, \tag{6}$$

and on each cell C_k, the full network map Φ is linear. The total number of $\{C_k\}$ is related to the specific architecture and parameters of the GFM network.

Theorem 2 (Prismatic Folding and Intrinsic Dimension)**.** *The global map $\Phi : \mathcal{M}_0 \to \mathcal{M}^{(L)}$ is piecewise linear. The intrinsic dimension of the final representation space is bounded by the maximum over linear regions of the minimum rank achieved across layers:*

$$d_{int}(\mathcal{M}^{(L)}) \leq \max_k \min_\ell rank(\boldsymbol{J}^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}). \tag{7}$$

Furthermore, the map Φ is piecewise constant on its rank. The final output $\mathcal{M}^{(L)}$ is a prismatic space embedded in $\mathbb{R}^{N \times d_L}$, likely with a much lower intrinsic dimension than D_0.

The proof of Theorem 2 in Appendix B.11 provides a method for analyzing the upper bound of the intrinsic dimension of the output prismatic space. This bound is analytical, derived from the partition of the input manifold induced by the GFM network as defined in Definition 6, making it difficult to compute numerically. Building on the local measure computation derived in Theorem 1, we formulate the definition of a global measure on the prismatic space in the following theorem.

Theorem 3 (Measure of the Final Prismatic Space)**.** *Assume the piecewise linear map $\Phi = F^{(L)} \circ \cdots \circ F^{(1)}$ is injective on the partition C_k of the input manifold $\mathcal{M}_0$, where each C_k is a cell in the linear region partition. Then, the d_{int}-dimensional Hausdorff measure of the final prismatic space $\mathcal{M}^{(L)} = \Phi(\mathcal{M}_0)$ is given by:*

$$\mathcal{H}^{d_{int}}(\mathcal{M}^{(L)}) = \sum_k \mathcal{H}^{d_{int}}(\Phi(C_k)) = \sum_k \left(\prod_{\ell=1}^{L} \prod_{i=1}^{d_{int}} \sigma_{i,k}^{(\ell)} \right) \mathcal{H}^{d_{int}}(C_k), \tag{8}$$

where for each layer ℓ and cell C_k, $\sigma_{i,k}^{(\ell)}$ for $i = 1, \ldots, d_{int}$ are the d_{int} largest singular values of the Jacobian $\boldsymbol{J}^{(\ell)}$ of $F^{(\ell)}$ restricted to the tangent space of $\Phi^{(\ell-1)}(C_k)$ (which is d_{int}-dimensional). If Φ is not injective, the formula provides an upper bound.

The proof of Theorem 3 is provided in Appendix B.12. This theorem precisely quantifies the prismatic effect: the total "volume" of the final representation is the sum of the volumes of all fragments of the input manifold, each shrunk by the product of the singular values of the Jacobians along its path through the network. At this point, we have established a mathematical framework (PS-Theory) for analyzing the output prismatic space of GFM. However, corresponding theoretical results on adaptation capacity still require integration with specific adaptation methods, such as prompt tuning.

3.4 Adaptation Capacity of Prompt Tuning

Without loss of generality, and in alignment with Wang et al. (2025a), our theoretical analysis in this subsection focuses on pre-prompt methods. As demonstrated by Lemma 1 in Wang et al. (2025a), prompt tuning methods, such as GPF and All-in-One, are equivalent to a transformation of the node feature matrix $\boldsymbol{X}$. This transformation can be simplified to the form $\boldsymbol{X}_\omega = \tilde{\boldsymbol{X}} + \boldsymbol{c}\boldsymbol{p}^\top$, where $\boldsymbol{c} \geq \boldsymbol{0}$ can be referred to as the coefficient vector and $\tilde{\boldsymbol{X}}$ can be either $\boldsymbol{X}$ or the natural extension of $\boldsymbol{X}$: $[\begin{smallmatrix}\boldsymbol{X} \\ \boldsymbol{0}\end{smallmatrix}]$. Therefore, modeling prompt tuning as a perturbation on the input manifold is a natural and direct approach within the PS-Theory framework.

Definition 7 (Prompt Perturbation Manifold)**.** *Assume the original input manifold $\mathcal{M}_0$ is perturbed by a prompt $\boldsymbol{P}$, forming a compact smooth manifold $\mathcal{M}_0(\boldsymbol{P})$, e.g., $\mathcal{M}_0(\boldsymbol{P}) = \{\boldsymbol{X} + \boldsymbol{P} \mid \boldsymbol{X} \in \mathcal{M}_0\}$. The prompt space $\mathcal{P}$ defines a manifold family: $\{\mathcal{M}_0(\boldsymbol{P}) \mid \boldsymbol{P} \in \mathcal{P}\}$. (See Appendix B.13 for details.)*

Theorem 4 (The Prompt Efficacy Bound)**.** *The adaptation capacity of a prompt $\boldsymbol{P}$ to influence model output is bounded by the measure and diameter of $\mathcal{M}^{(L)}(\boldsymbol{P})$:*

$$\textit{(Measure Bound)} \quad \mathcal{H}^{d_{int}}(\mathcal{M}^{(L)}(\boldsymbol{P})) \leq \left(\sup_k \prod_{\ell=1}^{L} \prod_{i=1}^{d_{int}} \sigma_{i,k}^{(\ell)} \right) \cdot \mathcal{H}^{d_{int}}(\mathcal{M}_0(\boldsymbol{P})), \tag{9}$$

$$\textit{(Diameter Bound)} \quad \mathit{diam}(\mathcal{M}^{(L)}(\boldsymbol{P})) \leq \left(\prod_{\ell=1}^{L} \sup_k |\boldsymbol{J}_k^{(\ell)}|_{op} \right) \cdot \mathit{diam}(\mathcal{M}_0(\boldsymbol{P})), \tag{10}$$

where $|\cdot|_{op}$ is the spectral norm (the largest singular value), $\sigma_{i,k}^{(\ell)}$ are the singular values of the Jacobian of the ℓ-th layer in the k-th linear region, and d_{int} is the intrinsic dimension of $\mathcal{M}^{(L)}(\boldsymbol{P})$.

The proof of Theorem 4 is provided in Appendix B.15. This theorem reveals that prompt tuning is fundamentally constrained by the frozen network's architecture. The prompt's influence is compressed by the product of layer-wise Jacobian singular values, leading to irreversible information loss. Since the prismatic, piecewise linear structure of the pre-trained model is immutable, the prompt can only shift the input within this fixed, contracting geometric framework (details are in Appendix B.16).

The establishment of Prismatic Space Theory revolves around prompt tuning, yet it offers a more fundamental geometric perspective on how graph foundation models process input manifolds. The theory is constructed layer by layer, making it not limited to adaptation methods that operate solely at the input data level, but also applicable to the analysis of other types of adaptation approaches.

4 Message Tuning for GFMs

In this section, we introduce Message Tuning for GFMs (MTG), a novel lightweight adaptation approach that dynamically guides message fusion across all layers (Figure 1). Through our PS-Theory in Section 3, we rigorously prove that MTG has greater adaptation capacity than prompt tuning.

4.1 Core Mechanism

The concept of MTG shares a similar inspiration with prefix-tuning (Li & Liang, 2021), which is widely adopted in large language models. However, prefix-tuning is specifically designed for transformer architectures and generative tasks on sequential data, making it not directly applicable to graph-structured data. In this work, we introduce a general message tuning framework tailored for graph foundation models with diverse backbone architectures. The core mechanism of MTG is to inject a small set of learnable message prototypes into each layer, which then undergo a dynamic fusion with the native messages computed by the model, while the original parameters $\mathbf{\Theta}^{(\ell)} = \{\mathbf{\Theta}_a^{(\ell)}, \mathbf{\Theta}_m^{(\ell)}, \mathbf{\Theta}_u^{(\ell)}\}$ in Eq.(1) are kept frozen.

Learnable Message Prototypes. Formally, for each layer ℓ, we introduce a small set of m learnable prototype vectors, denoted as $\boldsymbol{M}^{(\ell)} = [\boldsymbol{m}_1^{(\ell)}, \boldsymbol{m}_2^{(\ell)}, \dots, \boldsymbol{m}_m^{(\ell)}]^\top \in \mathbb{R}^{m \times d_{\ell-1}}$. Then the GFM layer after injecting message prototypes can be expressed as:

$$\boldsymbol{H}^{(\ell)} = \mathfrak{U}^{(\ell)} \left(\mathfrak{M}^{(\ell)} \left(\mathfrak{A}^{(\ell)} \left(\boldsymbol{A}, \boldsymbol{H}_M^{(\ell-1)}; \mathbf{\Theta}_a^{(\ell)} \right), \boldsymbol{H}_M^{(\ell-1)}; \mathbf{\Theta}_m^{(\ell)} \right), \boldsymbol{H}_M^{(\ell-1)}; \mathbf{\Theta}_u^{(\ell)} \right), \tag{11}$$

$$\boldsymbol{H}_M^{(\ell-1)} = \mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{M}^{(\ell)}; \mathbf{\Theta}_f^{(\ell)}), \tag{12}$$

where $\mathfrak{F}^{(\ell)}$ denotes dynamic message fusion operator, $\boldsymbol{M}^{(\ell)}$ and $\mathbf{\Theta}_f^{(\ell)}$ are the learnable parameters. This is equivalent to replacing $\boldsymbol{H}^{(\ell-1)}$ in Eq.(1) with $\boldsymbol{H}_M^{(\ell-1)}$ defined in Eq.(12), resulting in Eq.(11).

Dynamic Message Fusion. While both are message fusion operators, $\mathfrak{F}^{(\ell)}$ differs from $\mathfrak{M}^{(\ell)}$ in that it dynamically fuses learnable message prototypes with the input message representations at each layer, instead of fusing messages between nodes. We simply employ a linear projection followed by a row-wise Softmax operation to compute the attention for fusing $\boldsymbol{H}^{(\ell-1)}$ with $\boldsymbol{M}^{(\ell)}$. Thus, $\mathfrak{F}^{(\ell)}$ can be expressed as:

$$\mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{M}^{(\ell)}; \mathbf{\Theta}_f^{(\ell)}) = \boldsymbol{H}^{(\ell-1)} + \text{Softmax}(\boldsymbol{H}^{(\ell-1)} \boldsymbol{W}_p^{(\ell)}) \cdot \boldsymbol{M}^{(\ell)} \tag{13}$$

where $\mathbf{\Theta}_f^{(\ell)} = \boldsymbol{W}_p^{(\ell)} \in \mathbb{R}^{d_{\ell-1} \times m}$ is the projection matrix. Alternatively, one may consider replacing linear projections with MLPs or employing dot-product attention, though this may introduce higher computational complexity.

4.2 Theoretical Analysis

Consider a pre-trained GFM Φ with L layers as defined in Definition 1, and let $\mathcal{M}_0 \subset \mathbb{R}^{N \times d_0}$ be the compact smooth input manifold with intrinsic dimension D_0. Let $\mathcal{P}$ be the set of possible prompts for prompt tuning, and for any prompt $\boldsymbol{P} \in \mathcal{P}$, let $\mathcal{M}_0(\boldsymbol{P})$ be the perturbed input manifold. The final representation space under prompt tuning is $\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P}) = \Phi(\mathcal{M}_0(\boldsymbol{P}))$.

Theorem 5 (Message Tuning Has Greater Adaptation Capacity). *For message tuning, we inject learnable message prototypes $\boldsymbol{M}^{(\ell)} \in \mathbb{R}^{m \times d_{\ell-1}}$ and fusion parameters $\mathbf{\Theta}_f^{(\ell)}$ at each layer ℓ, resulting in a modified network Φ_{MTG}. Let $\mathcal{M}_{MTG}^{(L)}$ be the final representation space under message tuning with optimally chosen parameters. Then, the following inequalities hold:*

$$\textit{(Intrinsic Dimension Comparison)} \quad d_{int}(\mathcal{M}_{MTG}^{(L)}) \geq d_{int}(\mathcal{M}_{PT}^{(L)}(\boldsymbol{P})) \quad \textit{for all } \boldsymbol{P} \in \mathcal{P}, \tag{14}$$

$$\textit{(Measure Comparison)} \quad \mathcal{H}^{d_{int}}(\mathcal{M}_{MTG}^{(L)}) \geq \mathcal{H}^{d_{int}}(\mathcal{M}_{PT}^{(L)}(\boldsymbol{P})) \quad \textit{for all } \boldsymbol{P} \in \mathcal{P}, \tag{15}$$

$$\textit{(Diameter Comparison)} \quad diam(\mathcal{M}_{MTG}^{(L)}) \geq diam(\mathcal{M}_{PT}^{(L)}(\boldsymbol{P})) \quad \textit{for all } \boldsymbol{P} \in \mathcal{P}. \tag{16}$$

Moreover, there exists a message tuning configuration such that the inequalities are strict.

In the semantic context of the geometric properties of the prismatic space output by the GFM, this theorem reveals that MTG has greater adaptation capacity than prompt tuning. The proof of Theorem 5 and further theoretical analysis are provided in Appendix C.

5 EXPERIMENTS

In this section, we conduct extensive experiments to evaluate our proposed MTG on the Graph Prompt Learning benchmark ProG (Zi et al., 2024) by answering the following five research questions:

Q1: How does MTG perform compared to prompt tuning baselines? (Section 5.2)
Q2: How do different pre-training strategies affect MTG's adaptation capability? (Section 5.3)
Q3: Can MTG effectively mitigate negative transfer during adaptation? (Section 5.4)
Q4: How does MTG perform on different backbone models? (Appendix F.2)
Q5: What is MTG's computational efficiency compared to prompt tuning methods? (Appendix F.3)

5.1 EXPERIMENT SETTING

Datasets. To investigate the adaptability of MTG across diverse graphs, we conduct experiments across 15 datasets from the Graph Prompt Learning benchmark ProG (Zi et al., 2024). We evaluate our method over 7 node classification benchmarks spanning homophilic graphs (Cora, Citeseer, PubMed) (Sen et al., 2008), heterophilic graphs (Texas, Actor, Wisconsin) (Pei et al., 2020), and large-scale graphs (ogbn-arxiv) (Hu et al., 2020a). For graph-level tasks, we employ 8 graph classification datasets across diverse domains, including biological datasets (D&D, ENZYMES, PROTEINS) (Dobson & Doig, 2003; Borgwardt et al., 2005; Wang et al., 2022), small molecule datasets (BZR, COX2, MUTAG) (Kriege & Mutzel, 2012; Rossi & Ahmed, 2015), and social networks (COLLAB, IMDB-B) (Yanardag & Vishwanathan, 2015). Table 6 summarizes the statistics of all datasets and more dataset details are provided in Appendix D.2.

Backbones. Since the latest studies (Luo et al., 2024; 2025) have once again validated the powerful capabilities of GCN (Kipf & Welling, 2017) as the most classic and widely used graph neural network, we choose GCN as the baseline to compare MTG with prompt tuning. We also investigate other models commonly used as backbones for GFMs, such as GraphSAGE (Hamilton et al., 2017), GAT (Veličković et al., 2018), GIN (Xu et al., 2019), and Graph Transformer (Ying et al., 2021), and the results can be found in Appendix F.2.

Pre-training Strategies. Following ProG (Zi et al., 2024), we adopt six representative pre-training strategies across three levels: DGI (Veličković et al., 2019) maximizes node-graph mutual information while GraphMAE (Hou et al., 2022) reconstructs masked features at the node level; EdgePreGPPT (Sun et al., 2022) computes link probabilities and EdgePreGprompt (Liu et al., 2023b) learns triplet-based similarities for edge-level tasks; GraphCL (You et al., 2020) enforces augmentation consistency and SimGRACE (Xia et al., 2022) performs parameter perturbations at the graph level.

Prompt Tuning Baselines. We first adopt supervised learning as the baseline for evaluating positive transfer, where negative transfer is identified when adaptation methods fail to surpass supervised performance. Following ProG (Zi et al., 2024), we compare MTG to fine-tuning and the following prevalent prompt tuning methods: GPPT (Sun et al., 2022), Gprompt (Liu et al., 2023b), All-in-one (Sun et al., 2023a), GPF and GPF-plus (Fang et al., 2023). Baseline results combine those from ProG with our own reproductions, and details are provided in Appendix E.

Implementation. For node tasks, we use 90% of the data to the test set, while for graph tasks, we use an 80% test split. To ensure robustness, we repeat sampling five times to construct k-shot tasks and report average and standard deviation over these five results. Evaluation metrics include Accuracy (primary metric in Section 5), Macro F1 Score, and AUROC. Hyperparameters are optimized via random search. A comprehensive description of the experimental setup is provided in Appendix D.

5.2 UPPER BOUND PERFORMANCE OF MESSAGE TUNING

One-shot node/graph classification is one of the most challenging downstream adaptation tasks for graph foundation models, as it requires learning the characteristics of an entire class using only one sample. In Table 1 and Table 2, we present the best results achieved by various adaptation methods across 15 datasets, which represent the top adaptation performance from pre-trained models with different strategies. This offers an intuitive reflection of the upper bound performance of each type of adaptation method. The results in the tables demonstrate that our adaptation method MTG achieves a higher performance upper bound across all 15 datasets compared to state-of-the-art prompt tuning methods, which aligns with our theoretical insights. Despite being trained on only a small number

Table 1: Performance comparison of adaptation methods on 1-shot node classification (accuracy±std %, 5 runs). The first, second and third best results are shaded in red, with descending color saturation.

Method	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	$26.56_{\pm5.55}$	$21.78_{\pm7.32}$	$39.37_{\pm16.34}$	$41.60_{\pm3.10}$	$37.97_{\pm5.80}$	$20.57_{\pm4.47}$	$10.99_{\pm3.19}$
Fine-tuning	$52.61_{\pm1.73}$	$35.05_{\pm4.37}$	$46.74_{\pm14.89}$	$40.69_{\pm4.13}$	$46.88_{\pm4.69}$	$20.74_{\pm4.12}$	$16.21_{\pm3.82}$
GPPT	$43.15_{\pm9.44}$	$37.26_{\pm6.17}$	$48.31_{\pm17.72}$	$30.40_{\pm6.81}$	$31.81_{\pm15.33}$	$22.58_{\pm1.97}$	$14.65_{\pm3.07}$
Gprompt	$56.66_{\pm11.22}$	$53.21_{\pm10.94}$	$39.74_{\pm15.35}$	$77.07_{\pm5.93}$	$33.25_{\pm40.11}$	$25.26_{\pm1.10}$	$75.72_{\pm4.95}$
All-in-one	$52.39_{\pm10.17}$	$40.41_{\pm2.80}$	$45.17_{\pm6.45}$	$66.29_{\pm19.11}$	$65.49_{\pm7.06}$	$24.61_{\pm2.80}$	$13.16_{\pm5.98}$
GPF	$38.57_{\pm5.41}$	$31.16_{\pm8.05}$	$49.99_{\pm8.86}$	$78.35_{\pm4.07}$	$73.54_{\pm18.50}$	$28.70_{\pm3.35}$	$65.11_{\pm5.70}$
GPF-plus	$55.77_{\pm10.30}$	$59.67_{\pm11.87}$	$46.64_{\pm18.97}$	$82.11_{\pm13.95}$	$76.10_{\pm20.35}$	$29.32_{\pm8.56}$	$71.98_{\pm12.23}$
MTG (Ours)	$58.54_{\pm7.89}$	$62.31_{\pm18.90}$	$50.70_{\pm11.68}$	$83.32_{\pm12.46}$	$79.13_{\pm17.18}$	$29.44_{\pm7.31}$	$75.97_{\pm4.29}$

Table 2: Performance comparison of adaptation methods on 1-shot graph classification.

Method	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	$57.30_{\pm0.98}$	$47.23_{\pm0.61}$	$56.36_{\pm7.97}$	$65.20_{\pm6.70}$	$20.58_{\pm2.00}$	$27.08_{\pm1.95}$	$25.80_{\pm6.53}$	$55.33_{\pm6.22}$
Fine-tuning	$57.75_{\pm1.22}$	$48.10_{\pm0.23}$	$63.44_{\pm3.64}$	$65.47_{\pm5.89}$	$22.21_{\pm2.79}$	$76.19_{\pm5.41}$	$34.69_{\pm8.50}$	$57.15_{\pm4.32}$
GPPT	$50.15_{\pm0.75}$	$47.18_{\pm5.93}$	$60.92_{\pm2.47}$	$60.40_{\pm15.43}$	$21.29_{\pm3.79}$	$78.23_{\pm1.38}$	$59.32_{\pm11.22}$	$57.69_{\pm6.89}$
Gprompt	$54.75_{\pm12.43}$	$48.25_{\pm13.64}$	$59.17_{\pm11.26}$	$73.60_{\pm4.76}$	$22.29_{\pm3.50}$	$54.64_{\pm9.94}$	$55.43_{\pm13.69}$	$57.81_{\pm2.68}$
All-in-one	$60.07_{\pm4.81}$	$51.66_{\pm0.26}$	$66.49_{\pm6.26}$	$75.20_{\pm6.33}$	$23.96_{\pm1.45}$	$76.14_{\pm5.51}$	$64.38_{\pm9.32}$	$59.72_{\pm1.52}$
GPF	$59.65_{\pm5.06}$	$47.42_{\pm11.22}$	$63.91_{\pm3.26}$	$68.40_{\pm5.09}$	$22.00_{\pm1.25}$	$65.79_{\pm17.72}$	$71.67_{\pm14.71}$	$59.36_{\pm1.18}$
GPF-plus	$57.93_{\pm1.62}$	$47.24_{\pm0.29}$	$62.92_{\pm2.78}$	$65.20_{\pm6.04}$	$22.92_{\pm1.64}$	$33.78_{\pm1.52}$	$71.17_{\pm14.92}$	$57.62_{\pm2.42}$
MTG (Ours)	$62.25_{\pm3.72}$	$52.25_{\pm0.56}$	$66.98_{\pm2.17}$	$75.80_{\pm5.49}$	$26.08_{\pm4.31}$	$78.27_{\pm2.01}$	$74.81_{\pm13.96}$	$60.68_{\pm2.42}$

of parameters, MTG still exhibits a substantial advantage over supervised learning and fine-tuning approaches, which underscores its high parameter efficiency. Among node-level tasks, GPF-plus is the prompt tuning method that performs closest to MTG, while on graph-level tasks, All-in-one ranks as the second most effective method after MTG. Additional experimental details are provided in Appendix F, including results for few-shot node/graph classification tasks under **3-shot and 5-shot settings** (Appendix F.1), a comparative analysis of **computational efficiency** between MTG and prompt tuning methods (Appendix F.3), along with a **sensitivity analysis** (Appendix F.4).

5.3 Robustness Performance of MTG across Pre-training Strategies

In addition to validating the upper bound performance of MTG, we further analyze whether MTG exhibits strong robustness across different pre-training strategies through more detailed experimental results. In Section 5.2, we have verified that GPF-plus and All-in-one are the best-performing prompt tuning methods for 1-shot node classification and 1-shot graph classification tasks, respectively. Therefore, we selected these two methods along with Fine-tuning for a more detailed comparison with MTG. In our experiments, we employ three pre-training strategies at the node, edge, and graph levels to obtain pre-trained models, which have varying impacts on different datasets. As shown in Tables 3 and 4, Fine-tuning experiences performance collapse on the ogbn-arxiv dataset under the DGI and EdgePreGprompt pre-training strategies, with the accuracy dropping as low as 4.65%. Similarly, GPF-plus exhibits performance degradation on the Cora dataset under the DGI and SimGRACE pre-training strategies, achieving an accuracy as low as 17.29%. All-in-one also shows relatively low performance on the IMDB-B dataset under the GraphMAE and EdgePreGprompt pre-training strategies. These results indicate that their adaptability varies significantly across different pre-training strategies. In contrast, MTG demonstrates relatively more stable performance across all datasets and all pre-training strategies, highlighting broader compatibility and better robustness when combined with various types of pre-training strategies.

5.4 Mitigation of Negative Transfer

Compared to visual images and natural language, fine-tuning pre-trained models on graph data for downstream tasks is more prone to negative transfer. Therefore, the ability to effectively mitigate negative transfer serves as an important criterion for evaluating the quality of an adaptation method. As shown in Tables 3 and 4, prompt tuning methods such as GPF-plus and All-in-one have already

Table 3: Performance comparison of Fine-tuning, GPF-plus, and MTG on 1-shot node classification. ↑/↓: positive/negative transfer vs. supervised learning baseline; **NTR** (Negative Transfer Rate): fraction of datasets with ↓ per daptation method.

Pre-training	**Adaptation**	**NTR**	**Cora**	**Citeseer**	**Pubmed**	**Wisconsin**	**Texas**	**Actor**	**ogbn-arxiv**
-	Supervised	**0%**	26.56$_{\pm5.55}$ (-)	21.78$_{\pm7.32}$ (-)	39.37$_{\pm16.34}$ (-)	41.60$_{\pm3.10}$ (-)	37.97$_{\pm5.80}$ (-)	20.57$_{\pm4.47}$ (-)	10.99$_{\pm3.19}$ (-)
DGI	Fine-tuning	**57%**	33.15$_{\pm7.84}$ (↑)	**21.64**$_{\pm3.92}$ (↓)	42.01$_{\pm12.54}$ (↑)	**37.49**$_{\pm7.56}$ (↓)	45.31$_{\pm5.01}$ (↑)	**19.76**$_{\pm3.53}$ (↓)	**7.21**$_{\pm2.91}$ (↓)
	GPF-plus	**29%**	**17.29**$_{\pm6.18}$ (↓)	26.60$_{\pm13.24}$ (↑)	**34.02**$_{\pm11.94}$ (↓)	74.68$_{\pm11.81}$ (↑)	71.44$_{\pm18.66}$ (↑)	22.42$_{\pm9.66}$ (↑)	16.83$_{\pm10.02}$ (↑)
	MTG	**0%**	49.48$_{\pm4.82}$ (↑)	62.31$_{\pm18.90}$ (↑)	46.18$_{\pm7.32}$ (↑)	67.72$_{\pm10.19}$ (↑)	62.96$_{\pm16.80}$ (↑)	25.48$_{\pm7.33}$ (↑)	25.06$_{\pm10.57}$ (↑)
GraphMAE	Fine-tuning	**57%**	32.93$_{\pm3.17}$ (↑)	**21.26**$_{\pm3.57}$ (↓)	42.99$_{\pm14.25}$ (↑)	**36.80**$_{\pm7.17}$ (↓)	**37.81**$_{\pm8.62}$ (↓)	**19.86**$_{\pm2.70}$ (↓)	12.35$_{\pm3.60}$ (↑)
	GPF-plus	**0%**	54.26$_{\pm7.48}$ (↑)	59.67$_{\pm11.87}$ (↑)	46.64$_{\pm18.57}$ (↑)	82.11$_{\pm13.95}$ (↑)	70.95$_{\pm18.63}$ (↑)	26.58$_{\pm7.84}$ (↑)	49.81$_{\pm2.62}$ (↑)
	MTG	**0%**	46.27$_{\pm6.66}$ (↑)	49.21$_{\pm12.95}$ (↑)	46.98$_{\pm10.02}$ (↑)	83.32$_{\pm12.46}$ (↑)	71.59$_{\pm18.67}$ (↑)	29.44$_{\pm7.31}$ (↑)	36.44$_{\pm9.59}$ (↑)
EdgePre -GPPT	Fine-tuning	**43%**	38.12$_{\pm5.29}$ (↑)	**18.09**$_{\pm5.39}$ (↓)	46.74$_{\pm14.09}$ (↑)	**35.31**$_{\pm9.31}$ (↓)	47.66$_{\pm2.37}$ (↑)	**19.17**$_{\pm2.53}$ (↓)	16.21$_{\pm3.82}$ (↑)
	GPF-plus	**14%**	**28.49**$_{\pm18.73}$ (↓)	28.04$_{\pm14.31}$ (↑)	46.51$_{\pm15.84}$ (↑)	72.66$_{\pm12.05}$ (↑)	70.67$_{\pm17.59}$ (↑)	29.32$_{\pm8.56}$ (↑)	71.98$_{\pm12.23}$ (↑)
	MTG	**0%**	46.68$_{\pm2.66}$ (↑)	33.22$_{\pm12.52}$ (↑)	44.85$_{\pm9.75}$ (↑)	73.80$_{\pm9.56}$ (↑)	71.11$_{\pm17.13}$ (↑)	20.96$_{\pm2.93}$ (↑)	75.97$_{\pm4.29}$ (↑)
EdgePre -Gprompt	Fine-tuning	**14%**	35.57$_{\pm5.83}$ (↑)	22.28$_{\pm3.80}$ (↑)	41.50$_{\pm7.54}$ (↑)	**40.69**$_{\pm4.13}$ (↓)	40.62$_{\pm7.95}$ (↑)	20.74$_{\pm4.16}$ (↑)	14.83$_{\pm2.38}$ (↑)
	GPF-plus	**0%**	55.77$_{\pm10.30}$ (↑)	49.43$_{\pm8.21}$ (↑)	42.79$_{\pm18.18}$ (↑)	78.76$_{\pm13.63}$ (↑)	68.75$_{\pm16.51}$ (↑)	22.68$_{\pm3.64}$ (↑)	57.44$_{\pm6.95}$ (↑)
	MTG	**0%**	46.29$_{\pm3.84}$ (↑)	45.30$_{\pm16.04}$ (↑)	50.70$_{\pm11.68}$ (↑)	72.75$_{\pm11.21}$ (↑)	79.13$_{\pm17.18}$ (↑)	21.34$_{\pm1.78}$ (↑)	21.08$_{\pm2.34}$ (↑)
GraphCL	Fine-tuning	**43%**	52.61$_{\pm1.73}$ (↑)	27.02$_{\pm4.31}$ (↑)	42.49$_{\pm11.29}$ (↑)	**33.94**$_{\pm7.74}$ (↓)	40.31$_{\pm13.68}$ (↑)	**20.19**$_{\pm1.98}$ (↓)	**4.65**$_{\pm1.19}$ (↓)
	GPF-plus	**29%**	34.18$_{\pm17.71}$ (↓)	28.86$_{\pm22.88}$ (↑)	37.02$_{\pm11.29}$ (↓)	52.35$_{\pm19.69}$ (↑)	75.40$_{\pm19.10}$ (↑)	22.82$_{\pm4.99}$ (↑)	32.11$_{\pm4.86}$ (↑)
	MTG	**0%**	58.54$_{\pm7.89}$ (↑)	50.96$_{\pm16.40}$ (↑)	40.00$_{\pm7.80}$ (↑)	48.41$_{\pm16.10}$ (↑)	69.71$_{\pm16.42}$ (↑)	24.77$_{\pm8.45}$ (↑)	38.96$_{\pm6.82}$ (↑)
SimGRACE	Fine-tuning	**57%**	40.40$_{\pm4.66}$ (↑)	35.05$_{\pm4.37}$ (↑)	**37.59**$_{\pm8.17}$ (↓)	**37.37**$_{\pm3.68}$ (↓)	46.88$_{\pm4.64}$ (↑)	**19.78**$_{\pm1.89}$ (↓)	**8.13**$_{\pm3.26}$ (↓)
	GPF-plus	**29%**	**21.33**$_{\pm14.86}$ (↓)	24.61$_{\pm21.21}$ (↑)	**35.90**$_{\pm9.06}$ (↓)	73.49$_{\pm14.17}$ (↑)	76.10$_{\pm20.35}$ (↑)	20.51$_{\pm4.24}$ (↑)	46.71$_{\pm3.17}$ (↑)
	MTG	**0%**	45.93$_{\pm7.67}$ (↑)	57.60$_{\pm9.01}$ (↑)	43.29$_{\pm10.80}$ (↑)	72.98$_{\pm9.75}$ (↑)	73.17$_{\pm16.68}$ (↑)	22.03$_{\pm3.59}$ (↑)	37.90$_{\pm5.83}$ (↑)

Table 4: Performance comparison of Fine-tuning, All-in-one, and MTG on 1-shot graph classification.

Pre-training	**Adaptation**	**NTR**	**IMDB-B**	**COLLAB**	**PROTEINS**	**MUTAG**	**ENZYMES**	**COX2**	**BZR**	**D&D**
-	Supervised	**0%**	57.30$_{\pm0.98}$ (-)	47.23$_{\pm0.61}$ (-)	56.36$_{\pm7.97}$ (-)	65.20$_{\pm6.70}$ (-)	20.58$_{\pm2.00}$ (-)	27.08$_{\pm11.94}$ (-)	25.80$_{\pm6.53}$ (-)	55.33$_{\pm6.22}$ (-)
DGI	Fine-tuning	**38%**	57.32$_{\pm0.90}$ (↑)	**42.22**$_{\pm0.73}$ (↓)	64.65$_{\pm2.10}$ (↑)	**64.13**$_{\pm7.90}$ (↓)	**17.83**$_{\pm1.88}$ (↓)	29.44$_{\pm9.68}$ (↑)	26.48$_{\pm7.61}$ (↑)	57.15$_{\pm4.32}$ (↑)
	All-in-one	**13%**	60.07$_{\pm4.81}$ (↑)	**39.56**$_{\pm5.00}$ (↓)	62.58$_{\pm7.07}$ (↑)	73.87$_{\pm6.13}$ (↑)	23.96$_{\pm1.45}$ (↑)	50.72$_{\pm9.93}$ (↑)	64.38$_{\pm9.32}$ (↑)	55.97$_{\pm6.52}$ (↑)
	MTG	**13%**	59.15$_{\pm5.44}$ (↑)	**43.46**$_{\pm6.83}$ (↓)	62.78$_{\pm2.36}$ (↑)	65.60$_{\pm7.29}$ (↑)	24.71$_{\pm1.88}$ (↑)	51.74$_{\pm13.90}$ (↑)	74.81$_{\pm13.96}$ (↑)	56.39$_{\pm3.27}$ (↑)
GraphMAE	Fine-tuning	**0%**	57.70$_{\pm1.13}$ (↑)	48.10$_{\pm0.23}$ (↑)	63.57$_{\pm3.57}$ (↑)	65.20$_{\pm5.00}$ (-)	22.21$_{\pm2.79}$ (↑)	28.47$_{\pm14.72}$ (↑)	25.80$_{\pm6.53}$ (-)	57.54$_{\pm4.41}$ (↑)
	All-in-one	**25%**	**52.62**$_{\pm3.04}$ (↓)	**40.82**$_{\pm14.63}$ (↓)	66.49$_{\pm6.26}$ (↑)	69.67$_{\pm9.13}$ (↑)	23.21$_{\pm1.72}$ (↑)	56.68$_{\pm7.38}$ (↑)	58.64$_{\pm19.59}$ (↑)	58.77$_{\pm1.05}$ (↑)
	MTG	**0%**	58.10$_{\pm5.72}$ (↑)	48.24$_{\pm9.56}$ (↑)	59.62$_{\pm6.41}$ (↑)	66.93$_{\pm7.03}$ (↑)	22.71$_{\pm2.58}$ (↑)	58.93$_{\pm12.05}$ (↑)	54.07$_{\pm18.34}$ (↑)	58.01$_{\pm5.85}$ (↑)
EdgePre -GPPT	Fine-tuning	**63%**	**57.20**$_{\pm0.85}$ (↓)	**47.14**$_{\pm0.55}$ (↓)	58.27$_{\pm10.66}$ (↑)	**64.27**$_{\pm4.73}$ (↓)	**19.79**$_{\pm2.17}$ (↓)	27.83$_{\pm13.44}$ (↑)	72.10$_{\pm14.30}$ (↑)	**52.82**$_{\pm9.38}$ (↓)
	All-in-one	**13%**	59.12$_{\pm0.77}$ (↑)	**42.74**$_{\pm4.65}$ (↓)	65.71$_{\pm5.49}$ (↑)	75.20$_{\pm6.33}$ (↑)	20.92$_{\pm2.04}$ (↑)	60.27$_{\pm16.97}$ (↑)	59.69$_{\pm9.90}$ (↑)	56.24$_{\pm2.46}$ (↑)
	MTG	**13%**	62.25$_{\pm3.72}$ (↑)	**45.15**$_{\pm6.00}$ (↓)	62.71$_{\pm2.30}$ (↑)	67.20$_{\pm6.36}$ (↑)	26.08$_{\pm4.31}$ (↑)	60.16$_{\pm10.63}$ (↑)	62.28$_{\pm10.13}$ (↑)	56.37$_{\pm8.33}$ (↑)
EdgePre -Gprompt	Fine-tuning	**38%**	57.35$_{\pm0.92}$ (↑)	**47.20**$_{\pm0.53}$ (↓)	61.84$_{\pm2.59}$ (↑)	**62.67**$_{\pm2.67}$ (↓)	**19.75**$_{\pm2.33}$ (↓)	27.13$_{\pm12.05}$ (↑)	29.44$_{\pm11.20}$ (↑)	56.16$_{\pm5.10}$ (↑)
	All-in-one	**25%**	**53.78**$_{\pm2.82}$ (↓)	**42.87**$_{\pm6.19}$ (↓)	61.82$_{\pm7.53}$ (↑)	68.27$_{\pm3.88}$ (↑)	21.88$_{\pm0.56}$ (↑)	49.06$_{\pm5.53}$ (↑)	32.65$_{\pm10.08}$ (↑)	57.60$_{\pm4.37}$ (↑)
	MTG	**0%**	59.45$_{\pm5.45}$ (↑)	47.72$_{\pm8.45}$ (↑)	65.66$_{\pm1.56}$ (↑)	75.80$_{\pm5.49}$ (↑)	22.29$_{\pm1.94}$ (↑)	57.75$_{\pm10.76}$ (↑)	49.94$_{\pm9.08}$ (↑)	60.68$_{\pm2.42}$ (↑)
GraphCL	Fine-tuning	**25%**	57.75$_{\pm1.02}$ (↑)	**39.62**$_{\pm0.63}$ (↓)	63.44$_{\pm3.64}$ (↑)	**65.07**$_{\pm8.38}$ (↓)	23.96$_{\pm1.99}$ (↑)	53.14$_{\pm21.32}$ (↑)	29.07$_{\pm7.00}$ (↑)	60.62$_{\pm1.56}$ (↑)
	All-in-one	**13%**	58.75$_{\pm0.80}$ (↑)	51.66$_{\pm2.60}$ (↑)	66.00$_{\pm8.79}$ (↑)	66.00$_{\pm8.79}$ (↑)	**19.46**$_{\pm2.85}$ (↓)	52.55$_{\pm13.51}$ (↑)	42.65$_{\pm14.43}$ (↑)	59.72$_{\pm1.52}$ (↑)
	MTG	**0%**	57.65$_{\pm7.05}$ (↑)	47.81$_{\pm3.73}$ (↑)	63.70$_{\pm2.87}$ (↑)	66.20$_{\pm7.52}$ (↑)	20.96$_{\pm1.97}$ (↑)	50.36$_{\pm12.97}$ (↑)	51.05$_{\pm15.50}$ (↑)	55.46$_{\pm4.77}$ (↑)
SimGRACE	Fine-tuning	**38%**	57.33$_{\pm0.96}$ (↑)	**46.89**$_{\pm0.42}$ (↓)	60.07$_{\pm3.21}$ (↑)	65.47$_{\pm5.89}$ (↑)	**19.71**$_{\pm1.76}$ (↓)	76.19$_{\pm5.41}$ (↑)	28.48$_{\pm6.49}$ (↑)	**53.23**$_{\pm9.71}$ (↓)
	All-in-one	**0%**	58.83$_{\pm0.85}$ (↑)	47.60$_{\pm3.90}$ (↑)	66.20$_{\pm7.52}$ (↑)	66.67$_{\pm5.73}$ (↑)	22.50$_{\pm1.56}$ (↑)	76.14$_{\pm5.51}$ (↑)	59.01$_{\pm12.34}$ (↑)	58.26$_{\pm1.18}$ (↑)
	MTG	**0%**	61.82$_{\pm3.49}$ (↑)	52.25$_{\pm0.56}$ (↑)	66.98$_{\pm2.17}$ (↑)	68.87$_{\pm5.01}$ (↑)	21.33$_{\pm1.92}$ (↑)	78.27$_{\pm2.01}$ (↑)	65.68$_{\pm16.41}$ (↑)	57.26$_{\pm2.01}$ (↑)

relatively alleviated negative transfer compared to fine-tuning, while our proposed MTG demonstrates a more significant advantage in mitigating negative transfer. Under all pre-training strategies across the two downstream tasks, MTG achieves a lower negative transfer rate than GPF-plus and All-in-one, and is markedly superior to fine-tuning. It is particularly noteworthy that MTG completely eliminates negative transfer in the 1-shot node classification task, highlighting its exceptional capability in mitigating such issues. A theoretical analysis of **negative transfer** is presented in Appendix C.2.

6 Conclusion

In this paper, we propose Prismatic Space Theory to quantify the capacity of adaptation approaches and establish the upper bound for the adaptation capacity of graph prompt tuning. Building on these insights, we introduce Message Tuning for GFMs (MTG), a lightweight adaptation method that dynamically guides message fusion across GNN layers while keeping pre-trained weights frozen. Theoretical and empirical results demonstrate MTG's consistent superiority over prompt tuning.

Ethics Statement

We acknowledge the ICLR Code of Ethics and confirm that our work adheres to its principles. Our research prioritizes societal benefit, avoids harm, and respects privacy and intellectual property. All data used in this study comply with ethical guidelines and relevant licenses.

Reproducibility Statement

To ensure the reproducibility of our work, we have made substantial efforts to provide comprehensive details and resources across our main paper, appendix, and supplementary materials.

Code and Resources. We have developed a reproducible codebase MTG based on the ProG library (Zi et al., 2024), extended to support our message tuning. Our code is available at https://anonymous.4open.science/r/MTG. Anonymous, downloadable source code also includes scripts for pre-training, adaptation, and evaluation on all datasets used in our experiments.

Data Processing. A detailed description of all datasets used in the experiments and dataset preprocessing steps, including feature extraction, graph normalization, and train/validation/test splits for few-shot settings (1/3/5-shot), is provided in Appendix D.

Computational Resources. Hardware specifications and software environments are described in Appendix D.1 to facilitate replication of computational experiments.

Theoretical Proofs. All theoretical claims, including the Prismatic Space Theory and its application to the analysis of prompt tuning and message tuning, are rigorously proven in Appendix B and C.

We believe these efforts collectively ensure the reproducibility of our work and encourage the community to build upon our findings.

References

Raman Arora, Amitabh Basu, Poorya Mianjy, and Anirbit Mukherjee. Understanding deep neural networks with rectified linear units. *International Conference on Learning Representations*, 2018.

Kosio Beshkov. A relative homology theory of representation in neural networks. *arXiv preprint arXiv:2502.01360*, 2025.

Karsten M. Borgwardt, Cheng Soon Ong, Stefan Schönauer, S. V. N. Vishwanathan, Alex J. Smola, and Hans-Peter Kriegel. Protein function prediction via graph kernels. *Bioinformatics*, 2005.

Haibo Chen, Xin Wang, Zeyang Zhang, Haoyang Li, Ling Feng, and Wenwu Zhu. Autogfm: Automated graph foundation model with adaptive architecture customization. In *International Conference on Machine Learning*, 2025.

Cameron Diao, Kaixiong Zhou, Zirui Liu, Xiao Huang, and Xia Hu. Molcpt: Molecule continuous prompt tuning to generalize molecular representation learning. *arXiv preprint arXiv:2212.10614*, 2023.

Paul D. Dobson and Andrew J. Doig. Distinguishing enzyme structures from non-enzymes without alignments. *Journal of Molecular Biology*, 2003.

Taoran Fang, Yunchao Zhang, Yang Yang, Chunping Wang, and Lei Chen. Universal prompt tuning for graph neural networks. *Advances in Neural Information Processing Systems*, 2023.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. In *International Conference on Learning Representations*, 2019.

Yaoying Fu. Toric geometry of relu neural networks. *arXiv preprint arXiv:2509.05894*, 2025.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, 2017.

Werner Greub. Linear algebra. In *Graduate Texts in Mathematics*. Springer, 1975.

Anchun Gui, Jinqiang Ye, and Han Xiao. G-adapter: Towards structure-aware parameter-efficient transfer learning for graph transformer networks. In *Thirty-Eighth AAAI Conference on Artificial Intelligence*, 2024.

Paul R. Halmos. Measure theory. In *Graduate Texts in Mathematics*. Springer, 1950.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 2017.

Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders. In *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in Neural Information Processing Systems*, 33:22118–22133, 2020a.

Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. Gpt-gnn: Generative pre-training of graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2020b.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Steven Krantz and Harold Parks. *Geometric Integration Theory*. Birkhäuser Boston, MA, 2008.

Nils Kriege and Petra Mutzel. Subgraph matching kernels for attributed graphs. In *International Conference on Machine Learning*, 2012.

Serge Lang. Real and functional analysis. In *Graduate Texts in Mathematics*. Springer, 1993.

John M. Lee. Introduction to topological manifolds. In *Graduate Texts in Mathematics*. Springer, 2011.

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.

Shengrui Li, Xueting Han, and Jing Bai. Adaptergnn: Parameter-efficient fine-tuning improves generalization in gnns. In *Thirty-Eighth AAAI Conference on Artificial Intelligence*, 2024.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association of Computational Linguistics*, 2021.

Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S. Yu, and Chuan Shi. Graph foundation models: Concepts, opportunities and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Xiao Liu, Kaixuan Ji, Yicheng Fu, Weng Lam Tam, Zhengxiao Du, Zhilin Yang, and Jie Tang. P-tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks. In *Proceedings of the 60th Annual Meeting of the Association of Computational Linguistics*, 2022.

Yajing Liu, Christina M Cole, Chris Peterson, and Michael Kirby. Relu neural networks, polyhedral decompositions, and persistent homolog. In *International Conference on Machine Learning*, 2023a.

Zemin Liu, Xingtong Yu, Yuan Fang, and Xinming Zhang. Graphprompt: Unifying pre-training and downstream tasks for graph neural networks. In *The ACM Web Conference*, 2023b.

Yuankai Luo, Lei Shi, and Xiao-Ming Wu. Classic gnns are strong baselines: Reassessing gnns for node classification. *Advances in Neural Information Processing Systems*, 2024.

Yuankai Luo, Lei Shi, and Xiao-Ming Wu. Can classic gnns be strong baselines for graph-level tasks? simple architectures meet excellence. In *International Conference on Machine Learning*, 2025.

Maas, Awni Y. Hannun, and Andrew Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *International Conference on Machine Learning*, 2013.

André F. T. Martins and Ramón Fernandez Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *International Conference on Machine Learning*, 2016.

Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *International Conference on Machine Learning*, 2010.

Chaoxi Niu, Guansong Pang, Ling Chen, and Bing Liu. Replay-and-forget-free graph class-incremental learning: A task profiling and prompting approach. *Advances in Neural Information Processing Systems*, 2024.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 2019.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. *International Conference on Learning Representations*, 2020.

Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. Gcc: Graph contrastive coding for graph neural network pre-training. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2020.

Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems*, 2020.

Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI Conference on Artificial Intelligence*, 2015.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina EliassiRad. Collective classification in network data. *AI magazine*, pp. 29(3), 2008.

Mingchen Sun, Kaixiong Zhou, Xin He, Ying Wang, and Xin Wang. Gppt: Graph pre-training and prompt tuning to generalize graph neural networks. In *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022.

Xiangguo Sun, Hong Cheng, Jia Li, Bo Liu, and Jihong Guan. All in one: Multi-task prompting for graph neural networks. In *Proceedings of the 29th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2023a.

Xiangguo Sun, Jiawen Zhang, Xixi Wu, Hong Cheng, Yun Xiong, and Jia Li. Graph prompt learning: A comprehensive survey and beyond. In *IEEE Transactions on Knowledge and Data Engineering*, 2023b.

Yifei Sun, Qi Zhu, Yang Yang, Chunping Wang, Tianyu Fan, Jiajun Zhu, and Lei Chen. Fine-tuning graph neural networks by preserving graph generative patterns. In *Thirty-Eighth AAAI Conference on Artificial Intelligence*, 2024.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. In *International Conference on Learning Representations*, 2019.

Liyuan Wang, Mingtian Zhang, Zhongfan Jia, Qian Li, Chenglong Bao, Kaisheng Ma, Jun Zhu, and Yi Zhong. Afec: Active forgetting of negative transfer in continual learning. *Advances in Neural Information Processing Systems*, 2021.

Qunzhong Wang, Xiangguo Sun, and Hong Cheng. Does graph prompt work? a data operation perspective with theoretical analysis. In *International Conference on Machine Learning*, 2025a.

Song Wang, Yushun Dong, Xiao Huang, Chen Chen, and Jundong Li. Faith: Few-shot graph classification with hierarchical task graphs. In *International Joint Conference on Artificial Intelligence*, 2022.

Zehong Wang, Zheyuan Zhang, Nitesh V Chawla, Chuxu Zhang, and Yanfang Ye. Gft: Graph foundation model with transferable tree vocabulary. *Advances in Neural Information Processing Systems*, 2024.

Zehong Wang, Zheyuan Liu, Tianyi Ma, Jiazheng Li, Zheyuan Zhang, Xingbo Fu, Yiyang Li, Zhengqing Yuan, Wei Song, Yijun Ma, Qingkai Zeng, Xiusi Chen, Jianan Zhao, Jundong Li, Meng Jiang, Pietro Lio, Nitesh Chawla, Chuxu Zhang, and Yanfang Ye. Graph foundation models: A comprehensive survey. *arXiv preprint arXiv:2505.15116*, 2025b.

Jun Xia, Lirong Wu, Jintao Chen, Bozhen Hu, and Stan Z. Li. Simgrace: A simple framework for graph contrastive learning without data augmentation. In *The ACM Web Conference*, 2022.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

Pinar Yanardag and Svn V N Vishwanathan. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

Haoran Yang, Xiangyu Zhao, Yicong Li, Hongxu Chen, and Guandong Xu. An empirical study towards prompt-tuning for graph contrastive pre-training in recommendations. In *Advances in Neural Information Processing Systems*, 2023.

Zhe-Rui Yang, Jindong Han, Chang-Dong Wang, and Hao Liu. Graphlora: Structure-aware contrastive low-rank adaptation for cross-graph transfer learning. In *Proceedings of the 31th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2025.

Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. In *International Conference on Machine Learning*, 2016.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? *Advances in Neural Information Processing Systems*, 2021.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 2020.

Xingtong Yu, Jie Zhang, Yuan Fang, and Renhe Jiang. Non-homophilic graph pre-training and prompt learning. In *Proceedings of the 31th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2025.

Chuxu Zhang, Kaize Ding, Jundong Li, Xiangliang Zhang, Yanfang Ye, Nitesh V. Chawla, and Huan Liu. Few-shot learning on graphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022.

Xiao Zhang and Dongrui Wu. Empirical studies on the properties of linear regions in deep neural networks. *International Conference on Learning Representations*, 2020.

Chenyi Zi, Haihong Zhao, Xiangguo Sun, Yiqing Lin, Hong Cheng, and Jia Li. Prog: A graph prompt learning benchmark. *Advances in Neural Information Processing Systems*, 2024.

Appendix

A The Use of Large Language Models

In the preparation of this work, we use large language models (LLMs) to assist with proofreading, grammatical correction, and language polishing. The LLM serves solely as a tool to enhance the clarity and readability of our writing. We meticulously review and edit all AI-generated content, and we accept full responsibility for the final version of the manuscript.

B Extra Materials for Prismatic Space Theory

Reading Guideline: Appendix B is organized in strict accordance with the order in which definitions, theorems, and corollaries appear in the main text, serving as a detailed supplement. This includes supplementary explanations of definitions, lemmas required for proving theorems, interpretations of theorems, and more. We recommend that readers first review the related work, such as other related theoretical works and relevant mathematical textbooks, to establish a theoretical foundation before proceeding through the main text in sequence with the aid of Appendix B.

Notation: The notation used in this paper has been aligned as closely as possible with the standardized notation recommended by the ICLR conference (https://github.com/goodfeli/dlbook_notation/). A summary of the primary notation used is provided in the table below.

Table 5: Primary Notation.

Notation	**Description**
$\mathcal{G}$	A graph
$\mathcal{V}$	The set of nodes
$\mathcal{E}$	The set of edges
N	Number of nodes
ℓ	Number of layers
$\mathbb{R}$	The set of real numbers
$\{0, 1\}$	The set containing 0 and 1
$\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{H}^{(\ell)}$	The matrices
$\mathfrak{A}, \mathfrak{M}, \mathfrak{U}$	The operators
$\boldsymbol{\Theta}$	The parameters
$\mathbb{V}, \mathbb{S}, \mathbb{H}^{(\ell)}$	The sets
$F^{(\ell)} : \mathbb{H}^{(\ell-1)} \rightarrow \mathbb{H}^{(\ell)}$	The map $F^{(\ell)}$ with domain $\mathbb{H}^{(\ell-1)}$ and range $\mathbb{H}^{(\ell)}$
$\mathcal{X}, \mathcal{M}$	The manifold or space
$\Phi^{(\ell)} = F^{(\ell)} \circ \cdots \circ F^{(1)}$	The composition of maps
d_{int}	The intrinsic dimension
$\boldsymbol{J}^{(\ell)}(\boldsymbol{H})$	The Jacobian matrix
$\sigma_i^{(\ell)}$	The singular values
$\mathcal{H}^s$	The s-dimensional Hausdorff measure
$\{C_k\}$	The set of cells
$R_1, ..., R_L$	The regions
$\Omega^{(\ell)}$	The set of polytopic regions

B.1 Related Work on Prismatic Space Theory

We introduce, for the first time, the Prismatic Space Theory to provide a unified analysis of adaptation methods for graph foundation models. In constructing this theoretical framework, we adopt the perspective of piecewise linear maps, an approach that is not entirely new, as several outstanding theoretical studies have employed similar ideas to analyze ReLU neural networks (Arora et al., 2018; Zhang & Wu, 2020; Liu et al., 2023a; Fu, 2025; Beshkov, 2025).

Previously, only Wang et al. (2025a) conducted theoretical research on prompt tuning for GFMs, explaining why prompt tuning works from the perspective of data operations, primarily using mathematical tools from linear algebra, convex optimization, and probability. In contrast, our Prismatic Space Theory offers a more profound and fundamental geometric perspective to quantify the upper bound of prompt tuning's capability. Some mathematical concepts not explicitly defined or elaborated in this paper can be found in Halmos (1950); Greub (1975); Lang (1993); Krantz & Parks (2008); Lee (2011).

B.2 Details of Definition 1

The unified GFM layer formulation provided in Definition 1 offers a general framework that encapsulates a wide range of popular GNN architectures. The three core operators $\mathfrak{A}^{(\ell)}$ (attention), $\mathfrak{M}^{(\ell)}$ (message fusion), and $\mathfrak{U}^{(\ell)}$ (update) can be instantiated in different ways to recover specific models. Below, we delineate how several classic models are special cases of this unified formulation.

GCN (Graph Convolutional Network) (Kipf & Welling, 2017) employs a fixed, non-learnable attention mechanism based on the normalized adjacency matrix and a simple update function.

Attention Operator $\mathfrak{A}^{(\ell)}$ computes a static, structural attention weight for each edge (i, j) based on the normalized adjacency matrix:

$$\mathfrak{A}^{(\ell)}\left(\boldsymbol{A}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_a^{(\ell)}\right) = \tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}, \tag{17}$$

where $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_N$ is the adjacency matrix with self-loops and $\tilde{\boldsymbol{D}}$ is the corresponding degree matrix.

Message Fusion Operator $\mathfrak{M}^{(\ell)}$ performs a weighted sum of the neighbors' features using the normalized adjacency matrix:

$$\mathfrak{M}^{(\ell)}\left(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_m^{(\ell)}\right) = \tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}^{(\ell)}, \tag{18}$$

where $\boldsymbol{W}^{(\ell)} \in \mathbb{R}^{d_{\ell-1}\times d_\ell}$ is a learnable weight matrix ($\boldsymbol{\Theta}_m^{(\ell)} = \boldsymbol{W}^{(\ell)}$).

Update Operator $\mathfrak{U}^{(\ell)}$ applies a non-linear activation function σ (e.g., ReLU) to the aggregated messages:

$$\mathfrak{U}^{(\ell)}\left(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}^{(\ell)}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_u^{(\ell)}\right) = \sigma\left(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}^{(\ell)}\right), \tag{19}$$

where the previous representation $\boldsymbol{H}^{(\ell-1)}$ is not explicitly used in the update, making the update a direct transformation of the messages ($\boldsymbol{\Theta}_u^{(\ell)} = \varnothing$).

The resulting layer formulation is:

$$\boldsymbol{H}^{(\ell)} = \sigma\left(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}^{(\ell)}\right). \tag{20}$$

GraphSAGE (Hamilton et al., 2017) employs a uniform (or degree-based) attention weight over the sampled neighborhood, a configurable message aggregation function (e.g., mean, pool, LSTM), and an update function that concatenates the node's previous representation with the aggregated message.

Attention Operator $\mathfrak{A}^{(\ell)}$ often uses a static, uniform attention weight $\frac{1}{|\mathcal{N}(i)|}$ for each sampled neighbor of node i, or a learned weight based on node degree in some variants. It can be represented as a matrix $\boldsymbol{S}^{(\ell)}$:

$$\mathfrak{A}^{(\ell)}\left(\boldsymbol{A}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_a^{(\ell)}\right) = \boldsymbol{S}^{(\ell)}, \quad \boldsymbol{S}_{ij}^{(\ell)} = \begin{cases} \frac{1}{|\mathcal{N}(i)|} & \text{if } j \in \mathcal{N}(i) \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

where $\mathcal{N}(i)$ denotes the sampled neighbors of node i.

Message Fusion Operator $\mathfrak{M}^{(\ell)}$ aggregates messages from the sampled neighborhood using the specified aggregator $\mathrm{AGGREGATE}^{(\ell)}$ (e.g., mean, pool, LSTM). For the mean aggregator:

$$\mathfrak{M}^{(\ell)}\left(\boldsymbol{S}^{(\ell)}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_m^{(\ell)}\right) = \boldsymbol{S}^{(\ell)}\boldsymbol{H}^{(\ell-1)}, \tag{22}$$

where parameters $\boldsymbol{\Theta}_m^{(\ell)}$ depend on the choice of aggregator.

Update Operator $\mathfrak{U}^{(\ell)}$ concatenates the node's previous representation $\boldsymbol{H}^{(\ell-1)}$ with the aggregated neighborhood message, applies a linear transformation $\boldsymbol{W}^{(\ell)}$, and a non-linear activation function σ:

$$\mathfrak{U}^{(\ell)}\left(\boldsymbol{S}^{(\ell)}\boldsymbol{H}^{(\ell-1)}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_u^{(\ell)}\right) = \sigma\left(\boldsymbol{W}^{(\ell)} \cdot \mathrm{CONCAT}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{S}^{(\ell)}\boldsymbol{H}^{(\ell-1)})\right), \tag{23}$$

where $\boldsymbol{\Theta}_u^{(\ell)} = \boldsymbol{W}^{(\ell)}$.

The resulting layer formulation for the mean aggregator is:

$$\boldsymbol{H}^{(\ell)} = \sigma\left(\boldsymbol{W}^{(\ell)} \cdot \mathrm{CONCAT}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{S}^{(\ell)}\boldsymbol{H}^{(\ell-1)})\right). \tag{24}$$

GAT (Graph Attention Network) (Veličković et al., 2018) introduces a learnable self-attention mechanism to compute dynamic attention weights between nodes.

Attention Operator $\mathfrak{A}^{(\ell)}$ computes pairwise attention coefficients α_{ij} for nodes i and j using a learnable function (a shared attentional mechanism a):

$$e_{ij}^{(\ell)} = a(\boldsymbol{W}^{(\ell)}\boldsymbol{h}_i^{(\ell-1)}, \boldsymbol{W}^{(\ell)}\boldsymbol{h}_j^{(\ell-1)}) = \mathrm{LeakyReLU}\left(\tilde{\mathbf{a}}^{(\ell)T}[\boldsymbol{W}^{(\ell)}\boldsymbol{h}_i^{(\ell-1)} \| \boldsymbol{W}^{(\ell)}\boldsymbol{h}_j^{(\ell-1)}]\right), \tag{25}$$

$$\alpha_{ij}^{(\ell)} = \mathrm{Softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})}, \tag{26}$$

$$\mathfrak{A}^{(\ell)}\left(\boldsymbol{A}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_a^{(\ell)}\right) = \boldsymbol{A}_a^{(\ell)}, \quad \boldsymbol{A}_a^{(\ell)}{}_{ij} = \begin{cases} \alpha_{ij}^{(\ell)} & \text{if } j \in \mathcal{N}(i) \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

where $\boldsymbol{h}_i^{(\ell-1)}$ represents the vector of node i at the $\ell - 1$-th layer, ^{T} represents transposition, $\|$ is the concatenation operation, and the attention mechanism a is a single-layer feedforward neural network parametrized by a weight vector $\tilde{\mathbf{a}}^{(\ell)}$ ($\boldsymbol{\Theta}_a^{(\ell)} = \{\tilde{\mathbf{a}}^{(\ell)}, \boldsymbol{W}^{(\ell)}\}$).

Message Fusion Operator $\mathfrak{M}^{(\ell)}$ performs a weighted sum of the transformed neighbor features based on the computed attention weights:

$$\mathfrak{M}^{(\ell)}\left(\boldsymbol{A}_a^{(\ell)}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_m^{(\ell)}\right) = \boldsymbol{A}_a^{(\ell)} \cdot (\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}^{(\ell)}), \tag{28}$$

where $\boldsymbol{W}^{(\ell)}$ is a shared linear transformation applied to every node and is also used by the Attention Operator ($\boldsymbol{\Theta}_m^{(\ell)} = \boldsymbol{W}^{(\ell)}$). The operations of the Attention Operator and the Message Fusion Operator are partially overlapping.

Update Operator $\mathfrak{U}^{(\ell)}$ combines the aggregated representations from multiple attention heads, typically through concatenation (for intermediate layers) or averaging (for the output layer), followed by application of a non-linear activation function σ to produce the new node representations:

$$\mathfrak{U}^{(\ell)}\left(\{{\boldsymbol{A}_a^{(\ell)}}^k \cdot (\boldsymbol{H}^{(\ell-1)}{\boldsymbol{W}^{(\ell)}}^k)\}_{k=1}^K, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_u^{(\ell)}\right) = \mathop{\Big\|}_{k=1}^{K} \sigma\left({\boldsymbol{A}_a^{(\ell)}}^k \cdot (\boldsymbol{H}^{(\ell-1)}{\boldsymbol{W}^{(\ell)}}^k)\right), \tag{29}$$

where $\|$ represents concatenation, K represents the number of attention heads and no parameters are used in this operator ($\boldsymbol{\Theta}_u^{(\ell)} = \varnothing$).

The resulting layer formulation is:

$$\boldsymbol{H}^{(\ell)} = \mathop{\Big\|}_{k=1}^{K} \sigma\left({\boldsymbol{A}_a^{(\ell)}}^k \cdot (\boldsymbol{H}^{(\ell-1)}{\boldsymbol{W}^{(\ell)}}^k)\right). \tag{30}$$

GIN (Graph Isomorphism Network) (Xu et al., 2019) uses a fixed, uniform attention weight for neighbors and a powerful update function based on an MLP to achieve expressiveness equivalent to the Weisfeiler-Lehman graph isomorphism test.

Attention Operator $\mathfrak{A}^{(\ell)}$ employs a static attention weight of 1 for all neighbors and a weight of $(1+\epsilon^{(\ell)})$ for the central node itself:

$$\mathfrak{A}^{(\ell)}\left(\boldsymbol{A}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_a^{(\ell)}\right) = \tilde{\boldsymbol{A}}_\epsilon = \boldsymbol{A} + (1+\epsilon^{(\ell)})\boldsymbol{I} = \tilde{\boldsymbol{A}}_\epsilon, \tag{31}$$

where $\boldsymbol{\Theta}_a^{(\ell)} = \epsilon^{(\ell)}$ is a potentially learnable parameter.

Message Fusion Operator $\mathfrak{M}^{(\ell)}$ sums the neighbor messages and the scaled central node's message:

$$\mathfrak{M}^{(\ell)}\left(\tilde{\boldsymbol{A}}_\epsilon, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_m^{(\ell)}\right) = \tilde{\boldsymbol{A}}_\epsilon \boldsymbol{H}^{(\ell-1)}, \tag{32}$$

where no parameters are used ($\boldsymbol{\Theta}_m^{(\ell)} = \varnothing$).

Update Operator $\mathfrak{U}^{(\ell)}$ applies a multi-layer perceptron ($\mathrm{MLP}^{(\ell)}$) to the fused message:

$$\mathfrak{U}^{(\ell)}\left(\tilde{\boldsymbol{A}}_\epsilon \boldsymbol{H}^{(\ell-1)}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_u^{(\ell)}\right) = \mathrm{MLP}^{(\ell)}\left(\tilde{\boldsymbol{A}}_\epsilon \boldsymbol{H}^{(\ell-1)}\right), \tag{33}$$

where $\boldsymbol{\Theta}_u^{(\ell)}$ are the parameters of the MLP.

The resulting layer formulation is:

$$\boldsymbol{H}^{(\ell)} = \mathrm{MLP}^{(\ell)}\left(\left(\boldsymbol{A} + (1+\epsilon^{(\ell)})\boldsymbol{I}\right)\boldsymbol{H}^{(\ell-1)}\right). \tag{34}$$

GT (Graph Transformer) (Ying et al., 2021) enhances the standard transformer architecture to incorporate structural information of graphs, often by augmenting the self-attention mechanism with structural biases.

Attention Operator $\mathfrak{A}^{(\ell)}$ computes the query, key matrices $\boldsymbol{Q}^{(\ell)}, \boldsymbol{K}^{(\ell)} \in \mathbb{R}^{d_{\ell-1}\times d_\ell}$ via linear projections, with the core attention weight $\hat{\boldsymbol{A}}^{(\ell)}$ formulated as a sum of standard semantic attention and a structural attention component $\boldsymbol{B}^{(\ell)}$ (e.g., from positional encodings, edge features, connectivity patterns or node degrees).

$$\boldsymbol{Q}^{(\ell)} = \boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_Q^{(\ell)}, \quad \boldsymbol{K}^{(\ell)} = \boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_K^{(\ell)}, \tag{35}$$

$$\hat{\boldsymbol{A}}^{(\ell)} = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}^{(\ell)}(\boldsymbol{K}^{(\ell)})^T}{\sqrt{d_\ell}} + \boldsymbol{B}^{(\ell)}\right), \tag{36}$$

$$\mathfrak{A}^{(\ell)}\left(\boldsymbol{A}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_a^{(\ell)}\right) = \hat{\boldsymbol{A}}^{(\ell)}, \tag{37}$$

where $\boldsymbol{\Theta}_a^{(\ell)}$ including the projection weights for $\boldsymbol{Q}^{(\ell)}, \boldsymbol{K}^{(\ell)}$ and parameters for computing $\boldsymbol{B}^{(\ell)}$.

Message Fusion Operator $\mathfrak{M}^{(\ell)}$ computes the value matrice $\boldsymbol{V}^{(\ell)} \in \mathbb{R}^{d_{\ell-1}\times d_\ell}$ via linear projection $\boldsymbol{W}_V^{(\ell)}$ and performs the weighted aggregation of the value vectors using the computed attention matrix $\hat{\boldsymbol{A}}^{(\ell)}$:

$$\mathfrak{M}^{(\ell)}\left(\hat{\boldsymbol{A}}^{(\ell)}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_m^{(\ell)}\right) = \hat{\boldsymbol{A}}^{(\ell)}\boldsymbol{V}^{(\ell)} = \hat{\boldsymbol{A}}^{(\ell)} \cdot (\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_V^{(\ell)}), \tag{38}$$

where $\boldsymbol{\Theta}_m^{(\ell)} = \boldsymbol{W}_V^{(\ell)}$.

Update Operator $\mathfrak{U}^{(\ell)}$ applies a residual connection, layer normalization (LN), a position-wise feed-forward network (FFN), another residual connection, and layer normalization.

$$\tilde{\boldsymbol{H}}^{(\ell)} = \mathrm{LN}\left(\boldsymbol{H}^{(\ell-1)} + \hat{\boldsymbol{A}}^{(\ell)} \cdot (\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_V^{(\ell)})\right), \tag{39}$$

$$\mathrm{FFN}^{(l)}(\tilde{\boldsymbol{H}}^{(\ell)}) = \sigma\left(\tilde{\boldsymbol{H}}^{(\ell)}\boldsymbol{W}_1^{(l)} + \boldsymbol{b}_1^{(l)}\right)\boldsymbol{W}_2^{(l)} + \boldsymbol{b}_2^{(l)} \tag{40}$$

$$\mathfrak{U}^{(\ell)}\left(\hat{\boldsymbol{A}}^{(\ell)}\boldsymbol{V}^{(\ell)}, \boldsymbol{H}^{(\ell-1)}; \boldsymbol{\Theta}_u^{(\ell)}\right) = \mathrm{LN}\left(\tilde{\boldsymbol{H}}^{(\ell)} + \mathrm{FFN}^{(\ell)}(\tilde{\boldsymbol{H}}^{(\ell)})\right), \tag{41}$$

where $\boldsymbol{\Theta}_u^{(\ell)}$ are the parameters of the $\mathrm{FFN}^{(\ell)}$. Multi-head self-attention (MHA) can also correspond to the Update Operator.

The resulting layer formulation is:

$$\boldsymbol{H}^{(\ell)} = \mathrm{LN}\left(\mathrm{LN}\left(\boldsymbol{H}^{(\ell-1)} + \hat{\boldsymbol{A}}^{(\ell)}\boldsymbol{V}^{(\ell)}\right) + \mathrm{FFN}^{(\ell)}\left(\mathrm{LN}\left(\boldsymbol{H}^{(\ell-1)} + \hat{\boldsymbol{A}}^{(\ell)}\boldsymbol{V}^{(\ell)}\right)\right)\right). \tag{42}$$

This analysis demonstrates that the proposed unified GFM layer provides a powerful and expressive framework that generalizes a broad spectrum of prevalent GNN architectures. The specific choices of the operators $\mathfrak{A}^{(\ell)}$, $\mathfrak{M}^{(\ell)}$, and $\mathfrak{U}^{(\ell)}$ determine the particular inductive biases and capabilities of the resulting model.

B.3 Details of Definition 2

Definition 8 (Polyhedral Region). *In the context of Euclidean spaces, a polyhedral region (or polyhedron) is a subset of $\mathbb{R}^n$ defined by a finite set of linear inequalities. Formally, a set $R \subseteq \mathbb{R}^n$ is a polyhedral region if there exist matrices $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and vectors $\boldsymbol{b} \in \mathbb{R}^m$ such that:*

$$R = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}, \tag{43}$$

where the inequality is applied component-wise.

Remark 1. *A polyhedral region may be described as the intersection of finitely many closed half-spaces and/or hyperplanes, making it a convex polytope (possibly unbounded). In many analytical contexts, polyhedral regions are assumed to be non-empty and may be required to have a non-empty interior to avoid degenerate cases.*

Definition 9 (Polyhedral Region in Matrix Space). *A set $R \subseteq \mathbb{R}^{N\times d}$ is a polyhedral region if there exists a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times Nd}$ and a vector $\boldsymbol{b} \in \mathbb{R}^m$ such that:*

$$R = \left\{\boldsymbol{H} \in \mathbb{R}^{N\times d} \mid \boldsymbol{A} \cdot vec(\boldsymbol{H}) \leq \boldsymbol{b}\right\}, \tag{44}$$

where $vec(\boldsymbol{H}) \in \mathbb{R}^{Nd}$ denotes the vectorization of the matrix $\boldsymbol{H}$ (i.e., the column vector obtained by stacking the columns of $\boldsymbol{H}$). The inequality $\leq$ is applied component-wise.

Remark 2. *Since the spaces $\mathbb{R}^{\alpha\times\beta}$ and $\mathbb{R}^{\alpha\beta}$ are isomorphic as vector spaces via the vectorization operation $vec : \mathbb{R}^{\alpha\times\beta} \to \mathbb{R}^{\alpha\beta}$ (which stacks the columns of a matrix into a vector) and its inverse $unvec : \mathbb{R}^{\alpha\beta} \to \mathbb{R}^{\alpha\times\beta}$, many theorems and proofs in this paper do not strictly distinguish between the matrix form and the vectorized form. This isomorphism allows us to apply concepts from Euclidean geometry and measure theory directly to matrix-valued functions by considering their vectorized counterparts, without loss of generality. Consequently, in the following analysis, we may interchangeably use matrix or vector representations as convenient, ensuring that all results hold equivalently in both forms.*

Definition 10 (Piecewise Linear Function). *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called piecewise linear if there exists a finite set of polyhedral regions $\{R_i\}_{i=1}^K$ such that $\mathbb{R}^n = \bigcup_{i=1}^K R_i$ and f is affine on each R_i, i.e., $f(\boldsymbol{x}) = \boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i$ for all $\boldsymbol{x} \in R_i$, where $\boldsymbol{A}_i \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b}_i \in \mathbb{R}^m$.*

Definition 11 (Jacobian of a Matrix Map). *For a function $F : \mathbb{R}^{N\times d_{in}} \to \mathbb{R}^{N\times d_{out}}$ that is differentiable at a point $\boldsymbol{H}$, the Jacobian of F at $\boldsymbol{H}$ is defined as the Jacobian matrix of the vectorized function. Specifically, let $f : \mathbb{R}^{Nd_{in}} \to \mathbb{R}^{Nd_{out}}$ be given by $f(\boldsymbol{h}) = vec(F(unvec(\boldsymbol{h})))$. Then, the Jacobian matrix $\boldsymbol{J}_F(\boldsymbol{H}) \in \mathbb{R}^{Nd_{out}\times Nd_{in}}$ is:*

$$\boldsymbol{J}_F(\boldsymbol{H}) = \left.\frac{\partial f}{\partial \boldsymbol{h}}\right|_{\boldsymbol{h}=vec(\boldsymbol{H})}. \tag{45}$$

This matrix contains all first-order partial derivatives of the vectorized output with respect to the vectorized input.

Lemma 1 (Composition of Piecewise Linear Functions). *If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ are piecewise linear functions, then the composition $g \circ f : \mathbb{R}^n \to \mathbb{R}^p$ is also piecewise linear.*

Proof. Since f is piecewise linear, there exists a partition of $\mathbb{R}^n$ into polyhedral regions R_i such that f is affine on each R_i. Similarly, g is piecewise linear with polyhedral regions S_j in $\mathbb{R}^m$ where g is affine. For each i and j, consider the set $R_i \cap f^{-1}(S_j)$. Since f is affine on R_i, $f(R_i)$ is a polyhedral set, and $f^{-1}(S_j) \cap R_i$ is polyhedral (as the intersection of polyhedral sets). On $R_i \cap f^{-1}(S_j)$, $g \circ f$ is affine because f is affine and g is affine on S_j. The collection of all such sets $R_i \cap f^{-1}(S_j)$ covers $\mathbb{R}^n$, and there are finitely many such sets. Thus, $g \circ f$ is piecewise linear. □

B.4 Proof of Proposition 1

Proof. Based on the derivations in Appendix B.1, we observe that most operators primarily involve matrix multiplication or can be approximated by matrix multiplications. Some operators further apply a piecewise linear activation function (e.g., ReLU or LeakyReLU) or an MLP based on ReLU after the matrix multiplication. As a result, these operators are generally piecewise linear functions. Their piecewise linearity stems directly from the piecewise linear activation functions used. Since the activation functions are continuous, the continuity of these operators is obvious. Piecewise linear functions are differentiable a.e. because they are differentiable in the interior of each polyhedral region (where they are affine) and non-differentiable only on the boundaries, which have Lebesgue measure zero. □

Remark 3. *Linear functions are considered a special case of piecewise linear functions. If operator $\mathfrak{A}^{(\ell)}$ computes scaled dot-product attention, it somewhat exceeds the scope of our theoretical framework. Alternatively, approximating the computation of dynamic attention using piecewise linear mappings may, from a mathematical limit perspective, exhibit certain compatibility with the theoretical framework presented in this paper. This constitutes a promising direction for future research aimed at extending the current theory.*

B.5 Proof of Proposition 2

Proof. The layer map $F^{(\ell)}$ is defined by the composition of the operators $\mathfrak{A}^{(\ell)}, \mathfrak{M}^{(\ell)}, \mathfrak{U}^{(\ell)}$, as given in Definition 1. This can be viewed as a function $F^{(\ell)}$ that maps $\boldsymbol{H}^{(\ell-1)}$ to $\boldsymbol{H}^{(\ell)}$. Since each operator is piecewise linear, and by Lemma 1, the composition of piecewise linear functions is itself piecewise linear. Therefore, $F^{(\ell)}$ is a piecewise linear function. More formally, let $f_1 = \mathfrak{A}^{(\ell)}$, $f_2 = \mathfrak{M}^{(\ell)}$, and $f_3 = \mathfrak{U}^{(\ell)}$. Then $F^{(\ell)} = f_3 \circ (f_2 \circ (f_1, \mathrm{id}), \mathrm{id})$, where id denotes the identity function (which is linear and thus piecewise linear). The composition involves piecewise linear functions and Cartesian products (which preserve piecewise linearity), so $F^{(\ell)}$ is piecewise linear.

By Definition 2, there exists a finite set of polyhedral regions $\{R_i\}_{i=1}^K$ such that $\mathbb{R}^{N \times d_{\ell-1}} = \bigcup_{i=1}^K R_i$ and $F^{(\ell)}$ is affine on each R_i, i.e., $F^{(\ell)}(\boldsymbol{H}) = \mathrm{unvec}(\boldsymbol{A}_i \cdot \mathrm{vec}(\boldsymbol{H}) + \boldsymbol{b}_i)$ for all $\boldsymbol{H} \in R_i$, where $\boldsymbol{A}_i \in \mathbb{R}^{Nd_\ell \times Nd_{\ell-1}}$ and $\boldsymbol{b}_i \in \mathbb{R}^{Nd_\ell}$. An affine function is differentiable everywhere in the interior of its region. The polyhedral regions R_i are closed and have boundaries that are sets of measure zero (since they are defined by finite sets of linear inequalities). Therefore, $F^{(\ell)}$ is differentiable almost everywhere (a.e.)—specifically, in the interior of each region R_i. At any point $\boldsymbol{H}$ where $F^{(\ell)}$ is differentiable (i.e., in the interior of some R_i), the derivative is given by the constant matrix $\boldsymbol{A}_i$. The Jacobian matrix $\boldsymbol{J}^{(\ell)}(\boldsymbol{H})$ is precisely this matrix $\boldsymbol{A}_i$, which exists and has dimensions $\mathbb{R}^{Nd_\ell \times Nd_{\ell-1}}$ (since the input space has dimension $Nd_{\ell-1}$ and the output space has dimension Nd_ℓ). Hence, for any point $\boldsymbol{H}$ where $F^{(\ell)}$ is differentiable, the Jacobian $\boldsymbol{J}^{(\ell)}(\boldsymbol{H})$ exists. □

B.6 Details of Definition 3

Definition 12 (Compact Smooth Manifold). *A set $\mathcal{M} \subset \mathbb{R}^n$ is called a compact smooth manifold of dimension D_0 if it satisfies the following two conditions:*

1. *(Smooth Structure) For every point $p \in \mathcal{M}$, there exists an open neighborhood $U \subset \mathbb{R}^n$ containing p and a smooth (C^∞) mapping $F : U \to \mathbb{R}^{n-D_0}$ such that: $U \cap \mathcal{M} = F^{-1}(0) = \{x \in U \mid F(x) = 0\}$ and the Jacobian matrix $DF(x) \in \mathbb{R}^{(n-D_0)\times n}$ has full rank $(n - D_0)$ for all $x \in U \cap \mathcal{M}$.*

2. *(Compactness) $\mathcal{M}$ is compact in the subspace topology induced from $\mathbb{R}^n$, which by the Heine-Borel theorem is equivalent to being closed and bounded in $\mathbb{R}^n$.*

Definition 13 (Intrinsic Dimension). *The intrinsic dimension $D_0 = d_{int}(\mathcal{M}_0)$ of a manifold $\mathcal{M}_0$ is the minimum number of parameters needed to locally parameterize the manifold. Formally, it is the dimension of the tangent space $T_p\mathcal{M}_0$ at any point $p \in \mathcal{M}_0$, which is constant for smooth connected manifolds.*

Remark 4. *For a more basic definition of manifold, please refer to introductory mathematics textbook (Lee, 2011). In our subsequent discussion of prismatic space, we generalize the concept of intrinsic* *dimension. Since prismatic space lacks the well-behaved mathematical properties of smooth manifold, we define the intrinsic dimension as the maximum of the dimensions at all locally smooth points of the space.*

B.7 Proof of Proposition 3

Proof. We proceed by leveraging the definitions provided and establishing the piecewise linearity of the composite map $\Phi^{(\ell)}$, then analyzing its image on the input manifold $\mathcal{M}_0$.

By Proposition 2, each layer map $F^{(\ell)} : \mathbb{H}^{(\ell-1)} \to \mathbb{H}^{(\ell)}$ is a piecewise linear function. This follows from the assumptions that the operators $\mathfrak{A}^{(\ell)}$, $\mathfrak{M}^{(\ell)}$, and $\mathfrak{U}^{(\ell)}$ are piecewise linear and almost everywhere differentiable, and that $\mathfrak{U}^{(\ell)}$ uses piecewise linear activations. Since the composition of piecewise linear functions is piecewise linear (Lemma 1), the composite map $\Phi^{(\ell)} = F^{(\ell)} \circ \cdots \circ F^{(1)}$ is also piecewise linear. Formally, there exists a finite set of polyhedral regions $\{R_i\}_{i=1}^{K}$ covering the domain of $\Phi^{(\ell)}$ such that for each i, the restriction of $\Phi^{(\ell)}$ to R_i is affine:

$$\Phi^{(\ell)}(\boldsymbol{H}) = \text{unvec}(\boldsymbol{A}_i \cdot \text{vec}(\boldsymbol{H}) + \boldsymbol{b}_i) \quad \text{for all } \boldsymbol{H} \in R_i, \tag{46}$$

where $\boldsymbol{A}_i \in \mathbb{R}^{Nd_\ell \times Nd_0}$ and $\boldsymbol{b}_i \in \mathbb{R}^{Nd_\ell}$ are constants specific to region R_i.

The input manifold $\mathcal{M}_0 \subset \mathbb{R}^{N\times d_0}$ is compact and smooth by Definition 3. Consider the intersection of $\mathcal{M}_0$ with the polyhedral regions R_i:

$$\mathcal{M}_0^{(i)} = \mathcal{M}_0 \cap R_i. \tag{47}$$

Since $\mathcal{M}_0$ is a smooth manifold and each R_i is polyhedral, the sets $\mathcal{M}_0^{(i)}$ are submanifolds with boundaries (possibly with corners). The collection $\{\mathcal{M}_0^{(i)}\}_{i=1}^{K}$ forms a finite cover of $\mathcal{M}_0$.

On each $\mathcal{M}_0^{(i)}$, the map $\Phi^{(\ell)}$ is affine. Therefore, the image $\Phi^{(\ell)}(\mathcal{M}_0^{(i)})$ is an affine transformation of $\mathcal{M}_0^{(i)}$:

$$\Phi^{(\ell)}(\mathcal{M}_0^{(i)}) = \{\text{unvec}(\boldsymbol{A}_i \cdot \text{vec}(\boldsymbol{H}) + \boldsymbol{b}_i) \mid \boldsymbol{H} \in \mathcal{M}_0^{(i)}\}. \tag{48}$$

Assuming that $\Phi^{(\ell)}$ is injective on each R_i, it is also injective on each $\mathcal{M}_0^{(i)}$. Since affine maps preserve linear structures and injectivity ensures that the map is an embedding on each piece, $\Phi^{(\ell)}(\mathcal{M}_0^{(i)})$ is itself a submanifold with boundary (possibly with corners) in $\mathbb{R}^{N\times d_\ell}$.

The full representation space is the union of these images:

$$\mathcal{M}^{(\ell)} = \bigcup_{i=1}^{K} \Phi^{(\ell)}(\mathcal{M}_0^{(i)}). \tag{49}$$

Such a union is termed a prismatic space.

Singularities occur at the boundaries between the regions. Specifically:

- The boundaries between different $\mathcal{M}_0^{(i)}$ correspond to points where $\Phi^{(\ell)}$ transitions from one affine piece to another.
- At these boundaries, the Jacobian of $\Phi^{(\ell)}$ may be discontinuous or undefined, leading to non-smooth points in $\mathcal{M}^{(\ell)}$.
- Since $\mathcal{M}_0$ is compact and smooth, it generically intersects multiple regions R_i, making such singularities typical. For example, if $\mathcal{M}_0$ is transversal to the boundaries of R_i, the intersections will be lower-dimensional manifolds where the image under $\Phi^{(\ell)}$ may not be smooth.

Thus, $\mathcal{M}^{(\ell)}$ is a prismatic space and may have singularities along the boundaries of the pieces $\Phi^{(\ell)}(\mathcal{M}_0^{(i)})$. □

Remark 5. *The prismatic space we define constitutes a geometric structure more complex than a conventional topological manifold. While its interior may largely exhibit the properties of a smooth manifold, its boundary can contain intricate corners or even singularities. As a result, it is highly unlikely that the prismatic space satisfies the standard definitions of a topological manifold. It should be emphasized that constructing a rigorous topological definition of this geometric structure is highly challenging. Therefore, within the framework of this paper, we adopt a simplified definition grounded in piecewise linear map.*

B.8 Details of Definition 5

Remark 6. *The prismatic effect of different singular values on space:*

- *A singular value $\sigma_i^{(\ell)} \approx 1$ represents an unrefracted dimension, typically corresponding to node features preserved through linear identity paths or attention mechanisms that remain active.*
- *A singular value $0 < \sigma_i^{(\ell)} < 1$ represents a contracted dimension, potentially arising from the scaling of weight matrices ($\|\boldsymbol{W}^{(\ell)}\| < 1$) and the gradient attenuation of activation functions like ReLU/LeakyReLU in their unsaturated regimes.*
- *A singular value $\sigma_i^{(\ell)} = 0$ represents a nullified dimension, resulting directly from the sparsity induced by ReLU activations which reduces the rank of the layer's Jacobian.*
- *A singular value $\sigma_i^{(\ell)} > 1$ represents an expanded dimension, potentially arising from feature amplification in weight matrices ($\|\boldsymbol{W}^{(\ell)}\| > 1$) or certain graph convolution operations.*

B.9 Proof of Theorem 1

Proof. Since $F^{(\ell)}$ is linear on $\mathbb{S}$, there exists a matrix $\boldsymbol{A}^{(\ell)} \in \mathbb{R}^{Nd_\ell \times Nd_{\ell-1}}$ and a vector $\boldsymbol{b}^{(\ell)}$ such that for all $\boldsymbol{X} \in \mathbb{S}$:

$$F^{(\ell)}(\boldsymbol{X}) = \text{unvec}(\boldsymbol{A}^{(\ell)} \cdot \text{vec}(\boldsymbol{X}) + \boldsymbol{b}^{(\ell)}). \tag{50}$$

The Jacobian $\boldsymbol{J}^{(\ell)}$ is constant and equal to $\boldsymbol{A}^{(\ell)}$. By assumption, $\boldsymbol{A}^{(\ell)}$ has rank r_ℓ, and its singular value decomposition is:

$$\boldsymbol{A}^{(\ell)} = \boldsymbol{U}^{(\ell)} \boldsymbol{\Sigma}^{(\ell)} \boldsymbol{V}^{(\ell)\top}, \tag{51}$$

where $\boldsymbol{U}^{(\ell)}$ and $\boldsymbol{V}^{(\ell)}$ are orthogonal matrices, and $\boldsymbol{\Sigma}^{(\ell)} = \text{diag}(\sigma_1^{(\ell)}, \ldots, \sigma_{r_\ell}^{(\ell)}, 0, \ldots, 0)$ with $\sigma_1^{(\ell)} \geq \sigma_2^{(\ell)} \geq \cdots \geq \sigma_{r_\ell}^{(\ell)} > 0$.

Let $\boldsymbol{V}_s^{(\ell)}$ be the first s columns of $\boldsymbol{V}^{(\ell)}$, spanning the subspace $\mathbb{V}^{(\ell)}$. The restriction of $\boldsymbol{A}^{(\ell)}$ to $\mathbb{V}^{(\ell)}$ is the linear map $L^{(\ell)} : \mathbb{V}^{(\ell)} \to \mathbb{R}^M$ defined by $L^{(\ell)}(\boldsymbol{x}) = \boldsymbol{A}^{(\ell)}\boldsymbol{x}$.

Since $\boldsymbol{A}^{(\ell)}$ is injective on $\mathbb{V}^{(\ell)}$ (as $\mathbb{V}^{(\ell)}$ is spanned by right singular vectors corresponding to positive singular values), $L^{(\ell)}$ is injective. The image $L^{(\ell)}(\mathbb{V}^{(\ell)})$ is an s-dimensional subspace of $\mathbb{R}^M$, spanned by the first s columns of $\boldsymbol{U}^{(\ell)}$.

Let $\{\boldsymbol{v}_1^{(\ell)}, \ldots, \boldsymbol{v}_s^{(\ell)}\}$ be an orthonormal basis for $\mathbb{V}^{(\ell)}$ (e.g., the columns of $\boldsymbol{V}_s^{(\ell)}$). Then $\{L^{(\ell)}(\boldsymbol{v}_1^{(\ell)}), \ldots, L^{(\ell)}(\boldsymbol{v}_s^{(\ell)})\}$ is a basis for $L^{(\ell)}(\mathbb{V}^{(\ell)})$, and:

$$L^{(\ell)}(\boldsymbol{v}_i^{(\ell)}) = \sigma_i^{(\ell)} \boldsymbol{u}_i^{(\ell)}, \tag{52}$$

where $\boldsymbol{u}_i^{(\ell)}$ is the i-th column of $\boldsymbol{U}^{(\ell)}$. Thus, $\{\boldsymbol{u}_1^{(\ell)}, \ldots, \boldsymbol{u}_s^{(\ell)}\}$ is an orthonormal basis for $L^{(\ell)}(\mathbb{V}^{(\ell)})$.

The s-dimensional Hausdorff measure $\mathcal{H}^s$ is equivalent to the s-dimensional Lebesgue measure on s-dimensional subspaces. Consider the linear map $L^{(\ell)} : \mathbb{V}^{(\ell)} \to L^{(\ell)}(\mathbb{V}^{(\ell)})$. Since $\mathbb{V}^{(\ell)}$ and $L^{(\ell)}(\mathbb{V}^{(\ell)})$ are s-dimensional Euclidean spaces, we can compute the change in measure using the determinant of $L^{(\ell)}$ (in orthonormal coordinates).

Let $\boldsymbol{x} \in \mathbb{V}^{(\ell)}$ have coordinates $\boldsymbol{x} = \sum_{i=1}^{s} x_i \boldsymbol{v}_i^{(\ell)}$. Then:

$$L^{(\ell)}(\boldsymbol{x}) = \sum_{i=1}^{s} x_i L^{(\ell)}(\boldsymbol{v}_i) = \sum_{i=1}^{s} x_i \sigma_i^{(\ell)} \boldsymbol{u}_i^{(\ell)}. \tag{53}$$

Thus, the matrix representation of $L^{(\ell)}$ with respect to the bases $\boldsymbol{v}_i^{(\ell)}$ and $\boldsymbol{u}_i^{(\ell)}$ is the diagonal matrix $\operatorname{diag}(\sigma_1^{(\ell)}, \dots, \sigma_s^{(\ell)})$.

The absolute determinant of this matrix is $\prod_{i=1}^{s} \sigma_i^{(\ell)}$. Therefore, for any measurable set $\mathbb{S} \subset \mathbb{V}^{(\ell)}$:

$$\mathcal{H}^s(L^{(\ell)}(\mathbb{S})) = \left(\prod_{i=1}^{s} \sigma_i^{(\ell)} \right) \mathcal{H}^s(\mathbb{S}). \tag{54}$$

Since $F^{(\ell)}(\boldsymbol{X}) = L^{(\ell)}(\boldsymbol{X}) + \boldsymbol{b}^{(\ell)}$ and translation preserves Hausdorff measure, we have:

$$\mathcal{H}^s(F^{(\ell)}(\mathbb{S})) = \mathcal{H}^s(L^{(\ell)}(\mathbb{S}) + \boldsymbol{b}^{(\ell)}) = \mathcal{H}^s(L^{(\ell)}(\mathbb{S})) = \left(\prod_{i=1}^{s} \sigma_i^{(\ell)} \right) \mathcal{H}^s(\mathbb{S}). \tag{55}$$

When $s = r_\ell$, $\mathbb{V}^{(\ell)}$ is the entire row space of $\boldsymbol{A}^{(\ell)}$, and the product is over all positive singular values. This gives the volume contraction factor for the full rank part of the map. □

Remark 7. *We will not elaborate on mathematical concepts such as Hausdorff measure and Lebesgue measure in this article. For details, please refer to mathematics textbook (*Krantz & Parks*, 2008).*

B.10 Simple Linear Algebra

Lemma 2 (The Rank Inequality for Composition of Linear Maps)**.** *Let $A : \mathbb{V} \to \mathbb{W}$ and $B : \mathbb{W} \to \mathbb{U}$ be linear maps between vector spaces. The composition $B \circ A : \mathbb{V} \to \mathbb{U}$ is also a linear map. The rank of a linear map is defined as the dimension of its image:*

$$\textit{rank}(A) = \dim(\textit{im}(A)), \quad \textit{rank}(B) = \dim(\textit{im}(B)), \quad \textit{rank}(B \circ A) = \dim(\textit{im}(B \circ A)). \tag{56}$$

Then:

$$\textit{rank}(B \circ A) \le \min(\textit{rank}(A), \textit{rank}(B)). \tag{57}$$

Proof. Prove the first inequality: $\operatorname{rank}(B \circ A) \le \operatorname{rank}(A)$.

Observe that for any $\boldsymbol{v} \in \mathbb{V}$,

$$(B \circ A)(\boldsymbol{v}) = B(A(\boldsymbol{v})), \tag{58}$$

so the image of $B \circ A$ is:

$$\operatorname{im}(B \circ A) = \{B(A(\boldsymbol{v})) : \boldsymbol{v} \in \mathbb{V}\} = B(\{A(\boldsymbol{v}) : \boldsymbol{v} \in \mathbb{V}\}) = B(\operatorname{im}(A)). \tag{59}$$

Thus, $\operatorname{im}(B \circ A) = B(\operatorname{im}(A))$. Since $\operatorname{im}(A) \subseteq \mathbb{W}$, we can restrict B to $\operatorname{im}(A)$, obtaining a linear map:

$$B|_{\operatorname{im}(A)} : \operatorname{im}(A) \to \mathbb{U}. \tag{60}$$

The image of this restricted map is exactly $B(\operatorname{im}(A)) = \operatorname{im}(B \circ A)$. By the Rank-Nullity Theorem (or simply by the fact that the image of a linear map cannot exceed the dimension of its domain), we have:

$$\dim(B(\operatorname{im}(A))) \le \dim(\operatorname{im}(A)). \tag{61}$$

Therefore,

$$\operatorname{rank}(B \circ A) = \dim(\operatorname{im}(B \circ A)) \le \dim(\operatorname{im}(A)) = \operatorname{rank}(A). \tag{62}$$

Prove the second inequality: $\operatorname{rank}(B \circ A) \le \operatorname{rank}(B)$.

We now show that $\operatorname{im}(B \circ A) \subseteq \operatorname{im}(B)$. Let $\boldsymbol{u} \in \operatorname{im}(B \circ A)$. Then there exists $\boldsymbol{v} \in \mathbb{V}$ such that:

$$\boldsymbol{u} = (B \circ A)(\boldsymbol{v}) = B(A(\boldsymbol{v})). \tag{63}$$

Since $A(\boldsymbol{v}) \in \mathbb{W}$, it follows that $\boldsymbol{u} = B(\boldsymbol{w})$ for some $\boldsymbol{w} \in \mathbb{W}$, so $\boldsymbol{u} \in \operatorname{im}(B)$. Hence,

$$\operatorname{im}(B \circ A) \subseteq \operatorname{im}(B), \tag{64}$$

and therefore:

$$\dim(\operatorname{im}(B \circ A)) \le \dim(\operatorname{im}(B)) \quad \Rightarrow \quad \operatorname{rank}(B \circ A) \le \operatorname{rank}(B). \tag{65}$$

Combining both inequalities (62) and (65), we conclude:

$$\operatorname{rank}(B \circ A) \le \min(\operatorname{rank}(A), \operatorname{rank}(B)). \tag{66}$$

□

B.11 Proof of Theorem 2

Proof. From Proposition 2, each layer map $F^{(\ell)}$ is piecewise linear and differentiable almost everywhere. By Proposition 3, the composite map $\Phi = F^{(L)} \circ \cdots \circ F^{(1)}$ is also piecewise linear and $\mathcal{M}^{(L)}$ is a prismatic space. On each linear region C_k (as defined in Definition 6), Φ is linear, so its rank is constant on C_k. Thus, Φ is piecewise constant on its rank.

Let $\{C_k\}$ be the linear region partition of $\mathcal{M}_0$ from Definition 6. For each C_k, the map $\Phi|_{C_k}$ is linear. Let $T_k = \Phi|_{C_k}$ denote this linear map. The image $\Phi(C_k)$ is contained in a linear subspace of dimension $\text{rank}(T_k)$.

The local dimension of $\mathcal{M}^{(L)}$ at any point in $\Phi(C_k)$ is at most $\text{rank}(T_k)$. Since $\mathcal{M}^{(L)} = \bigcup_k \Phi(C_k)$, the intrinsic dimension $d_{\text{int}}(\mathcal{M}^{(L)})$ is the supremum of the local dimensions over all points in $\mathcal{M}^{(L)}$. Thus,

$$d_{\text{int}}(\mathcal{M}^{(L)}) \leq \max_k \text{rank}(T_k). \tag{67}$$

Now, we bound $\text{rank}(T_k)$. Since $T_k = F^{(L)} \circ \cdots \circ F^{(1)}|_{C_k}$, and each $F^{(\ell)}$ is linear on the relevant region, we have:

$$\text{rank}(T_k) \leq \min_\ell \text{rank}\left(F^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}\right). \tag{68}$$

This follows from Lemma 2: for linear maps A and B, $\text{rank}(B \circ A) \leq \min(\text{rank}(A), \text{rank}(B))$. By induction, this holds for the composition of L linear maps.

For each layer ℓ, rank $\left(F^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}\right)$ = rank $\left(\boldsymbol{J}^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}\right)$ because the Jacobian is constant on the region where $F^{(\ell)}$ is linear (from Definition 6).

Let $r_{\ell,k} = \text{rank}\left(\boldsymbol{J}^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}\right)$. Then,

$$\text{rank}(T_k) \leq \min_\ell r_{\ell,k}. \tag{69}$$

Therefore,

$$d_{\text{int}}(\mathcal{M}^{(L)}) \leq \max_k \text{rank}(T_k) \leq \max_k \min_\ell r_{\ell,k}. \tag{70}$$

Due to the contraction effect of the layers (especially with ReLUs, which project dimensions to zero), the ranks $r_{\ell,k}$ are often much smaller than the input dimension D_0. Thus, $\max_k \min_\ell r_{\ell,k}$ is typically less than D_0, implying that $\mathcal{M}^{(L)}$ has a lower intrinsic dimension than D_0. □

B.12 Proof of Theorem 3

Proof. By Proposition 2, each layer map $F^{(\ell)}$ is piecewise linear and differentiable almost everywhere. By Proposition 3, the composite map $\Phi = F^{(L)} \circ \cdots \circ F^{(1)}$ is piecewise linear. By Definition 6, the input manifold $\mathcal{M}_0$ is partitioned into cells C_k such that on each C_k, Φ is linear.

We assume Φ is injective on C_k. This implies that for each C_k, Φ restricted to C_k is a linear injection, so $d_{\text{int}}(\Phi(C_k)) = d_{\text{int}}(C_k) = d_{\text{int}}$, where $d_{\text{int}} = D_0$ is the intrinsic dimension of $\mathcal{M}_0$ (Definition 3). Since Φ is injective, each layer $F^{(\ell)}$ must be injective on $\Phi^{(\ell-1)}(C_k)$ for all ℓ and k. Otherwise, the composition would not be injective. Thus, for each ℓ and k, the Jacobian $\boldsymbol{J}^{(\ell)}$ of $F^{(\ell)}$ restricted to the tangent space of $\Phi^{(\ell-1)}(C_k)$ has rank at least d_{int}. Since the tangent space is d_{int}-dimensional, $\boldsymbol{J}^{(\ell)}$ has exactly d_{int} positive singular values $\sigma^{(\ell)}_{1,k} \geq \sigma^{(\ell)}_{2,k} \geq \cdots \geq \sigma^{(\ell)}_{d_{\text{int}},k} > 0$ on the region corresponding to C_k.

Consider a fixed cell C_k. Since Φ is linear on C_k, we can write $\Phi(\boldsymbol{X}) = \text{unvec}(\boldsymbol{J}_k \cdot \text{vec}(\boldsymbol{X}) + \boldsymbol{b}_k)$ for $\boldsymbol{X} \in C_k$, where $\boldsymbol{J}_k$ is the Jacobian of Φ on C_k (constant). However, to understand the layer-wise measure change, we use the composition structure.

For the first layer $F^{(1)}$, since it is linear on C_k, it maps C_k to $F^{(1)}(C_k)$. By Theorem 1, the d_{int}-dimensional Hausdorff measure changes as:

$$\mathcal{H}^{d_{\text{int}}}(F^{(1)}(C_k)) = \Big(\prod_{i=1}^{d_{\text{int}}} \sigma^{(1)}_{i,k}\Big)\mathcal{H}^{d_{\text{int}}}(C_k), \tag{71}$$

where $\sigma_{i,k}^{(1)}$ are the singular values of $\boldsymbol{J}^{(1)}$ restricted to the tangent space of C_k (which is d_{int}-dimensional).

For the second layer $F^{(2)}$, it is linear on $F^{(1)}(C_k)$ (which is d_{int}-dimensional). It maps $F^{(1)}(C_k)$ to $F^{(2)}(F^{(1)}(C_k))$. Again, by Theorem 1:

$$\mathcal{H}^{d_{\text{int}}}(F^{(2)}(F^{(1)}(C_k))) = \Big(\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(2)}\Big)\mathcal{H}^{d_{\text{int}}}(F^{(1)}(C_k)) = \Big(\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(2)}\Big)\Big(\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(1)}\Big)\mathcal{H}^{d_{\text{int}}}(C_k), \quad (72)$$

where $\sigma_{i,k}^{(2)}$ are the singular values of $\boldsymbol{J}^{(2)}$ restricted to the tangent space of $F^{(1)}(C_k)$.

Proceeding inductively for all L layers, we get:

$$\mathcal{H}^{d_{\text{int}}}(\Phi(C_k)) = \Big(\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(C_k). \quad (73)$$

This is because each layer's measure change factor is multiplicative, and the composition preserves the d_{int}-dimensional measure up to the product of the singular values.

Since the cells C_k form a partition of $\mathcal{M}_0$ (Definition 6), and Φ is injective on C_k, the images $\Phi(C_k)$ are disjoint and cover $\mathcal{M}^{(L)}$ (up to sets of measure zero, due to piecewise linearity). Therefore, by the additivity of the Hausdorff measure:

$$\mathcal{H}^{d_{\text{int}}}(\mathcal{M}^{(L)}) = \sum_k \mathcal{H}^{d_{\text{int}}}(\Phi(C_k)) = \sum_k \Big(\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(C_k). \quad (74)$$

This establishes the desired formula.

If Φ is not injective, then the images $\Phi(C_k)$ may overlap. Since the Hausdorff measure is subadditive, we have:

$$\mathcal{H}^{d_{\text{int}}}(\mathcal{M}^{(L)}) \leq \sum_k \mathcal{H}^{d_{\text{int}}}(\Phi(C_k)) = \sum_k \Big(\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(C_k). \quad (75)$$

Thus, the formula provides an upper bound. □

B.13 Details of Definition 7

Remark 8. *This definition formalizes the notion of how a prompt $\boldsymbol{P}$ modifies the input data manifold in the context of prompt tuning. The original input manifold $\mathcal{M}_0$, which represents the natural data distribution (e.g., graph node features), is typically assumed to be a compact smooth manifold embedded in $\mathbb{R}^{N\times d_0}$. The prompt $\boldsymbol{P}$ is a low-dimensional perturbation applied to every point in $\mathcal{M}_0$, resulting in a new manifold $\mathcal{M}_0(\boldsymbol{P})$. The operation $\mathcal{M}_0(\boldsymbol{P}) = \{\boldsymbol{X} + \boldsymbol{P} \mid \boldsymbol{X} \in \mathcal{M}_0\}$ is a translation of the entire manifold by $\boldsymbol{P}$, which preserves the topological and geometric properties of $\mathcal{M}_0$, such as compactness and smoothness, since translation is a diffeomorphism. The prompt space $\mathcal{P}$ is the set of all possible prompts, often constrained to be low-dimensional (e.g., a subspace of $\mathbb{R}^{N\times d_0}$), and each prompt $\boldsymbol{P} \in \mathcal{P}$ defines a distinct perturbed manifold. This family of manifolds $\{\mathcal{M}_0(\boldsymbol{P}) \mid \boldsymbol{P} \in \mathcal{P}\}$ encapsulates the variability introduced by prompt tuning, and the goal is to understand how the graph foundation model (GFM) transforms these manifolds through its layers.*

B.14 Lipschitz Continuous and Jacobian

Lemma 3 (Continuity of the Layer Map $F^{(\ell)}$)**.** *Assume the operators $\mathfrak{A}^{(\ell)}$, $\mathfrak{M}^{(\ell)}$, and $\mathfrak{U}^{(\ell)}$ defining the GFM layer in Definition 1 are continuous. Then, the layer map $F^{(\ell)}$ is continuous.*

Proof. Similar to the proof of Proposition 2, let $f_1 = \mathfrak{A}^{(\ell)}$, $f_2 = \mathfrak{M}^{(\ell)}$, and $f_3 = \mathfrak{U}^{(\ell)}$. Then $F^{(\ell)} = f_3 \circ (f_2 \circ (f_1, \text{id}), \text{id})$, where id denotes the identity function Since the composition of continuous functions is continuous, the overall layer map $F^{(\ell)}$ is continuous. □

Lemma 4 (Lipschitz Continuity of the GFM Map Φ)**.** *Let $\mathcal{M}_0(\boldsymbol{P}) \subset \mathbb{R}^{N\times d_0}$ be the compact prompt-perturbed input manifold as defined in Definition 7. The composite map $\Phi = F^{(L)} \circ F^{(L-1)} \circ \cdots \circ F^{(1)}$, where each $F^{(\ell)}$ is a piecewise linear layer map (Proposition 2), is Lipschitz continuous on $\mathcal{M}_0(\boldsymbol{P})$. That is, there exists a constant $L_\Phi < \infty$ such that for all $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{M}_0(\boldsymbol{P})$,*

$$\|\Phi(\boldsymbol{X}) - \Phi(\boldsymbol{Y})\| \leq L_\Phi \|\boldsymbol{X} - \boldsymbol{Y}\|. \tag{76}$$

Moreover, the Lipschitz constant L_Φ satisfies:

$$L_\Phi \leq \prod_{\ell=1}^{L} L_\ell, \tag{77}$$

where L_ℓ is the Lipschitz constant of the ℓ-th layer $F^{(\ell)}$ on the appropriate domain.

Proof. Prove the piecewise linear layers are Lipschitz continuous.

Each layer map $F^{(\ell)} : \mathbb{R}^{N\times d_{\ell-1}} \to \mathbb{R}^{N\times d_\ell}$ is piecewise linear and continuous by Proposition 2 and Lemma 3. Since $\mathcal{M}_0(\boldsymbol{P})$ is compact and each $F^{(\ell)}$ is continuous, the image $F^{(\ell)}(\mathcal{M}_0(\boldsymbol{P}))$ is also compact. The piecewise linearity implies that there exists a finite partition of the domain into polyhedral regions $R_k^{(\ell)}$ such that $F^{(\ell)}$ is linear on each region $R_k^{(\ell)}$. On each such region, for any $\boldsymbol{H}, \boldsymbol{H}' \in R_k^{(\ell)}$, we have

$$\|F^{(\ell)}(\boldsymbol{H}) - F^{(\ell)}(\boldsymbol{H}')\| = \|\boldsymbol{A}_k^{(\ell)}(\boldsymbol{H} - \boldsymbol{H}')\| \leq \|\boldsymbol{A}_k^{(\ell)}\|_{\text{op}} \|\boldsymbol{H} - \boldsymbol{H}'\|, \tag{78}$$

where $\boldsymbol{A}_k^{(\ell)}$ is the matrix representing the linear map on $R_k^{(\ell)}$ and $|\cdot|_{\text{op}}$ denotes the operator norm (spectral norm). Define the local Lipschitz constant for $F^{(\ell)}$ on region $R_k^{(\ell)}$ as $L_k^{(\ell)} = |\boldsymbol{A}_k^{(\ell)}|_{\text{op}}$. Since the number of regions intersecting the compact set $\mathcal{M}_0(\boldsymbol{P})$ is finite, the global Lipschitz constant for $F^{(\ell)}$ on $\mathcal{M}_0(\boldsymbol{P})$ is finite and given by

$$L_\ell = \max_k L_k^{(\ell)} < \infty. \tag{79}$$

Thus, for any $\boldsymbol{H}, \boldsymbol{H}' \in \mathcal{M}_0(\boldsymbol{P})$,

$$\|F^{(\ell)}(\boldsymbol{H}) - F^{(\ell)}(\boldsymbol{H}')\| \leq L_\ell \|\boldsymbol{H} - \boldsymbol{H}'\|. \tag{80}$$

Prove the composite map Φ is Lipschitz continuous.

The composite map $\Phi = F^{(L)} \circ F^{(L-1)} \circ \cdots \circ F^{(1)}$ is a composition of Lipschitz continuous maps. For any $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{M}_0(\boldsymbol{P})$, let $\boldsymbol{H}^{(\ell)} = F^{(\ell)} \circ \cdots \circ F^{(1)}(\boldsymbol{X})$ and $\boldsymbol{K}^{(\ell)} = F^{(\ell)} \circ \cdots \circ F^{(1)}(\boldsymbol{Y})$ denote the intermediate representations. Then,

$$\begin{aligned}
\|\boldsymbol{H}^{(1)} - \boldsymbol{K}^{(1)}\| &= \|F^{(1)}(\boldsymbol{X}) - F^{(1)}(\boldsymbol{Y})\| \leq L_1 \|\boldsymbol{X} - \boldsymbol{Y}\|, \\
\|\boldsymbol{H}^{(2)} - \boldsymbol{K}^{(2)}\| &= \|F^{(2)}(\boldsymbol{H}^{(1)}) - F^{(2)}(\boldsymbol{K}^{(1)})\| \leq L_2 \|\boldsymbol{H}^{(1)} - \boldsymbol{K}^{(1)}\| \leq L_2 L_1 \|\boldsymbol{X} - \boldsymbol{Y}\|, \\
&\vdots \\
\|\boldsymbol{H}^{(L)} - \boldsymbol{K}^{(L)}\| &= \|\Phi(\boldsymbol{X}) - \Phi(\boldsymbol{Y})\| \leq L_L \|\boldsymbol{H}^{(L-1)} - \boldsymbol{K}^{(L-1)}\| \leq \Big(\prod_{\ell=1}^{L} L_\ell\Big) \|\boldsymbol{X} - \boldsymbol{Y}\|.
\end{aligned} \tag{81}$$

Therefore, Φ is Lipschitz continuous with constant $L_\Phi = \prod_{\ell=1}^{L} L_\ell$.

This completes the proof. □

Lemma 5 (Existence of the Jacobian $\boldsymbol{J}_\Phi(\boldsymbol{X})$)**.** *In the context of the unified GFM framework, we aim to prove that the Jacobian of the composite map $\Phi = F^{(L)} \circ F^{(L-1)} \circ \cdots \circ F^{(1)}$ exists almost everywhere (a.e.) on the input manifold $\mathcal{M}_0(\boldsymbol{P})$, and that at points where it exists, it is given by the product of the layer Jacobians.*

Proof. By Proposition 2, each layer map $F^{(\ell)} : \mathbb{R}^{N\times d_{\ell-1}} \to \mathbb{R}^{N\times d_\ell}$ is piecewise linear. This means that the domain of $F^{(\ell)}$ can be partitioned into a finite number of polyhedral regions $R_k^{(\ell)}$ such that $F^{(\ell)}$ is linear on each region. Since linear functions are differentiable everywhere, $F^{(\ell)}$ is differentiable on the interior of each region. The boundaries between regions have Lebesgue measure zero in $\mathbb{R}^{N\times d_{\ell-1}}$ (as they are subsets of lower-dimensional affine spaces). Therefore, $F^{(\ell)}$ is differentiable almost everywhere in its domain. Let D_ℓ denote the set of points where $F^{(\ell)}$ is differentiable; then D_ℓ has full measure (i.e., its complement has measure zero).

The composite map Φ is defined as $\Phi = F^{(L)} \circ F^{(L-1)} \circ \cdots \circ F^{(1)}$. Consider the sets where each $F^{(\ell)}$ is differentiable. Since each $F^{(\ell)}$ is differentiable a.e., the set of points where all $F^{(\ell)}$ are differentiable along the composition path is also of full measure. More formally, define:

- $E_1 = D_1$ (the set where $F^{(1)}$ is differentiable).
- For $\ell = 2$ to L, define $E_\ell = \{\boldsymbol{X} \in E_{\ell-1} : F^{(\ell)}\}$ is differentiable at $\Phi^{(\ell-1)}(\boldsymbol{X})$, where $\Phi^{(\ell-1)} = F^{(\ell-1)} \circ \cdots \circ F^{(1)}$.

Since $F^{(\ell)}$ is differentiable a.e., and $\Phi^{(\ell-1)}$ is continuous and piecewise linear (hence Lipschitz), it preserves sets of measure zero. Thus, by induction, each E_ℓ has full measure. Therefore, the set $E = E_L$ where all $F^{(\ell)}$ are differentiable at the appropriate points has full measure in $\mathcal{M}_0(\boldsymbol{P})$. For any $\boldsymbol{X} \in E$, the composite map Φ is differentiable at $\boldsymbol{X}$ by the chain rule.

At a point $\boldsymbol{X} \in E$, the chain rule applies. Let $\boldsymbol{H}^{(0)} = \boldsymbol{X}$, and for $\ell = 1$ to L, define $\boldsymbol{H}^{(\ell)} = F^{(\ell)}(\boldsymbol{H}^{(\ell-1)})$. Then, the Jacobian of Φ at $\boldsymbol{X}$ is given by:

$$\boldsymbol{J}_\Phi(\boldsymbol{X}) = \boldsymbol{J}^{(L)}(\boldsymbol{H}^{(L-1)}) \cdot \boldsymbol{J}^{(L-1)}(\boldsymbol{H}^{(L-2)}) \cdots \boldsymbol{J}^{(1)}(\boldsymbol{X}), \tag{82}$$

where $\boldsymbol{J}^{(\ell)}(\boldsymbol{H}^{(\ell-1)})$ is the Jacobian of $F^{(\ell)}$ at $\boldsymbol{H}^{(\ell-1)}$. This product is well-defined because each Jacobian exists at the respective points.

Since $\mathcal{M}_0(\boldsymbol{P})$ is a compact smooth manifold embedded in $\mathbb{R}^{N\times d_0}$, it has a Lipschitz parameterization. The above argument holds for almost every point in $\mathcal{M}_0(\boldsymbol{P})$ with respect to the Lebesgue measure on the parameter space. Thus, $\boldsymbol{J}_\Phi(\boldsymbol{X})$ exists for almost every $\boldsymbol{X} \in \mathcal{M}_0(\boldsymbol{P})$. □

B.15 Proof of Theorem 4

Proof. Assume the input manifold $\mathcal{M}_0(\boldsymbol{P})$ is compact and smooth with intrinsic dimension d_{int}. By Definition 6, the piecewise linear map $\Phi = F^{(L)} \circ \cdots \circ F^{(1)}$ partitions $\mathcal{M}_0(\boldsymbol{P})$ into a countable collection of cells $\{C'_k\}$, where each C'_k is a connected subset of $\mathcal{M}_0(\boldsymbol{P})$ such that Φ is linear on C'_k. This partition exists because Φ is piecewise linear (Theorem 2).

Proof of the Measure Bound.

For each cell C'_k, since Φ is linear on C'_k, the Jacobian $\boldsymbol{J}_\Phi$ is constant on C'_k. By Theorem 3, the Hausdorff measure of the image $\Phi(C'_k)$ is given by:

$$\mathcal{H}^{d_{\text{int}}}(\Phi(C'_k)) = \Big(\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(C'_k), \tag{83}$$

where $\sigma_{i,k}^{(\ell)}$ are the first d_{int} singular values of the Jacobian of the ℓ-th layer evaluated in the linear region corresponding to C'_k. Note that the product $\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}$ is taken over the largest d_{int} singular values, as the tangent space has dimension d_{int}.

The total measure of $\mathcal{M}^{(L)}(\boldsymbol{P})$ is the sum over all cells:

$$\mathcal{H}^{d_{\text{int}}}(\mathcal{M}^{(L)}(\boldsymbol{P})) \leq \sum_k \mathcal{H}^{d_{\text{int}}}(\Phi(C'_k)) = \sum_k \Big(\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(C'_k). \tag{84}$$

Since $\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}}\sigma_{i,k}^{(\ell)} \leq \sup_{k'}\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}}\sigma_{i,k'}^{(\ell)}$ for all k, we have:

$$\mathcal{H}^{d_{\text{int}}}(\mathcal{M}^{(L)}(\boldsymbol{P})) \leq \Big(\sup_{k'}\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}}\sigma_{i,k'}^{(\ell)}\Big)\sum_{k}\mathcal{H}^{d_{\text{int}}}(C'_k) = \Big(\sup_{k'}\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}}\sigma_{i,k'}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(\mathcal{M}_0(\boldsymbol{P})). \quad (85)$$

This proves the measure bound.

Proof of the Diameter Bound.

Let diam$(\mathcal{M})$ denote the diameter of a set $\mathcal{M}$, defined as:

$$\text{diam}(\mathcal{M}) = \sup_{x,y\in\mathcal{M}} \|x-y\|. \quad (86)$$

By Theorem 2 and Lemma 4, the map Φ is piecewise linear and Lipschitz continuous on $\mathcal{M}_0(\boldsymbol{P})$. The global Lipschitz constant L_Φ satisfies:

$$\|\Phi(\boldsymbol{X}) - \Phi(\boldsymbol{Y})\| \leq L_\Phi\|\boldsymbol{X}-\boldsymbol{Y}\| \quad \forall \boldsymbol{X},\boldsymbol{Y}\in\mathcal{M}_0(\boldsymbol{P}). \quad (87)$$

The Lipschitz constant L_Φ can be bounded by the operator norms of the Jacobians of Φ. For any point $\boldsymbol{X}\in\mathcal{M}_0(\boldsymbol{P})$, the Jacobian $\boldsymbol{J}_\Phi(\boldsymbol{X})$ exists almost everywhere (by Definition 5) and is given by the product of the layer Jacobians:

$$\boldsymbol{J}_\Phi(\boldsymbol{X}) = \boldsymbol{J}^{(L)}(F^{(L-1)}(\boldsymbol{X}))\cdots\boldsymbol{J}^{(1)}(\boldsymbol{X}). \quad (88)$$

The operator norm of $\boldsymbol{J}_\Phi(\boldsymbol{X})$ satisfies:

$$|\boldsymbol{J}_\Phi(\boldsymbol{X})|_{\text{op}} \leq |\boldsymbol{J}^{(L)}(F^{(L-1)}(\boldsymbol{X}))|_{\text{op}}\cdots|\boldsymbol{J}^{(1)}(\boldsymbol{X})|_{\text{op}}. \quad (89)$$

Each layer Jacobian $|\boldsymbol{J}^{(\ell)}(\boldsymbol{X}_\ell)|_{\text{op}}$ (where $\boldsymbol{X}_\ell = F^{(\ell-1)}(\boldsymbol{X})$) is constant on linear regions. Let $|\boldsymbol{J}_k^{(\ell)}|_{\text{op}}$ be the operator norm of the Jacobian of the ℓ-th layer in the k-th linear region. Then:

$$|\boldsymbol{J}^{(\ell)}(\boldsymbol{X}_\ell)|_{\text{op}} \leq \sup_k |\boldsymbol{J}_k^{(\ell)}|_{\text{op}} \quad \forall \boldsymbol{X}_\ell. \quad (90)$$

Therefore,

$$|\boldsymbol{J}_\Phi(\boldsymbol{X})|_{\text{op}} \leq \prod_{\ell=1}^{L}\sup_k |\boldsymbol{J}_k^{(\ell)}|_{\text{op}} \quad \forall \boldsymbol{X}. \quad (91)$$

The global Lipschitz constant L_Φ is the supremum of $|\boldsymbol{J}_\Phi(\boldsymbol{X})|_{\text{op}}$ over $\boldsymbol{X}\in\mathcal{M}_0(\boldsymbol{P})$:

$$L_\Phi = \sup_{\boldsymbol{X}\in\mathcal{M}_0(\boldsymbol{P})}|\boldsymbol{J}_\Phi(\boldsymbol{X})|_{\text{op}} \leq \prod_{\ell=1}^{L}\sup_k |\boldsymbol{J}_k^{(\ell)}|_{\text{op}}. \quad (92)$$

Now, for any $\boldsymbol{X},\boldsymbol{Y}\in\mathcal{M}_0(\boldsymbol{P})$,

$$\|\Phi(\boldsymbol{X}) - \Phi(\boldsymbol{Y})\| \leq L_\Phi\|\boldsymbol{X}-\boldsymbol{Y}\| \leq \Big(\prod_{\ell=1}^{L}\sup_k |\boldsymbol{J}_k^{(\ell)}|_{\text{op}}\Big)\|\boldsymbol{X}-\boldsymbol{Y}\|. \quad (93)$$

Taking the supremum over $\boldsymbol{X},\boldsymbol{Y}\in\mathcal{M}_0(\boldsymbol{P})$, we get:

$$\text{diam}(\mathcal{M}^{(L)}(\boldsymbol{P})) \leq \Big(\prod_{\ell=1}^{L}\sup_k |\boldsymbol{J}_k^{(\ell)}|_{\text{op}}\Big)\cdot\text{diam}(\mathcal{M}_0(\boldsymbol{P})). \quad (94)$$

This proves the diameter bound. □

B.16 Theoretical Limitations of Prompt Tuning

The Prompt Efficacy Bound (Theorem 4) reveals fundamental theoretical limitations of prompt tuning in GFMs. Specifically, the measure and diameter bounds imply that the influence of a prompt $\boldsymbol{P}$ is constrained by the compositional prismatic effect of the frozen GFM layers.

Information Loss through Spectral Contraction.

The measure bound shows that the effective "volume" of the prompt-perturbed space $\mathcal{M}^{(L)}(\boldsymbol{P})$ is scaled by the product of singular values across layers and linear regions. Since deep GFMs often exhibit spectral decay (with many singular values $\sigma_i^{(\ell)} \ll 1$), the prompt-induced perturbations are compressed exponentially with depth. This irreversible contraction implies that fine-grained semantic nuances introduced by the prompt may be lost or distorted before reaching the output layer.

Intrinsic Dimensionality Collapse.

As shown in Theorem 2, the intrinsic dimension $d_{\text{int}}(\mathcal{M}^{(L)})$ of the final representation is bounded by the minimal rank achieved locally across layers. Prompt tuning operates on the input manifold $\mathcal{M}_0(\boldsymbol{P})$, but the frozen network's piecewise linear transformations inherently project the prompt into a lower-dimensional subspace. Thus, even if the prompt is high-dimensional, its effective influence is limited by the bottleneck rank of the Jacobians, reducing its capacity to encode complex instructions.

Sensitivity to Input Geometry.

The diameter bound depends on the operator norms of the layer Jacobians. If the network exhibits gradient explosion (large $\sup_k |\boldsymbol{J}_k^{(\ell)}|_{\text{op}}$) or vanishing (small singular values), the prompt's effect may be either amplified erratically or suppressed. This sensitivity makes prompt tuning highly dependent on the pre-trained model's architecture and parameterization, limiting its robustness.

Non-Adaptive Prismatic Structure.

Since the network is frozen, the prompt cannot alter the prismatic folding process (e.g., the partition into linear regions or the Jacobian spectra). The prompt is merely a shift in the input space, and its efficacy depends on how the fixed geometric transformation Φ distorts this shift. In contrast, full fine-tuning adapts Φ itself to preserve task-relevant information, which prompt tuning cannot achieve.

Trade-off Between Prompt Size and Expressivity.

While increasing the prompt dimension $\dim(\mathcal{P})$ might seem beneficial, the measure bound shows that the effective output scale is constrained by the product of Jacobian singular values. Thus, simply enlarging the prompt may not improve efficacy if the network's contraction forces are too strong. This suggests a fundamental trade-off between prompt complexity and the network's capacity to preserve prompt-induced variations.

In summary, prompt tuning is inherently limited by the frozen GFM's spectral properties and geometric structure. While it can induce some distributional shifts, its ability to convey nuanced instructions is bounded by the network's pre-existing prismatic contraction and rank collapse. These limitations motivate the need for architectural interventions (e.g., adding adapters) or alternative tuning strategies that can mitigate the loss of prompt information through deeper layers.

C Theoretical Analysis of Message Tuning

C.1 Proof of Theorem 5

Proof. We prove the theorem using the geometric measure theoretic framework of Prismatic Space Theory. The key idea is that message tuning, by injecting learnable parameters at each layer, can compensate for the measure contraction and intrinsic dimension reduction caused by the prismatic effect of the frozen GFM layers, and can additionally expand the diameter of the output space.

Intrinsic Dimension Comparison.

In Prismatic Space Theory, the intrinsic dimension $d_{\text{int}}(\mathcal{M}_{\text{MTG}}^{(L)})$ refers to the topological dimension or Hausdorff dimension of the final representation space $\mathcal{M}_{\text{MTG}}^{(L)}$, which is the inherent dimensionality of the space itself, not the dimension of the embedding space. This intrinsic dimension is defined by the geometric properties of space, but we can use the rank of the Jacobian matrix of the mapping to provide an upper bound.

Specifically, for the mapping $\Phi_{\text{MTG}} : \mathcal{M}_0 \to \mathcal{M}_{\text{MTG}}^{(L)}$ (where Φ_{MTG} is the composite layer mapping after message tuning), we have:

$$d_{\text{int}}(\mathcal{M}_{\text{MTG}}^{(L)}) \leq \sup_{\boldsymbol{X} \in \mathcal{M}_0} \text{rank}(\boldsymbol{J}_{\Phi_{\text{MTG}}}(\mathbf{X})), \tag{95}$$

where $\boldsymbol{J}_{\Phi_{\text{MTG}}}(\boldsymbol{X})$ is the Jacobian matrix of the mapping Φ_{MTG} at point $\boldsymbol{X}$. This means that the intrinsic dimension of the space cannot exceed the maximum rank of the Jacobian matrix across all input points.

From Theorem 2, for prompt tuning, the intrinsic dimension of the final space is bounded by:

$$d_{\text{int}}(\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P})) \leq \max_k \min_\ell \text{rank}(\boldsymbol{J}^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}), \tag{96}$$

where $\boldsymbol{J}^{(\ell)}$ is the Jacobian of the ℓ-th layer of the frozen GFM, and C_k are the linear regions of the input manifold.

For message tuning, the layer map is modified to include the fusion operation $\mathfrak{F}^{(\ell)}$. Specifically, at each layer ℓ, the input representation $\boldsymbol{H}^{(\ell-1)}$ is transformed to $\boldsymbol{H}_{\boldsymbol{M}}^{(\ell-1)} = \mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{M}^{(\ell)}; \boldsymbol{\Theta}_f^{(\ell)})$ before applying the standard layer map $F^{(\ell)}$. Thus, the effective layer map becomes $\Psi^{(\ell)} = F^{(\ell)} \circ \mathfrak{F}^{(\ell)}$.

The Jacobian of $\Psi^{(\ell)}$ at a point where it is differentiable is given by the chain rule:

$$\boldsymbol{J}_\Psi^{(\ell)} = \boldsymbol{J}_F^{(\ell)} \cdot \boldsymbol{J}_{\mathfrak{F}}^{(\ell)}, \tag{97}$$

where $\boldsymbol{J}_F^{(\ell)}$ is the Jacobian of $F^{(\ell)}$ and $\boldsymbol{J}_{\mathfrak{F}}^{(\ell)}$ is the Jacobian of $\mathfrak{F}^{(\ell)}$.

The core issue is that $\mathfrak{F}^{(\ell)}$ is not linear, but we can show that with learnable parameters, its Jacobian can be made full-rank, ensuring the desired rank inequality.

Recall that for message tuning, the fusion operation is defined as:

$$\mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{M}^{(\ell)}; \boldsymbol{\Theta}_f^{(\ell)}) = \boldsymbol{H}^{(\ell-1)} + \text{Softmax}(\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_p^{(\ell)}) \cdot \boldsymbol{M}^{(\ell)}, \tag{98}$$

where $\boldsymbol{H}^{(\ell-1)} \in \mathbb{R}^{N \times d_{\ell-1}}$, $\boldsymbol{W}_p^{(\ell)} \in \mathbb{R}^{d_{\ell-1} \times m}$, and $\boldsymbol{M}^{(\ell)} \in \mathbb{R}^{m \times d_{\ell-1}}$.

The Jacobian of $\mathfrak{F}^{(\ell)}$ with respect to $\boldsymbol{H}^{(\ell-1)}$ is a block-diagonal matrix composed of N blocks, each of size $d_{\ell-1} \times d_{\ell-1}$. For each node i, the block corresponds to the derivative of the i-th row of $\mathfrak{F}^{(\ell)}$ with respect to the i-th row of $\boldsymbol{H}^{(\ell-1)}$. Specifically, let $\boldsymbol{h}_i$ be the i-th row of $\boldsymbol{H}^{(\ell-1)}$, and let $\boldsymbol{a}_i = \boldsymbol{h}_i\boldsymbol{W}_p^{(\ell)}$. Then the Softmax output is $\alpha_i = \text{Softmax}(\boldsymbol{a}_i)$, and the i-th row of $\mathfrak{F}^{(\ell)}$ is $\boldsymbol{h}_i + \alpha_i\boldsymbol{M}^{(\ell)}$.

The Jacobian for node i is:

$$\boldsymbol{B}_i = \boldsymbol{I} + \boldsymbol{M}^{(\ell)\top}\boldsymbol{J}_{\text{softmax}}(\boldsymbol{a}_i)\boldsymbol{W}_p^{(\ell)\top}, \tag{99}$$

where $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{J}_{\text{softmax}}(\boldsymbol{a}_i) \in \mathbb{R}^{m \times m}$ is the Jacobian of Softmax at $\boldsymbol{a}_i$, which has rank $m - 1$.

Since $\boldsymbol{J}_{\text{softmax}}(\boldsymbol{a}_i)$ is bounded, we can choose $\boldsymbol{M}^{(\ell)}$ and $\boldsymbol{W}_p^{(\ell)}$ such that the spectral norm of $\boldsymbol{M}^{(\ell)\top}\boldsymbol{J}_{\text{softmax}}(\boldsymbol{a}_i)\boldsymbol{W}_p^{(\ell)\top}$ is less than 1 for all i. This ensures that $\boldsymbol{B}_i$ is invertible and thus full-rank for all i. Therefore, the full Jacobian $\boldsymbol{J}_{\mathfrak{F}}^{(\ell)}$ has rank $Nd_{\ell-1}$.

Now, for the composite map $\Psi^{(\ell)} = F^{(\ell)} \circ \mathfrak{F}^{(\ell)}$, the Jacobian is:

$$\boldsymbol{J}_\Psi^{(\ell)} = \boldsymbol{J}_F^{(\ell)} \cdot \boldsymbol{J}_{\mathfrak{F}}^{(\ell)}. \tag{100}$$

Since $\boldsymbol{J}_{\mathfrak{F}}^{(\ell)}$ has full rank $Nd_{\ell-1}$, and $\boldsymbol{J}_F^{(\ell)}$ has rank r, we have Sylvester's rank inequality:

$$\text{rank}(\boldsymbol{J}_\Psi^{(\ell)}) \geq \text{rank}(\boldsymbol{J}_F^{(\ell)}) + \text{rank}(\boldsymbol{J}_{\mathfrak{F}}^{(\ell)}) - Nd_{\ell-1} = \text{rank}(\boldsymbol{J}_F^{(\ell)}) + Nd_{\ell-1} - Nd_{\ell-1} = \text{rank}(\boldsymbol{J}_F^{(\ell)}). \tag{101}$$

Thus, the rank of $\boldsymbol{J}_\Psi^{(\ell)}$ is at least the rank of $\boldsymbol{J}_F^{(\ell)}$:

$$\text{rank}(\boldsymbol{J}_\Psi^{(\ell)}) \geq \text{rank}(\boldsymbol{J}_F^{(\ell)}). \tag{102}$$

Moreover, by optimizing the fusion parameters, we can ensure that $\text{rank}(\boldsymbol{J}_\Psi^{(\ell)}) \geq \text{rank}(\boldsymbol{J}_F^{(\ell)})$ for all ℓ. Since $\Psi^{(\ell)}$ does not introduce additional linear region partitions, meaning it does not generate more boundaries, corners, or singular points, the inequality holds pointwise. For any k, there always exists a point $\boldsymbol{X}_k$ such that:

$$\min_{\ell} \operatorname{rank}(\boldsymbol{J}_{\Psi}^{(\ell)}|_{\boldsymbol{X}_k}) \geq \min_{\ell} \operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}). \tag{103}$$

This implies that the upper bound on the intrinsic dimension for message tuning is at least as large as that for prompt tuning:

$$\max_{k} \min_{\ell} \operatorname{rank}(\boldsymbol{J}_{\Psi}^{(\ell)}|_{\boldsymbol{X}_k}) \geq \max_{k} \min_{\ell} \operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}). \tag{104}$$

Although $\boldsymbol{J}_{\mathfrak{F}}^{(\ell)}$ is full-rank and thus $\operatorname{rank}(\boldsymbol{J}_{\Psi}^{(\ell)}) = \operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)})$ at any point where both are defined, the key to strict inequality lies in the distribution of points across linear regions of the frozen layers. The message fusion operation $\mathfrak{F}^{(\ell)}$ can map inputs to different linear regions of $F^{(\ell)}$ where the rank of $\boldsymbol{J}_{F}^{(\ell)}$ is higher.

Suppose that for some layer ℓ, the frozen Jacobian $\boldsymbol{J}_{F}^{(\ell)}$ has varying rank across its linear regions. Specifically, there exist linear regions R_{low} and R_{high} such that:

$$\operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)}|_{R_{\text{low}}}) < \operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)}|_{R_{\text{high}}}). \tag{105}$$

In prompt tuning, the input to $F^{(\ell)}$ may fall primarily into R_{low} due to the shift caused by the prompt, resulting in a lower minimum rank. However, in message tuning, the learnable parameters $\boldsymbol{\Theta}_f^{(\ell)}$ and $\boldsymbol{M}^{(\ell)}$ can be optimized to steer the input to $F^{(\ell)}$ into R_{high}, thereby increasing the rank at that layer.

Assume that $\Phi^{(\ell-1)}(C_k)$ does not lie in a linear region that maximizes $\operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)})$. This assumption is realistic because $\Phi^{(\ell-1)}$ is pre-trained and lacks the ability to adjust its output range. Formally, by optimizing the fusion parameters, we can ensure that for each layer ℓ, the input $\mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)})$ lies in a region where $\operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)})$ is maximized. Consequently, for any k, there always exists a point $\boldsymbol{Y}_k$ such that:

$$\min_{\ell} \operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)}|_{\boldsymbol{Y}_k \in \mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)})}) > \min_{\ell} \operatorname{rank}(\boldsymbol{J}_{F}^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}). \tag{106}$$

This implies that the upper bound for message tuning is strictly greater:

$$\max_{k} \min_{\ell} \operatorname{rank}(\boldsymbol{J}_{\Psi}^{(\ell)}|_{\boldsymbol{Y}_k}) > \max_{k} \min_{\ell} \operatorname{rank}(\boldsymbol{J}^{(\ell)}|_{\Phi^{(\ell-1)}(C_k)}). \tag{107}$$

Therefore, the actual intrinsic dimension satisfies:

$$d_{\text{int}}(\mathcal{M}_{\text{MTG}}^{(L)}) > d_{\text{int}}(\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P})). \tag{108}$$

This strict inequality holds when the fusion parameters are optimized to avoid low-rank linear regions of the frozen layers, which is achievable through gradient-based training that maximizes the rank of the Jacobians during adaptation.

Thus, message tuning provides strictly greater adaptation capacity in terms of intrinsic dimension compared to prompt tuning.

$$d_{\text{int}}(\mathcal{M}_{\text{MTG}}^{(L)}) \geq d_{\text{int}}(\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P})) \tag{109}$$

and the inequality is strict for some configuration.

Hausdorff Measure Comparison.

Recall that the pre-trained GFM Φ is composed of L layers, each defined as in Definition 1. For prompt tuning, the input manifold is perturbed by a prompt $\boldsymbol{P}$, resulting in $\mathcal{M}_0(\boldsymbol{P})$. The final space is $\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P}) = \Phi(\mathcal{M}_0(\boldsymbol{P}))$.

For message tuning, we introduce learnable message prototypes $\boldsymbol{M}^{(\ell)} \in \mathbb{R}^{m \times d_{\ell-1}}$ and fusion parameters $\boldsymbol{\Theta}_f^{(\ell)}$ at each layer ℓ, modifying the layer map to:

$$\boldsymbol{H}^{(\ell)} = \mathfrak{U}^{(\ell)}\left(\mathfrak{M}^{(\ell)}\left(\mathfrak{A}^{(\ell)}\left(\boldsymbol{A}, \boldsymbol{H}_{\boldsymbol{M}}^{(\ell-1)}; \boldsymbol{\Theta}_a^{(\ell)}\right), \boldsymbol{H}_{\boldsymbol{M}}^{(\ell-1)}; \boldsymbol{\Theta}_m^{(\ell)}\right), \boldsymbol{H}_{\boldsymbol{M}}^{(\ell-1)}; \boldsymbol{\Theta}_u^{(\ell)}\right), \tag{110}$$

where

$$\boldsymbol{H}_{\boldsymbol{M}}^{(\ell-1)} = \mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{M}^{(\ell)}; \boldsymbol{\Theta}_f^{(\ell)}) = \boldsymbol{H}^{(\ell-1)} + \text{Softmax}(\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_p^{(\ell)}) \cdot \boldsymbol{M}^{(\ell)}. \tag{111}$$

The modified network is denoted Φ_{MTG}, and the final space is $\mathcal{M}_{\text{MTG}}^{(L)} = \Phi_{\text{MTG}}(\mathcal{M}_0)$.

The introduction of the Softmax function in the fusion operation $\mathfrak{F}^{(\ell)}$ indeed breaks the strict piecewise linearity of the layer map, since Softmax is a smooth, nonlinear function. However, we can address this issue through analyzing the network as a piecewise-linear map with smooth activations, leveraging the fact that the Softmax can be effectively constant on large regions of the input space.

More generally, we can partition the input space into regions where the Softmax is approximately linear. For instance, if we use a linearized Softmax (e.g., by taking a first-order Taylor expansion around a point), we obtain a piecewise linear approximation. The error of this approximation can be made arbitrarily small by refining the partition.

Given the above, we may treat Φ_{MTG} as a piecewise linear map for the purpose of geometric analysis. Specifically, we define:

$$\mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{M}^{(\ell)}; \boldsymbol{W}_p^{(\ell)}) \approx \boldsymbol{H}^{(\ell-1)} + \text{Linear}(\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_p^{(\ell)})\boldsymbol{M}^{(\ell)}, \tag{112}$$

where Linear$(\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_p^{(\ell)})$ is a piecewise linear function (e.g., sparsemax (Martins & Astudillo, 2016) or a linearized Softmax). Then, the modified layer map is piecewise linear, and the entire network Φ_{MTG} is piecewise linear.

Under this approximation, by Theorem 3, the Hausdorff measures are:

$$\mathcal{H}^{d_{\text{int}}}(\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P})) = \sum_k \Big(\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(C_k), \tag{113}$$

$$\mathcal{H}^{d_{\text{int}}}(\mathcal{M}_{\text{MTG}}^{(L)}) = \sum_k \Big(\prod_{\ell=1}^{L}\prod_{i=1}^{d_{\text{int}}} \tilde{\sigma}_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(\tilde{C}_k), \tag{114}$$

where $\sigma_{i,k}^{(\ell)}$ and $\tilde{\sigma}_{i,k}^{(\ell)}$ are the singular values of the Jacobians of the original and modified layers, respectively, and C_k and $\tilde{C}_k$ are the linear regions of the input manifold under the original and modified networks.

Message tuning introduces learnable parameters $\boldsymbol{M}^{(\ell)}$ and $\boldsymbol{W}_p^{(\ell)}$ at each layer. Crucially, message tuning can simulate prompt tuning by appropriately setting these parameters. However, it also has additional degrees of freedom that allow it to reduce measure contraction.

For any layer ℓ and linear region k, message tuning can achieve:

$$\prod_{i=1}^{d_{\text{int}}} \tilde{\sigma}_{i,k}^{(\ell)} \geq \prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}. \tag{115}$$

This is because the product of singular values can be increased by adjusting the parameters to reduce contraction. Let us consider a specific example to illustrate this possibility, assuming that all mappings are constructed under the same partition.

Consider the modified layer map in message tuning:

$$\Psi^{(\ell)} = F^{(\ell)} \circ \mathfrak{F}^{(\ell)}, \tag{116}$$

where $F^{(\ell)}$ is the original layer map and $\mathfrak{F}^{(\ell)}$ is the fusion operation. In a linear region C_k, both maps are linear and injective on the tangent space of the input manifold, which has dimension d_{int}.

By Theorem 1, for a measurable set S in the tangent space, the d_{int}-dimensional Hausdorff measure transforms as:

$$\mathcal{H}^{d_{\text{int}}}(\mathfrak{F}^{(\ell)}(S)) = \Big(\prod_{i=1}^{d_{\text{int}}} \tau_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(S), \tag{117}$$

where $\tau_{1,k}^{(\ell)}, \ldots, \tau_{d_{\text{int}},k}^{(\ell)}$ are the largest d_{int} singular values of the Jacobian of $\mathfrak{F}^{(\ell)}$ restricted to the tangent space. Similarly,

$$\mathcal{H}^{d_{\text{int}}}(F^{(\ell)}(\mathfrak{F}^{(\ell)}(S))) = \Big(\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(\mathfrak{F}^{(\ell)}(S)) = \Big(\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\Big(\prod_{i=1}^{d_{\text{int}}} \tau_{i,k}^{(\ell)}\Big)\mathcal{H}^{d_{\text{int}}}(S). \tag{118}$$

Thus, for the composite map $\tilde{F}^{(\ell)}$, the product of singular values is:

$$\prod_{i=1}^{d_{\text{int}}} \tilde{\sigma}_{i,k}^{(\ell)} = \Big(\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\Big(\prod_{i=1}^{d_{\text{int}}} \tau_{i,k}^{(\ell)}\Big). \tag{119}$$

Consider the fusion operation $\mathfrak{F}^{(\ell)}$. Its Jacobian with respect to $\boldsymbol{H}^{(\ell-1)}$ is:

$$\boldsymbol{J}_{\mathfrak{F}}^{(\ell)} = \boldsymbol{I} + \frac{\partial}{\partial \boldsymbol{H}^{(\ell-1)}}\left(\text{Softmax}(\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_p^{(\ell)}) \cdot \boldsymbol{M}^{(\ell)}\right). \tag{120}$$

By training $\boldsymbol{W}_p^{(\ell)}$ and $\boldsymbol{M}^{(\ell)}$, we can influence the singular values of $\boldsymbol{J}_{\mathfrak{F}}^{(\ell)}$. For example:

- If $\boldsymbol{W}_p^{(\ell)} = \boldsymbol{O}$ and $\boldsymbol{M}^{(\ell)} = \boldsymbol{O}$, then $\mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}) = \boldsymbol{H}^{(\ell-1)}$, so $\boldsymbol{J}_{\mathfrak{F}}^{(\ell)} = \boldsymbol{I}$, and the singular values are 1.
- If $\boldsymbol{W}_p^{(\ell)}$ and $\boldsymbol{M}^{(\ell)}$ are trained such that the second term is positive definite, then the singular values can be greater than 1.

Thus, by parameter choice, we can ensure:

$$\prod_{i=1}^{d_{\text{int}}} \tau_{i,k}^{(\ell)} \geq 1. \tag{121}$$

From the above, we have:

$$\prod_{i=1}^{d_{\text{int}}} \tilde{\sigma}_{i,k}^{(\ell)} = \Big(\prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}\Big)\Big(\prod_{i=1}^{d_{\text{int}}} \tau_{i,k}^{(\ell)}\Big) \geq \prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}, \tag{122}$$

This proves that message tuning can achieve the desired inequality for any layer ℓ and linear region k. Moreover, if $\prod_{i=1}^{d_{\text{int}}} \tau_{i,k}^{(\ell)} > 1$, the inequality is strict.

The input manifold $\mathcal{M}_0$ is fixed. Prompt tuning shifts it to $\mathcal{M}_0(\boldsymbol{P})$, but the fusion operation $\mathfrak{F}^{(1)}$ in the first layer also possesses the capability to adjust the input manifold, we may reasonably assume that $\mathcal{H}^{d_{\text{int}}}(\mathcal{M}_0(\boldsymbol{P})) = \mathcal{H}^{d_{\text{int}}}(\mathcal{M}_0)$.

The linear regions C_k and $\tilde{C}_k$ are partitions of $\mathcal{M}_0(\boldsymbol{P})$ and $\mathcal{M}_0$ induced by the piecewise linear maps Φ and Φ_{MTG}, respectively. Message tuning modifies the network architecture, which may refine the linear regions. However, the total measure of the input manifold is conserved:

$$\sum_k \mathcal{H}^{d_{\text{int}}}(C_k) = \mathcal{H}^{d_{\text{int}}}(\mathcal{M}_0(\boldsymbol{P})) = \mathcal{H}^{d_{\text{int}}}(\mathcal{M}_0) = \sum_k \mathcal{H}^{d_{\text{int}}}(\tilde{C}_k). \tag{123}$$

While individual regions may change, the overall sum remains unchanged. Therefore, for the purpose of comparing the sums, we have:

$$\sum_k \mathcal{H}^{d_{\text{int}}}(\tilde{C}_k) = \sum_k \mathcal{H}^{d_{\text{int}}}(C_k). \tag{124}$$

From the above, for any prompt $\boldsymbol{P}$, message tuning can choose parameters such that for each layer ℓ and region k:

$$\prod_{i=1}^{d_{\text{int}}} \tilde{\sigma}_{i,k}^{(\ell)} \geq \prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}. \tag{125}$$

Moreover, since the input measures are equal, we have:

$$\mathcal{H}^{d_{\text{int}}}(\mathcal{M}_{\text{MTG}}^{(L)}) = \sum_{k} \Big(\prod_{\ell=1}^{L} \prod_{i=1}^{d_{\text{int}}} \tilde{\sigma}_{i,k}^{(\ell)} \Big) \mathcal{H}^{d_{\text{int}}}(\tilde{C}_k) \geq \sum_{k} \Big(\prod_{\ell=1}^{L} \prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)} \Big) \mathcal{H}^{d_{\text{int}}}(C_k) = \mathcal{H}^{d_{\text{int}}}(\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P})). \tag{126}$$

The inequality holds term-wise due to the non-decrease in singular value products and the conservation of input measure.

There exists a message tuning configuration where the inequality is strict. For example, if we train $\boldsymbol{W}_p^{(\ell)}$ and $\boldsymbol{M}^{(\ell)}$ such that for some layer ℓ and region k, $\prod_{i=1}^{d_{\text{int}}} \tilde{\sigma}_{i,k}^{(\ell)} > \prod_{i=1}^{d_{\text{int}}} \sigma_{i,k}^{(\ell)}$, and since the input measure is positive, the overall measure increases strictly.

Thus, we conclude that:

$$\mathcal{H}^{d_{\text{int}}}(\mathcal{M}_{\text{MTG}}^{(L)}) \geq \mathcal{H}^{d_{\text{int}}}(\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P})) \quad \text{for all } \boldsymbol{P} \in \mathcal{P}, \tag{127}$$

and the inequality is strict for some configuration.

Diameter Comparison.

The diameter of a set $\mathcal{M}$ is:

$$\text{diam}(\mathcal{M}) = \sup_{x,y \in \mathcal{M}} \|x - y\|. \tag{128}$$

For any prompt $\boldsymbol{P}$, message tuning can simulate prompt tuning by setting:

- $\mathfrak{F}^{(1)}(\boldsymbol{H}^{(0)}, \boldsymbol{M}^{(1)}; \boldsymbol{W}_p^{(1)}) \overset{\sim}{\longrightarrow} \mathcal{M}_0(\boldsymbol{P})$,
- $\mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}, \boldsymbol{M}^{(\ell)}; \boldsymbol{W}_p^{(\ell)}) = \boldsymbol{H}^{(\ell-1)}$ for $\ell \geq 2$.

This reduces message tuning to prompt tuning, giving:

$$\mathcal{M}_{\text{MTG}}^{(L)} = \mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P}), \tag{129}$$

and hence:

$$\text{diam}(\mathcal{M}_{\text{MTG}}^{(L)}) = \text{diam}(\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P})). \tag{130}$$

Thus, the inequality holds with equality for this configuration.

We now show that message tuning can achieve a strictly larger diameter by leveraging its additional parameters to expand the output space.

Message tuning can expand the distance between representations layer-wise. Consider the fusion operation:

$$\mathfrak{F}^{(\ell)}(\boldsymbol{H}^{(\ell-1)}) = \boldsymbol{H}^{(\ell-1)} + \text{Softmax}(\boldsymbol{H}^{(\ell-1)} \boldsymbol{W}_p^{(\ell)}) \cdot \boldsymbol{M}^{(\ell)}. \tag{131}$$

By choosing $\boldsymbol{W}_p^{(\ell)}$ and $\boldsymbol{M}^{(\ell)}$ appropriately, we can make $\mathfrak{F}^{(\ell)}$ an expanding map. For example:

- Set $\boldsymbol{W}_p^{(\ell)}$ to have orthonormal columns.
- Set $\boldsymbol{M}^{(\ell)} = c \cdot \boldsymbol{W}_p^{(\ell)}$ for some $c > 0$.

Then, the Jacobian of $\mathfrak{F}^{(\ell)}$ satisfies:

$$\boldsymbol{J}_{\mathfrak{F}}^{(\ell)}(\boldsymbol{H}) = \boldsymbol{I} + c \cdot \boldsymbol{W}_p^{(\ell)} \cdot \boldsymbol{J}_{\text{softmax}}(\boldsymbol{H} \boldsymbol{W}_p^{(\ell)}) \cdot \boldsymbol{W}_p^{(\ell)\top}, \tag{132}$$

which has eigenvalues ≥ 1 (since $\boldsymbol{J}_{\text{softmax}}$ is positive semidefinite). By choosing c large, we can make $\mathfrak{F}^{(\ell)}$ arbitrarily expansive and Φ_{MTG} results from the superposition of such expansion effects.

Thus, for the pair $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{M}_0(\boldsymbol{P})$ achieving the diameter of $\Phi(\mathcal{M}_0)$, message tuning can ensure:

$$\text{diam}(\mathcal{M}_{\text{MTG}}^{(L)}) \geq \|\Phi_{\text{MTG}}(\boldsymbol{X}) - \Phi_{\text{MTG}}(\boldsymbol{Y})\| > \|\Phi(\boldsymbol{X}) - \Phi(\boldsymbol{Y})\| = \text{diam}(\Phi(\mathcal{M}_0(\boldsymbol{P}))). \tag{133}$$

Thus, we have:

$$\text{diam}(\mathcal{M}_{\text{MTG}}^{(L)}) > \text{diam}(\mathcal{M}_{\text{PT}}^{(L)}(\boldsymbol{P})) \quad \text{for all } \boldsymbol{P} \in \mathcal{P}. \tag{134}$$

For any prompt $\boldsymbol{P}$, message tuning can simulate prompt tuning, so:

$$\mathrm{diam}(\mathcal{M}_{\mathrm{MTG}}^{(L)}) \geq \mathrm{diam}(\mathcal{M}_{\mathrm{PT}}^{(L)}(\boldsymbol{P})). \tag{135}$$

Moreover, by choosing parameters to expand inter-point distances, message tuning can achieve:

$$\mathrm{diam}(\mathcal{M}_{\mathrm{MTG}}^{(L)}) > \mathrm{diam}(\mathcal{M}_{\mathrm{PT}}^{(L)}(\boldsymbol{P})) \quad \text{for all } \boldsymbol{P} \in \mathcal{P}. \tag{136}$$

This completes the proof. □

C.2 Analysis of Negative Transfer

Definition 14 (Negative Transfer from a Manifold Perspective). *Let $\mathcal{M}_0^s \subset \mathbb{R}^{N \times d_0}$ and $\mathcal{M}_0^t \subset \mathbb{R}^{N \times d_0}$ be the compact smooth input manifolds of the source and target domains, respectively, with intrinsic dimensions D_s and D_t. Let $\Phi = F^{(L)} \circ \cdots \circ F^{(1)}$ be the map of the GFM, and let $\mathcal{M}_s^{(L)} = \Phi(\mathcal{M}_0^s)$ and $\mathcal{M}_t^{(L)} = \Phi(\mathcal{M}_0^t)$ be the representation spaces. Negative transfer is said to occur if the map Φ causes a geometric misalignment or structural distortion between the transformed spaces, such that:*

- *Information Loss: The intrinsic dimension or geometric measure (e.g., volume) of $\Phi(\mathcal{M}_0^t)$ is significantly reduced compared to $\Phi(\mathcal{M}_0^s)$.*
- *Poor Alignment: The transformed spaces $\Phi(\mathcal{M}_0^s)$ and $\Phi(\mathcal{M}_0^t)$ are poorly aligned, as quantified by a large Hausdorff distance or a small intersection measure.*

Remark 9. *Fine-tuning severely exacerbates negative transfer in graph data because it aggressively warps the target space's geometry to fit the source domain's feature space. This often collapses the intrinsic structure of the target graph, leading to catastrophic information loss and misalignment. Prompt tuning alleviates negative transfer by gently realigning the target space within the frozen source feature space, preserving its intrinsic geometry and measure to prevent catastrophic distortion or collapse.*

Corollary 2 (Message Tuning Mitigates Negative Transfer). *Negative transfer often arises when the model's capacity is insufficient to capture the target domain's distribution, leading to interference from source domain features. The higher intrinsic dimension $d_{int}(\mathcal{M}_{MTG}^{(L)})$ indicates that MTG can learn more diverse features, reducing reliance on source-specific patterns. The greater Hausdorff measure $\mathcal{H}^{d_{int}}(\mathcal{M}_{MTG}^{(L)})$ implies a larger "volume" of the space, accommodating a wider range of target domain variations. The increased diameter $diam(\mathcal{M}_{MTG}^{(L)})$ signifies that the representations span a broader range, enhancing model flexibility. In contrast, prompt tuning only perturbs the input manifold $\mathcal{M}_0$ via prompts, which constrains adaptation to superficial layers and may insufficiently adjust internal representations, thus more likely to lead to negative transfer.*

By further refining and extending Prismatic Space Theory, a theoretical characterization of negative transfer in GFMs can be established. We identify this as a direction for future research.

D Datasets and Experimental Details

D.1 Configuration

The experiments are conducted on a Linux server equipped with an Intel(R) Xeon(R) Gold 6240 CPU @ 2.60GHz, 256GB RAM and 2 NVIDIA A100-SXM4-40GB GPUs. Our implementation is based on PyTorch (Paszke et al., 2019) version 2.2.1, PyG (Fey & Lenssen, 2019) version 2.6.1 with CUDA version 12.1 and Python 3.12.7.

D.2 Details of Datasets

Homophilic Graphs. Cora and Citeseer datasets (Sen et al., 2008) represent computer science publications, with nodes encoded as bag-of-words features and labeled by research topics. Pubmed (Yang et al., 2016) contains diabetes-related articles from PubMed database, with nodes represented by TF/IDF-weighted word vectors and classified by diabetes type. ogbn-arxiv (Hu et al., 2020a) is a large-scale citation network of CS arXiv papers, where nodes represent papers with 128-dimensional title+abstract embeddings, and directed edges denote citations.

Heterophilic Graphs. Texas and Wisconsin (Pei et al., 2020) datasets are WebKB subgraphs comprising university web pages, where nodes represent pages with bag-of-words features and edges indicate hyperlinks. Pages are classified into five categories: student, project, course, staff, and faculty. Actor dataset (Pei et al., 2020) forms a co-occurrence network with actors as nodes and Wikipedia page co-appearances as edges.

Biological Graphs. D&D dataset (Dobson & Doig, 2003) contains 1,178 protein graphs where nodes represent amino acids connected by edges, classified as enzymes/non-enzymes. ENZYMES (Borgwardt et al., 2005) comprises 600 enzyme structures from BRENDA, categorized into 6 EC classes. PROTEINS (Wang et al., 2022) represents tertiary protein structures with nodes as secondary structure elements and edges indicating sequence/3D proximity, yielding binary graph classification.

Small Molecule Graphs. BZR dataset (Rossi & Ahmed, 2015) contains 405 benzodiazepine receptor ligand graphs with binary classification. COX2 (Rossi & Ahmed, 2015) comprises 467 cyclooxygenase-2 inhibitor molecular graphs, where nodes represent atoms and edges encode bond types (single/double/triple/aromatic), also yielding binary classification. MUTAG (Kriege & Mutzel, 2012) includes 188 mutagenic aromatic compounds classified into 7 categories.

Social Network Graphs. COLLAB (Yanardag & Vishwanathan, 2015) represents scientific collaboration networks, where nodes denote researchers, edges indicate co-authorships, and graphs are classified by research fields. IMDB-B (Yanardag & Vishwanathan, 2015) captures actor collaboration networks, with nodes representing performers, edges signifying co-appearances in films, and binary graph labels distinguishing Action versus Romance genres.

Table 6: Statistics of all datasets.

Dataset	Task	# Graphs	# Nodes	# Edges	# Features	# Classes	Graph Type
Cora	Node	1	2,708	5,429	1,433	7	Homophilic
CiteSeer	Node	1	3,327	9,104	3,703	6	Homophilic
Pubmed	Node	1	19,717	88,648	500	3	Homophilic
Texas	Node	1	183	325	1703	5	Heterophilic
Actor	Node	1	7600	30019	932	5	Heterophilic
Wisconsin	Node	1	251	515	1703	5	Heterophilic
ogbn-arxiv	Node	1	169,343	1,166,243	128	40	Large-scale
D&D	Graph	1,178	284.1	715.7	89	2	Proteins
ENZYMES	Graph	600	32.6	62.1	3	6	Proteins
PROTEINS	Graph	1,113	39.1	72.8	3	2	Proteins
BZR	Graph	405	35.8	38.4	3	2	Small Molecule
COX2	Graph	467	41.2	43.5	3	2	Small Molecule
MUTAG	Graph	188	17.9	19.8	7	2	Small Molecule
COLLAB	Graph	5000	74.5	2457.8	0	3	Social Network
IMDB-B	Graph	1000	19.8	96.53	0	2	Social Network

D.3 Data Split.

We adopt the same dataset processing methodology as ProG (Zi et al., 2024) to ensure consistency and comparability with prior work. For the node classification task, we adopt a 90% test set allocation to rigorously evaluate model performance. In contrast, for the graph classification task, we employ an 80 % test set split to maintain a balance between evaluation rigor and training data availability. To ensure statistical robustness and mitigate potential sampling bias, we repeat the random sampling procedure five times to construct distinct k-shot learning tasks for both task types. The final performance metrics are reported as the mean and standard deviation across these five independent trials, providing a comprehensive assessment of model stability and generalization capability.

D.4 Evaluation Metrics

In node and graph classification tasks, AUROC (Area Under the Receiver Operating Characteristic Curve) and F1-score serve as two critical evaluation metrics. AUROC quantifies a model's class discrimination capability, where 1 represents perfect classification and 0.5 indicates random guessing. The F1-score, which harmonizes precision (correctness of positive predictions) and recall (coverage of actual positives), ranges from 0 to 1, with higher values indicating better performance. This metric is particularly valuable for imbalanced datasets. For multi-class scenarios, we employ a macro-averaging approach, where each class is iteratively treated as positive while aggregating results. Both metrics are computed via a one-vs-rest strategy for class-wise evaluation.

D.5 Hyperparameter Configuration

In most experiments, the model architecture consists of 2 layers with a hidden dimension of 128. We develop a systematic random search strategy to identify optimal hyperparameters for each adaptation method across all datasets, extending beyond default configurations. Considering the substantial heterogeneity in hyperparameter requirements among different adaptation approaches, we concentrate on tuning three key hyperparameters through random search: (1) learning rate, sampled from $\{0.001, 0.005, 0.01, 0.05, 0.1\}$; (2) weight decay, selected from $\{0, 0.00001, 0.0001, 0.001, 0.01\}$; and (3) batch size, uniformly sampled from $\{32, 64, 128\}$ in each experimental trial. This comprehensive search strategy ensures robust parameter optimization while maintaining methodological consistency across diverse experimental conditions.

D.6 Implementation Details

To ensure experimental fairness and demonstrate the compatibility of our approach, we implement MTG based on the ProG library (Zi et al., 2024). We have made some modifications to the ProG library to adapt it to MTG, but these changes do not affect the original prompt tuning method at all.

E Details of Baselines

E.1 Backbones of Graph Foundation Models

GCN (Graph Convolutional Network) (Kipf & Welling, 2017) employs convolutional operations to aggregate and transform feature information from a node's immediate neighborhood. This localized message-passing mechanism allows the network to iteratively refine node representations by incorporating structural and attribute information from adjacent nodes, effectively capturing the graph's topological properties.

GraphSAGE (Hamilton et al., 2017) is an inductive learning framework that computes node embeddings through a localized feature aggregation process. Instead of relying on fixed graph convolutions, it operates by sampling neighboring nodes and hierarchically aggregating their features using learnable functions. This approach enables the model to generalize to unseen graph structures while capturing both node attributes and local topological patterns.

GAT (Graph Attention Network) (Veličković et al., 2018) introduces an attention mechanism into graph neural networks, dynamically computing attention weights between connected nodes during feature aggregation. By learning to assign differential importance to neighboring nodes, GAT can focus on more relevant connections while suppressing noisy or less informative edges. This adaptive weighting scheme enhances model expressiveness and interpretability compared to standard aggregation approaches.

GIN (Graph Isomorphism Network) (Xu et al., 2019) is a theoretically motivated GNN architecture designed to maximize discriminative power in graph representation learning. By employing injective multiset aggregation functions and MLP-based transformations, GIN achieves provable expressiveness equivalent to the Weisfeiler-Lehman graph isomorphism test. This framework demonstrates superior capability in distinguishing graph structures while maintaining efficient computation through neighborhood aggregation.

GT (Graph Transformer) (Ying et al., 2021) adapts the Transformer architecture to graph-structured data by incorporating structural biases into the self-attention mechanism. Through masked attention patterns that respect graph connectivity, the model efficiently captures both local and global dependencies while maintaining the parallelizability of standard Transformers. This approach enables the simultaneous modeling of node features and graph topology through position-aware attention computations. The architecture demonstrates particular effectiveness in scenarios requiring long-range dependency modeling across graph structures.

E.2 Pre-training Strategies

DGI(Veličković et al., 2019) is a self-supervised learning framework that employs mutual information maximization for graph representation learning. The method optimizes the mutual information between patch-level node representations and global graph summaries through a contrastive objective. By leveraging negative sampling and discriminator functions, DGI learns informative node embeddings that preserve both local structural patterns and global graph characteristics. This approach demonstrates particular effectiveness in scenarios with limited labeled data, enabling effective transfer learning across graph-based tasks.

GraphMAE(Hou et al., 2022) adopts a self-supervised pretraining approach based on feature reconstruction of masked nodes. The framework randomly masks portions of node features and learns to recover them through an encoder-decoder architecture, forcing the model to develop robust structural understanding from contextual patterns. This denoising objective promotes the learning of generalized graph representations that capture both local neighborhood characteristics and global topological properties. The method demonstrates particular effectiveness in scenarios requiring transferable graph representations across different downstream tasks.

EdgePreGPPT(Sun et al., 2022) introduces a novel graph pre-training paradigm that fundamentally reconfigures structural knowledge acquisition in graph neural networks. The framework employs masked edge prediction as its foundational pretext task, where the model learns to reconstruct randomly obscured connections through an edge prediction module. This pre-training phase focuses on optimizing pairwise node similarity computations, enabling the model to develop robust representations of graph topology and connectivity patterns. The methodology's effectiveness stems from its direct optimization of structural relationships between nodes, training the network to evaluate connection probabilities through learned embedding similarities.

EdgePreGprompt(Liu et al., 2023b) establishes a novel paradigm for learning transferable structural representations from label-free graph data. At its core, the framework employs link prediction as its self-supervised pretext task, leveraging the abundant connectivity patterns naturally available in graph structures without requiring additional annotation. The methodology operates by first constructing contextual subgraphs for nodes, which capture not only node-specific features but also rich topological information from their local neighborhoods. Specifically, the framework optimizes a contrastive objective that maximizes the similarity between linked node pairs while minimizing similarity for non-linked pairs, thereby encoding fundamental graph connectivity patterns into the learned representations.

GraphCL(You et al., 2020) introduces a graph contrastive learning framework that learns transferable graph representations through self-supervised pre-training by maximizing agreement between different augmented views of the same graph. The method employs four key augmentation strategies—node dropping, edge perturbation, attribute masking, and subgraph sampling—each encoding domain-specific priors about structural invariance. These augmentations generate correlated views that are processed through a shared GNN encoder, projected via an MLP head, and optimized using an NT-Xent loss function to enhance similarity between positive pairs while contrasting negative samples. The framework theoretically maximizes mutual information between augmented views, unifying various contrastive learning approaches for graphs.

SimGRACE(Xia et al., 2022) presents a novel graph contrastive learning framework that eliminates the need for manual data augmentation by instead leveraging encoder perturbations to generate contrasting views. The core methodology involves feeding the original graph through both a standard GNN encoder and its perturbed version, where the perturbation is achieved by adding Gaussian noise to the encoder weights, thereby producing correlated representations without altering input data semantics. These dual representations are then projected through a shared MLP head and optimized using the NT-Xent loss to maximize agreement between positive pairs while contrasting with negative samples from the same batch.

E.3 Prompt Tuning Baselines

GPPT (Sun et al., 2022) introduces an innovative graph prompting function that bridges the gap between pre-training and downstream tasks by reformulating node classification as an edge prediction problem through token pair construction. The framework converts standalone nodes into structured token pairs composed of two components: a task token that represents candidate labels through trainable continuous vectors and a structure token that encodes neighborhood information by aggregating adjacent nodes with attention-based weighting. The approach fundamentally rethinks graph transfer learning by aligning task formulations rather than forcing downstream adaptation to mismatched pre-training objectives.

Gprompt (Liu et al., 2023b) introduces a unified prompting framework that bridges graph pre-training and downstream tasks through a subgraph similarity template. The core innovation involves learnable task-specific prompt vectors that dynamically reweight node features during subgraph aggregation operations such as READOUT, allowing downstream tasks including node classification and graph classification to selectively extract relevant knowledge from frozen pre-trained GNNs. The prompt vectors act as lightweight task adapters, preserving the pre-trained model's parameters while tailoring subgraph representations through dimension-wise feature importance scoring, demonstrating superior parameter efficiency and few-shot performance across diverse graph tasks.

All-in-one (Sun et al., 2023a) introduces a unified multi-task prompting framework for graph neural networks that effectively connects various downstream tasks at node, edge, and graph levels with graph pre-training through several key innovations. First, it employs task reformulation by transforming node and edge tasks into graph-level tasks through induced subgraph construction. Second, it incorporates a learnable prompt graph featuring tunable tokens, dynamic token structures, and adaptive insertion patterns to align downstream tasks with pre-training objectives. Third, it utilizes meta-learning optimization to generalize prompts across different tasks. The framework maintains frozen pre-trained GNNs while only tuning lightweight prompt parameters, enabling efficient knowledge transfer with task-specific adaptability.

GPF (Fang et al., 2023) introduces a unified approach to prompt tuning by focusing on feature space adaptation within graph neural networks. It employs a shared learnable vector that is added to all node features in the input graph, creating a consistent modification across the entire structure. This design allows the pre-trained model to maintain its frozen parameters while adapting to downstream tasks through subtle yet effective feature adjustments. The approach demonstrates theoretical equivalence to any form of prompting function, making it universally applicable across diverse pre-training strategies without requiring task-specific templates.

GPF-plus (Fang et al., 2023) enhances flexibility by assigning distinct learnable vectors to individual nodes through an attention-based mechanism. Rather than using a single global prompt, it generates node-specific prompts by combining a set of basis vectors with weights derived from each node's features. This architecture captures finer-grained adaptations while maintaining parameter efficiency through basis sharing. The method automatically adjusts to graphs of varying scales and complexities, offering improved expressiveness over GPF while preserving its universal applicability and theoretical guarantees for effective knowledge transfer.

F More Information on Experiments

F.1 Details of the experimental results on 1/3/5-shot node/graph classification

While one-shot node/graph classification presents the most challenging scenario for evaluating adaptation methods, few-shot node/graph classification remains a critical task for assessing the robustness and generalization capability of these methods. Therefore, we have extended our evaluation to include 3-shot and 5-shot node/graph classification tasks across various adaptation approaches. The optimal performance of various adaptation methods, alongside supervised learning baseline, under 3-shot and 5-shot settings, is summarized in Tables 9-12. Consistent with the earlier experimental findings presented in Subsection 5.2, these results demonstrate that MTG substantially enhances the performance of multiple pre-trained GFMs across 1/3/5-shot scenarios, thereby significantly improving the transferability of pre-trained knowledge within the "Pre-training and Adaptation" paradigm. Notably, MTG consistently exhibits superior compatibility with diverse pre-training strategies on both node-level and graph-level tasks. In contrast, among prompt tuning methods, while All-in-one demonstrates competitive performance on graph-level tasks, it suffers from severe performance degradation on certain node-level datasets, such as ogbn-arxiv.

Furthermore, a comprehensive evaluation of all adaptation methods under 1/3/5-shot settings, as measured by three key metrics including Accuracy, F1 score and AUROC, is provided in Tables 14-31. Through these more detailed experimental results, we observe that GPF-plus, as a simple and general prompt tuning approach, demonstrates strong overall performance across both downstream tasks under the 1/3/5-shot settings, making it the second-best adaptation method after MTG. It is worth noting that GPF-plus can be regarded, to some extent, as a special case of MTG in which learnable parameters are injected solely into the first pre-trained GFM layer, effectively equivalent to operating directly on the input graph data. The baseline results in this experimental section also combine those from ProG (Zi et al., 2024) with our own reproductions.

F.2 Performance with More Backbones for GFMs

For GNN-based GFMs, both prompt tuning and message tuning are universal adaptation methods that are not limited to specific model architectures. Therefore, in this subsection, we evaluate the performance of various adaptation methods on five of the most classic, popular, and widely used GFM backbone models. Tables 32 and 33 present the performance of different adaptation methods based on various backbone models on the representative datasets Wisconsin and PROTEINS. These results once again confirm that prompt tuning and message tuning outperform fine-tuning, while our proposed MTG demonstrates even more significant advantages. For more complex GFMs, such as models that integrate LLMs with GNNs, MTG can also be naturally adapted to the GNN module or the module responsible for fusing features obtained from LLMs and GNNs. The core idea of MTG is to perform layer-wise parameter injection for message fusion regulation, which is not constrained by any specific model architecture. We believe this represents a promising direction for future research.

Most experiments in this paper employ a relatively basic 2-layer backbone model, which may not fully demonstrate the performance advantages of MTG. To further investigate the impact of model depth on adaptation methods, we continue to use the GCN backbone model and representative datasets Cora and BZR to evaluate the performance of various adaptation methods when applied to models with 4, 8, 12, and 16 layers in downstream tasks. The results in Table 13 confirm that MTG still maintains significant advantages even with deeper model architectures.

F.3 Computational Efficiency of MTG

As a general adaptation method, MTG inherently possesses the advantage of parameter efficiency. It does not impose significant computational burden on the original model and requires substantially fewer parameters than fine-tuning to achieve effective adaptation on downstream tasks. In this subsection, we take the GCN backbone model as an example and first provide a theoretical analysis of the time complexity and trainable parameter complexity of both fine-tuning and MTG.

Fine-tuning. The time complexity per layer of a GCN with L layers, where each layer transforms input features of dimension $d_{\ell-1}$ to output dimension d_ℓ, comprises two main components: the feature transformation via matrix multiplication between the weight matrix $\boldsymbol{W}^{(\ell)} \in \mathbb{R}^{d_{\ell-1} \times d_\ell}$ and the node feature matrix $\boldsymbol{H}^{(\ell-1)} \in \mathbb{R}^{|\mathcal{V}|\times d_{\ell-1}}$, with complexity $O(|\mathcal{V}|d_{\ell-1}d_\ell)$, and the neighborhood aggregation through sparse matrix multiplication between the normalized adjacency matrix $\tilde{\boldsymbol{A}} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$ and the transformed features, requiring $O(|\mathcal{E}|d_\ell)$ operations, where $|\mathcal{E}|$ denotes the number of edges. Assuming all hidden dimensions are equal d, the total time complexity becomes $O(L(|\mathcal{V}|d^2 + |\mathcal{E}|d))$. The space complexity for trainable parameters is dominated by the weight matrices, yielding $O(Ld^2)$.

Message Tuning. MTG introduces three additional components per layer: message vectors $\boldsymbol{M}^{(\ell)} \in \mathbb{R}^{m\times d_{\ell-1}}$ containing m learnable message prototypes, a projection matrix $\boldsymbol{W}_p^{(\ell)} \in \mathbb{R}^{d_{\ell-1}\times m}$ to compute attention scores, and an attention mechanism $\alpha = \text{Softmax}(\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_p^{(\ell)}) \in \mathbb{R}^{|\mathcal{V}|\times m}$, with the corresponding computational overhead consisting of the projection operation $\boldsymbol{H}^{(\ell-1)}\boldsymbol{W}_p^{(\ell)}$ requiring $O(|\mathcal{V}|d_{\ell-1}m)$ operations, the attention computation including softmax and matrix multiplication $\alpha\boldsymbol{M}^{(\ell)}$ requiring $O(|\mathcal{V}|m^2)$ operations, and message integration via element-wise addition with original features requiring $O(|\mathcal{V}|d_{\ell-1})$ operations. Thus, the total time complexity becomes $O(L(|\mathcal{V}|d^2 + |\mathcal{E}|d + |\mathcal{V}|dm + |\mathcal{V}|m^2))$. Since $m \ll d$ typically holds, MTG does not introduce significant inference time overhead to the original GCN, and their time complexities remain essentially within the same order of magnitude. The trainable parameter complexity comes from $M^{(\ell)}$ and $P^{(\ell)}$ matrices, contributing $O(L(dm + md)) = O(Ldm)$ parameters, which is lower than fine-tuning the entire GCN model.

The above analysis offers a theoretical perspective on model inference time and trainable parameters; however, it should be noted that such theoretical estimates may differ from practical performance. Due to variations in their practical implementations, various prompt tuning methods are not amenable to straightforward computational complexity analysis. Therefore, we further conduct a comparative analysis of the actual training time per epoch and GPU memory consumption between prompt tuning methods and MTG on the large-scale dataset ogbn-arxiv, which has the largest number of nodes, and the COLLAB dataset, which contains the most graphs. The pre-training strategy uses DGI and experimental results are presented in Table 7. Due to its distinct data loading mechanism, GPPT exhibits significantly different GPU memory usage compared to other methods. Excluding GPPT, MTG demonstrates advantages in both training speed and memory consumption. It should be emphasized that MTG exhibits superior training efficiency compared to All-in-one.

F.4 Sensitivity Analysis

Message tuning injects m learnable message vectors at each layer of the model, making m a hyperparameter of MTG. We further conduct a sensitivity analysis on this hyperparameter m using the GCN backbone model and representative datasets Cora and BZR, evaluating the performance of MTG when $m = 3, 5, 10, 20, 30$. The optimal results described in Subsection 5.2 are presented in Table 8. These results demonstrate that MTG exhibits a certain degree of robustness to this hyperparameter, as no performance collapse occurs even with very small or large values of m. In our experiments, m is typically set to 10; nevertheless, careful selection of m remains necessary to fully exploit the potential of MTG across different datasets.

Table 7: Computational efficiency comparison of prompt tuning and message tuning.

Methods	**ogbn-arxiv (1-shot)**		**COLLAB (1-shot)**	
	Time (s)	Memory (MB)	Time (s)	Memory (MB)
GPPT	0.6032	3499	0.0204	32357
Gprompt	0.0326	10987	0.0081	3517
All-in-one	0.0559	11023	0.0147	3767
GPF	0.0067	10963	0.0045	3515
GPF-plus	0.0074	10983	0.0057	3517
MTG (Ours)	0.0053	10963	0.0036	3515

Table 8: Performance sensitivity to the number of message prototypes m in MTG.

Dataset	$m = 3$	$m = 5$	$m = 10$	$m = 20$	$m = 30$
Cora (1-shot)	$51.28_{\pm 7.29}$	$54.06_{\pm 4.49}$	$58.54_{\pm 7.89}$	$56.73_{\pm 5.51}$	$56.21_{\pm 7.02}$
BZR (1-shot)	$73.15_{\pm 15.26}$	$77.84_{\pm 2.22}$	$74.81_{\pm 13.96}$	$77.53_{\pm 2.39}$	$72.78_{\pm 17.86}$

Table 9: Performance comparison of adaptation methods on 3-shot node classification (accuracy±std %, 5 runs). The first, second and third best results are shaded in red, with descending color saturation.

Method	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	37.79$_{\pm 9.16}$	35.18$_{\pm 6.86}$	57.33$_{\pm 4.64}$	41.03$_{\pm 6.40}$	40.78$_{\pm 12.55}$	18.62$_{\pm 3.46}$	19.03$_{\pm 5.08}$
Fine-tuning	51.97$_{\pm 2.84}$	45.08$_{\pm 2.09}$	65.40$_{\pm 3.00}$	42.40$_{\pm 7.77}$	43.13$_{\pm 13.79}$	22.11$_{\pm 1.97}$	27.34$_{\pm 6.61}$
GPPT	43.84$_{\pm 6.11}$	42.34$_{\pm 8.31}$	67.43$_{\pm 2.96}$	34.29$_{\pm 4.71}$	38.90$_{\pm 8.86}$	21.65$_{\pm 3.39}$	22.46$_{\pm 4.05}$
Gprompt	63.78$_{\pm 5.77}$	60.00$_{\pm 6.18}$	66.68$_{\pm 3.53}$	92.52$_{\pm 5.38}$	39.00$_{\pm 47.08}$	29.67$_{\pm 2.53}$	73.92$_{\pm 2.75}$
All-in-one	48.09$_{\pm 4.83}$	48.09$_{\pm 8.18}$	65.79$_{\pm 5.79}$	89.62$_{\pm 4.38}$	88.69$_{\pm 1.08}$	24.23$_{\pm 1.39}$	31.15$_{\pm 2.25}$
GPF	34.84$_{\pm 19.83}$	25.92$_{\pm 12.30}$	71.20$_{\pm 2.82}$	93.85$_{\pm 3.71}$	95.47$_{\pm 2.75}$	37.44$_{\pm 3.43}$	59.67$_{\pm 12.69}$
GPF-plus	56.38$_{\pm 5.37}$	72.48$_{\pm 5.63}$	70.85$_{\pm 4.03}$	98.15$_{\pm 0.73}$	97.66$_{\pm 0.41}$	43.59$_{\pm 4.52}$	64.63$_{\pm 10.05}$
MTG (Ours)	66.11$_{\pm 6.37}$	73.81$_{\pm 8.56}$	71.38$_{\pm 3.21}$	98.58$_{\pm 0.93}$	98.17$_{\pm 1.40}$	37.62$_{\pm 4.72}$	76.01$_{\pm 5.39}$

Table 10: Performance comparison of adaptation methods on 3-shot graph classification.

Method	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	53.33$_{\pm 6.61}$	50.77$_{\pm 2.44}$	61.33$_{\pm 2.89}$	59.47$_{\pm 8.34}$	15.96$_{\pm 1.64}$	65.15$_{\pm 18.61}$	52.35$_{\pm 8.12}$	59.77$_{\pm 1.10}$
Fine-tuning	66.10$_{\pm 0.70}$	56.10$_{\pm 3.46}$	62.72$_{\pm 2.39}$	59.87$_{\pm 8.78}$	22.71$_{\pm 0.86}$	69.97$_{\pm 13.89}$	52.22$_{\pm 10.64}$	59.70$_{\pm 0.98}$
GPPT	59.48$_{\pm 5.42}$	50.88$_{\pm 6.31}$	64.74$_{\pm 1.99}$	64.13$_{\pm 18.31}$	19.12$_{\pm 2.43}$	71.90$_{\pm 14.28}$	70.93$_{\pm 16.35}$	59.00$_{\pm 6.34}$
Gprompt	64.35$_{\pm 1.21}$	54.95$_{\pm 9.47}$	64.94$_{\pm 2.92}$	66.53$_{\pm 14.84}$	22.08$_{\pm 3.57}$	51.53$_{\pm 13.08}$	54.63$_{\pm 2.95}$	55.99$_{\pm 7.53}$
All-in-one	65.67$_{\pm 0.58}$	57.12$_{\pm 1.99}$	69.84$_{\pm 6.02}$	80.00$_{\pm 5.67}$	23.96$_{\pm 0.62}$	66.06$_{\pm 18.23}$	61.98$_{\pm 11.32}$	58.96$_{\pm 5.93}$
GPF	65.97$_{\pm 0.69}$	53.87$_{\pm 3.44}$	63.35$_{\pm 2.45}$	74.27$_{\pm 1.55}$	23.87$_{\pm 3.45}$	65.31$_{\pm 19.45}$	74.38$_{\pm 11.62}$	59.07$_{\pm 0.65}$
GPF-plus	64.38$_{\pm 2.30}$	56.50$_{\pm 3.71}$	63.55$_{\pm 1.85}$	75.20$_{\pm 3.64}$	24.46$_{\pm 2.27}$	65.25$_{\pm 18.07}$	71.67$_{\pm 14.87}$	59.51$_{\pm 0.62}$
MTG (Ours)	66.95$_{\pm 0.59}$	57.49$_{\pm 2.52}$	70.49$_{\pm 0.68}$	78.13$_{\pm 6.36}$	29.71$_{\pm 2.06}$	73.86$_{\pm 9.74}$	74.65$_{\pm 12.14}$	60.85$_{\pm 6.39}$

Table 11: Performance comparison of adaptation methods on 5-shot node classification.

Method	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	50.25$_{\pm 8.37}$	41.22$_{\pm 6.30}$	67.88$_{\pm 2.18}$	39.43$_{\pm 5.86}$	43.91$_{\pm 6.47}$	21.92$_{\pm 1.86}$	22.38$_{\pm 3.05}$
Fine-tuning	62.66$_{\pm 3.55}$	39.54$_{\pm 3.54}$	70.91$_{\pm 4.87}$	42.97$_{\pm 8.99}$	47.19$_{\pm 7.37}$	22.92$_{\pm 1.22}$	28.84$_{\pm 3.11}$
GPPT	51.98$_{\pm 3.43}$	45.77$_{\pm 7.41}$	66.97$_{\pm 3.70}$	37.00$_{\pm 3.19}$	48.82$_{\pm 5.15}$	21.58$_{\pm 0.84}$	28.90$_{\pm 1.64}$
Gprompt	69.03$_{\pm 3.61}$	66.13$_{\pm 1.64}$	67.87$_{\pm 2.08}$	78.22$_{\pm 37.33}$	39.32$_{\pm 47.08}$	34.67$_{\pm 1.28}$	85.40$_{\pm 0.79}$
All-in-one	30.36$_{\pm 13.48}$	27.93$_{\pm 10.59}$	46.16$_{\pm 15.83}$	87.16$_{\pm 3.02}$	73.28$_{\pm 9.91}$	21.49$_{\pm 3.02}$	13.01$_{\pm 6.29}$
GPF	35.43$_{\pm 1.02}$	25.12$_{\pm 3.01}$	68.96$_{\pm 3.99}$	98.26$_{\pm 1.19}$	98.42$_{\pm 0.36}$	44.07$_{\pm 3.94}$	71.83$_{\pm 9.37}$
GPF-plus	66.22$_{\pm 6.20}$	75.73$_{\pm 2.19}$	69.59$_{\pm 4.33}$	99.01$_{\pm 1.43}$	99.12$_{\pm 0.95}$	44.58$_{\pm 5.95}$	66.88$_{\pm 6.14}$
MTG (Ours)	71.81$_{\pm 3.59}$	76.34$_{\pm 6.18}$	70.84$_{\pm 3.28}$	99.12$_{\pm 0.95}$	98.76$_{\pm 2.36}$	45.09$_{\pm 3.26}$	85.94$_{\pm 1.93}$

Table 12: Performance comparison of adaptation methods on 5-shot graph classification.

Method	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	62.60$_{\pm 4.01}$	55.23$_{\pm 4.26}$	62.90$_{\pm 5.03}$	73.47$_{\pm 3.92}$	25.67$_{\pm 0.48}$	64.99$_{\pm 10.42}$	51.48$_{\pm 2.29}$	63.59$_{\pm 2.86}$
Fine-tuning	65.40$_{\pm 3.33}$	60.72$_{\pm 2.09}$	63.33$_{\pm 4.13}$	75.33$_{\pm 1.89}$	7.46$_{\pm 1.29}$	73.19$_{\pm 9.53}$	72.96$_{\pm 11.98}$	64.71$_{\pm 3.22}$
GPPT	66.37$_{\pm 3.59}$	54.05$_{\pm 4.58}$	58.27$_{\pm 4.63}$	70.53$_{\pm 3.90}$	22.17$_{\pm 2.34}$	67.88$_{\pm 17.34}$	69.63$_{\pm 14.96}$	60.02$_{\pm 3.24}$
Gprompt	66.70$_{\pm 3.87}$	60.76$_{\pm 5.08}$	62.94$_{\pm 1.38}$	73.07$_{\pm 2.13}$	21.46$_{\pm 2.27}$	53.35$_{\pm 7.75}$	59.38$_{\pm 14.43}$	58.28$_{\pm 2.18}$
All-in-one	63.62$_{\pm 2.30}$	57.86$_{\pm 5.88}$	71.37$_{\pm 4.89}$	80.93$_{\pm 1.96}$	26.71$_{\pm 2.17}$	62.95$_{\pm 8.57}$	62.78$_{\pm 10.18}$	63.44$_{\pm 1.35}$
GPF	67.80$_{\pm 5.58}$	59.65$_{\pm 6.25}$	63.37$_{\pm 4.37}$	74.00$_{\pm 3.65}$	27.00$_{\pm 0.78}$	66.27$_{\pm 14.57}$	61.05$_{\pm 11.51}$	61.06$_{\pm 2.63}$
GPF-plus	68.13$_{\pm 3.31}$	60.68$_{\pm 4.67}$	63.51$_{\pm 2.89}$	73.87$_{\pm 3.51}$	26.87$_{\pm 1.89}$	72.87$_{\pm 10.17}$	71.54$_{\pm 14.81}$	64.80$_{\pm 3.45}$
MTG (Ours)	69.15$_{\pm 4.09}$	63.11$_{\pm 1.88}$	70.10$_{\pm 1.12}$	81.60$_{\pm 4.53}$	35.08$_{\pm 3.28}$	71.84$_{\pm 2.75}$	76.37$_{\pm 8.11}$	66.07$_{\pm 2.39}$

Table 13: Performance comparison of adaptation methods on deep backbone models.

Method	Cora (1-shot)				BZR (1-shot)			
	$L = 4$	$L = 8$	$L = 12$	$L = 16$	$L = 4$	$L = 8$	$L = 12$	$L = 16$
Fine-tuning	38.88$_{\pm 6.74}$	36.84$_{\pm 4.10}$	33.78$_{\pm 6.05}$	30.70$_{\pm 4.18}$	70.06$_{\pm 18.37}$	56.17$_{\pm 28.60}$	67.41$_{\pm 23.52}$	71.11$_{\pm 15.81}$
GPPT	30.68$_{\pm 5.78}$	33.82$_{\pm 1.99}$	33.89$_{\pm 8.06}$	24.32$_{\pm 4.60}$	68.95$_{\pm 8.69}$	77.90$_{\pm 23.15}$	67.59$_{\pm 12.99}$	69.20$_{\pm 14.51}$
Gprompt	38.53$_{\pm 5.57}$	43.55$_{\pm 4.87}$	41.42$_{\pm 8.95}$	33.74$_{\pm 4.10}$	67.04$_{\pm 12.70}$	71.67$_{\pm 7.01}$	76.60$_{\pm 7.37}$	72.65$_{\pm 7.40}$
All-in-one	29.42$_{\pm 4.09}$	29.68$_{\pm 6.53}$	26.02$_{\pm 4.38}$	30.93$_{\pm 4.38}$	61.23$_{\pm 7.94}$	62.53$_{\pm 10.26}$	69.32$_{\pm 9.94}$	77.04$_{\pm 2.93}$
GPF	33.84$_{\pm 9.28}$	36.82$_{\pm 13.61}$	28.12$_{\pm 2.39}$	30.68$_{\pm 3.43}$	75.74$_{\pm 7.02}$	73.95$_{\pm 10.60}$	72.59$_{\pm 8.94}$	78.83$_{\pm 0.75}$
GPF-plus	43.67$_{\pm 9.52}$	41.34$_{\pm 6.52}$	39.23$_{\pm 7.88}$	36.08$_{\pm 5.54}$	73.70$_{\pm 3.14}$	76.79$_{\pm 1.08}$	75.86$_{\pm 6.77}$	71.73$_{\pm 9.99}$
MTG (Ours)	**47.80**$_{\pm 7.07}$	**45.10**$_{\pm 7.90}$	**43.36**$_{\pm 6.54}$	**37.00**$_{\pm 5.66}$	**77.04**$_{\pm 9.93}$	**79.26**$_{\pm 0.45}$	**79.26**$_{\pm 0.45}$	**79.26**$_{\pm 0.45}$

Table 14: Accuracy (%) on 1-shot node classification.

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	$26.56_{\pm 5.55}$	$21.78_{\pm 7.32}$	$39.37_{\pm 16.34}$	$41.60_{\pm 3.10}$	$37.97_{\pm 5.80}$	$20.57_{\pm 4.47}$	$10.99_{\pm 3.19}$
Fine-tuning	DGI	$33.15_{\pm 7.84}$	$21.64_{\pm 3.92}$	$42.01_{\pm 12.54}$	$37.49_{\pm 5.13}$	$45.31_{\pm 5.01}$	$19.76_{\pm 3.53}$	$7.21_{\pm 2.91}$
	GraphMAE	$32.93_{\pm 3.17}$	$21.26_{\pm 3.57}$	$42.99_{\pm 14.25}$	$36.80_{\pm 7.17}$	$37.81_{\pm 8.62}$	$19.86_{\pm 2.70}$	$12.35_{\pm 3.60}$
	EdgePreGPPT	$38.12_{\pm 5.29}$	$18.09_{\pm 5.39}$	$46.74_{\pm 14.09}$	$35.31_{\pm 9.31}$	$47.66_{\pm 2.37}$	$19.17_{\pm 2.53}$	$16.21_{\pm 3.82}$
	EdgePreGprompt	$35.57_{\pm 5.83}$	$22.28_{\pm 3.80}$	$41.50_{\pm 7.54}$	$40.69_{\pm 4.13}$	$40.62_{\pm 7.95}$	$20.74_{\pm 4.12}$	$14.83_{\pm 2.38}$
	GraphCL	$52.61_{\pm 1.73}$	$27.02_{\pm 4.31}$	$42.49_{\pm 11.29}$	$33.94_{\pm 7.74}$	$40.31_{\pm 13.68}$	$20.19_{\pm 1.98}$	$4.65_{\pm 1.19}$
	SimGRACE	$40.40_{\pm 4.66}$	$35.05_{\pm 4.37}$	$37.59_{\pm 8.17}$	$37.37_{\pm 3.68}$	$46.88_{\pm 4.64}$	$19.78_{\pm 1.89}$	$8.13_{\pm 3.26}$
GPPTPrompt	DGI	$30.47_{\pm 3.53}$	$37.26_{\pm 6.17}$	$35.62_{\pm 7.74}$	$29.94_{\pm 10.40}$	$29.29_{\pm 14.57}$	$21.76_{\pm 2.00}$	$3.80_{\pm 6.19}$
	GraphMAE	$27.39_{\pm 10.26}$	$21.54_{\pm 4.00}$	$48.31_{\pm 17.72}$	$29.83_{\pm 9.34}$	$25.04_{\pm 10.38}$	$22.58_{\pm 1.97}$	$4.35_{\pm 6.01}$
	EdgePreGPPT	$30.37_{\pm 4.30}$	$21.06_{\pm 4.37}$	$39.64_{\pm 7.64}$	$23.89_{\pm 5.40}$	$30.39_{\pm 8.96}$	$19.85_{\pm 0.76}$	$14.65_{\pm 3.07}$
	EdgePreGprompt	$25.52_{\pm 4.42}$	$21.85_{\pm 4.30}$	$46.20_{\pm 10.76}$	$30.40_{\pm 6.81}$	$22.68_{\pm 12.82}$	$21.52_{\pm 1.13}$	$2.05_{\pm 1.43}$
	GraphCL	$43.15_{\pm 9.44}$	$26.73_{\pm 4.12}$	$38.34_{\pm 11.59}$	$25.03_{\pm 5.37}$	$31.81_{\pm 15.33}$	$22.51_{\pm 1.73}$	$7.15_{\pm 4.12}$
	SimGRACE	$27.86_{\pm 2.79}$	$25.06_{\pm 4.90}$	$36.70_{\pm 9.26}$	$29.83_{\pm 6.44}$	$25.67_{\pm 8.01}$	$20.97_{\pm 2.30}$	$5.50_{\pm 5.10}$
Gprompt	DGI	$36.46_{\pm 5.39}$	$36.25_{\pm 10.26}$	$33.65_{\pm 5.29}$	$67.71_{\pm 9.92}$	$31.00_{\pm 37.32}$	$23.85_{\pm 3.52}$	$48.47_{\pm 5.87}$
	GraphMAE	$50.58_{\pm 7.34}$	$42.84_{\pm 10.59}$	$39.74_{\pm 15.35}$	$67.62_{\pm 18.06}$	$23.31_{\pm 29.49}$	$22.34_{\pm 3.57}$	$66.66_{\pm 5.04}$
	EdgePreGPPT	$46.96_{\pm 6.22}$	$40.15_{\pm 7.04}$	$35.46_{\pm 14.12}$	$67.37_{\pm 12.32}$	$30.52_{\pm 36.73}$	$23.50_{\pm 4.16}$	$75.72_{\pm 4.95}$
	EdgePreGprompt	$48.11_{\pm 9.89}$	$48.07_{\pm 5.62}$	$33.54_{\pm 16.66}$	$74.38_{\pm 13.15}$	$33.25_{\pm 40.11}$	$19.89_{\pm 1.38}$	$70.55_{\pm 7.66}$
	GraphCL	$56.66_{\pm 11.22}$	$45.81_{\pm 7.04}$	$39.37_{\pm 14.95}$	$77.07_{\pm 5.93}$	$29.55_{\pm 35.56}$	$25.26_{\pm 1.10}$	$51.20_{\pm 6.40}$
	SimGRACE	$46.34_{\pm 6.75}$	$53.21_{\pm 10.94}$	$35.58_{\pm 9.03}$	$65.38_{\pm 13.70}$	$30.20_{\pm 36.49}$	$24.49_{\pm 4.38}$	$52.76_{\pm 5.30}$
All-in-one	DGI	$47.52_{\pm 2.50}$	$39.37_{\pm 3.12}$	$38.74_{\pm 2.15}$	$56.02_{\pm 13.12}$	$57.38_{\pm 12.82}$	$21.03_{\pm 1.96}$	$1.93_{\pm 1.48}$
	GraphMAE	$23.09_{\pm 4.92}$	$18.08_{\pm 5.23}$	$33.19_{\pm 11.98}$	$57.54_{\pm 10.66}$	$52.82_{\pm 11.47}$	$23.31_{\pm 2.01}$	$4.19_{\pm 6.04}$
	EdgePreGPPT	$49.63_{\pm 5.26}$	$35.06_{\pm 2.37}$	$40.73_{\pm 11.32}$	$66.29_{\pm 19.11}$	$58.62_{\pm 5.54}$	$21.49_{\pm 1.27}$	$5.62_{\pm 3.95}$
	EdgePreGprompt	$37.39_{\pm 3.31}$	$28.85_{\pm 4.32}$	$35.53_{\pm 9.07}$	$59.18_{\pm 12.30}$	$39.71_{\pm 25.31}$	$20.49_{\pm 3.90}$	$13.01_{\pm 6.29}$
	GraphCL	$52.39_{\pm 10.17}$	$37.37_{\pm 4.15}$	$45.17_{\pm 6.93}$	$39.14_{\pm 1.17}$	$65.49_{\pm 7.06}$	$24.61_{\pm 2.80}$	$6.70_{\pm 6.01}$
	SimGRACE	$35.99_{\pm 2.76}$	$40.41_{\pm 2.80}$	$30.23_{\pm 7.03}$	$55.56_{\pm 14.70}$	$59.22_{\pm 20.17}$	$21.03_{\pm 2.23}$	$5.72_{\pm 1.61}$
GPF	DGI	$27.83_{\pm 18.89}$	$16.50_{\pm 4.57}$	$38.33_{\pm 8.13}$	$62.69_{\pm 13.96}$	$60.54_{\pm 13.13}$	$28.17_{\pm 4.81}$	$19.36_{\pm 6.20}$
	GraphMAE	$38.57_{\pm 5.41}$	$25.61_{\pm 3.27}$	$48.52_{\pm 13.23}$	$76.84_{\pm 10.50}$	$69.51_{\pm 18.75}$	$28.37_{\pm 5.82}$	$43.28_{\pm 10.60}$
	EdgePreGPPT	$15.29_{\pm 8.41}$	$12.33_{\pm 5.33}$	$43.78_{\pm 6.02}$	$78.35_{\pm 4.07}$	$68.05_{\pm 17.34}$	$25.66_{\pm 3.33}$	$65.11_{\pm 5.70}$
	EdgePreGprompt	$26.60_{\pm 13.92}$	$31.16_{\pm 8.05}$	$48.98_{\pm 11.57}$	$75.20_{\pm 13.22}$	$69.48_{\pm 17.07}$	$25.27_{\pm 5.65}$	$41.87_{\pm 11.49}$
	GraphCL	$23.16_{\pm 5.11}$	$16.77_{\pm 1.39}$	$49.99_{\pm 8.86}$	$51.60_{\pm 20.06}$	$73.54_{\pm 18.50}$	$20.68_{\pm 6.70}$	$27.73_{\pm 5.12}$
	SimGRACE	$32.01_{\pm 11.21}$	$19.43_{\pm 2.10}$	$37.27_{\pm 6.09}$	$60.81_{\pm 26.52}$	$69.97_{\pm 16.76}$	$28.70_{\pm 3.35}$	$25.12_{\pm 4.50}$
GPF-plus	DGI	$17.29_{\pm 6.18}$	$26.60_{\pm 13.24}$	$34.02_{\pm 11.94}$	$74.68_{\pm 11.81}$	$71.44_{\pm 18.66}$	$22.42_{\pm 9.66}$	$16.83_{\pm 10.02}$
	GraphMAE	$54.26_{\pm 7.48}$	$59.67_{\pm 11.87}$	$46.64_{\pm 18.57}$	$82.11_{\pm 13.95}$	$70.95_{\pm 18.63}$	$26.58_{\pm 7.84}$	$49.81_{\pm 2.62}$
	EdgePreGPPT	$28.49_{\pm 18.73}$	$28.04_{\pm 14.31}$	$46.51_{\pm 15.84}$	$72.66_{\pm 12.05}$	$70.67_{\pm 17.59}$	$29.32_{\pm 8.56}$	$71.98_{\pm 12.23}$
	EdgePreGprompt	$55.77_{\pm 10.30}$	$49.43_{\pm 8.21}$	$42.79_{\pm 18.18}$	$78.76_{\pm 13.63}$	$68.75_{\pm 16.51}$	$22.68_{\pm 3.64}$	$57.44_{\pm 6.95}$
	GraphCL	$34.18_{\pm 17.71}$	$28.86_{\pm 22.88}$	$37.02_{\pm 11.29}$	$52.35_{\pm 19.69}$	$75.40_{\pm 19.10}$	$22.82_{\pm 4.99}$	$32.11_{\pm 4.86}$
	SimGRACE	$21.33_{\pm 14.86}$	$24.61_{\pm 21.21}$	$35.90_{\pm 9.06}$	$73.49_{\pm 14.17}$	$76.10_{\pm 20.35}$	$20.51_{\pm 4.24}$	$46.71_{\pm 3.17}$
MTG (Ours)	DGI	$49.48_{\pm 4.82}$	$62.31_{\pm 18.90}$	$46.18_{\pm 7.32}$	$67.72_{\pm 10.19}$	$62.96_{\pm 16.80}$	$25.48_{\pm 7.33}$	$25.06_{\pm 10.57}$
	GraphMAE	$46.27_{\pm 6.66}$	$49.21_{\pm 12.95}$	$46.98_{\pm 10.02}$	$83.32_{\pm 12.46}$	$71.59_{\pm 18.67}$	$29.44_{\pm 7.31}$	$36.44_{\pm 9.59}$
	EdgePreGPPT	$46.68_{\pm 2.66}$	$33.22_{\pm 12.52}$	$44.85_{\pm 9.75}$	$73.80_{\pm 9.56}$	$71.11_{\pm 17.13}$	$20.96_{\pm 2.93}$	$75.97_{\pm 4.29}$
	EdgePreGprompt	$46.29_{\pm 3.84}$	$45.30_{\pm 16.04}$	$50.70_{\pm 11.68}$	$72.75_{\pm 11.21}$	$79.13_{\pm 17.18}$	$21.34_{\pm 1.78}$	$21.08_{\pm 2.34}$
	GraphCL	$58.54_{\pm 7.89}$	$50.96_{\pm 16.40}$	$40.00_{\pm 7.80}$	$48.41_{\pm 16.10}$	$69.71_{\pm 16.42}$	$24.77_{\pm 8.45}$	$38.96_{\pm 6.82}$
	SimGRACE	$45.93_{\pm 7.67}$	$57.60_{\pm 9.01}$	$43.29_{\pm 10.80}$	$72.98_{\pm 9.75}$	$71.26_{\pm 17.71}$	$22.03_{\pm 3.59}$	$37.90_{\pm 5.83}$

Table 15: F1-score on 1-shot node classification.

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	$16.60_{\pm 2.54}$	$10.81_{\pm 4.90}$	$37.23_{\pm 15.48}$	$26.34_{\pm 4.01}$	$24.05_{\pm 5.12}$	$11.56_{\pm 3.08}$	$7.99_{\pm 0.99}$
Fine-tuning	DGI	$24.96_{\pm 6.01}$	$11.01_{\pm 5.94}$	$34.75_{\pm 13.75}$	$26.69_{\pm 3.39}$	$28.90_{\pm 6.81}$	$12.01_{\pm 1.43}$	$3.93_{\pm 1.32}$
	GraphMAE	$23.18_{\pm 2.85}$	$10.82_{\pm 3.83}$	$41.03_{\pm 13.36}$	$27.43_{\pm 4.47}$	$23.08_{\pm 6.07}$	$12.71_{\pm 1.24}$	$8.07_{\pm 1.08}$
	EdgePreGPPT	$35.92_{\pm 4.06}$	$10.86_{\pm 2.29}$	$40.62_{\pm 18.09}$	$23.56_{\pm 3.43}$	$29.03_{\pm 5.16}$	$14.59_{\pm 3.09}$	$12.13_{\pm 2.04}$
	EdgePreGprompt	$28.99_{\pm 6.35}$	$12.39_{\pm 3.56}$	$28.89_{\pm 6.74}$	$26.74_{\pm 3.28}$	$26.81_{\pm 6.66}$	$11.63_{\pm 2.64}$	$10.61_{\pm 1.45}$
	GraphCL	$47.14_{\pm 3.15}$	$21.86_{\pm 3.67}$	$38.30_{\pm 10.89}$	$14.27_{\pm 3.90}$	$24.52_{\pm 7.59}$	$15.91_{\pm 0.98}$	$2.88_{\pm 0.51}$
	SimGRACE	$33.22_{\pm 3.29}$	$30.78_{\pm 3.91}$	$32.79_{\pm 6.60}$	$23.71_{\pm 2.97}$	$29.53_{\pm 6.44}$	$15.73_{\pm 1.20}$	$2.88_{\pm 0.50}$
GPPTPrompt	DGI	$25.82_{\pm 4.78}$	$33.00_{\pm 6.49}$	$31.92_{\pm 8.94}$	$22.77_{\pm 6.29}$	$23.80_{\pm 8.89}$	$17.59_{\pm 1.13}$	$0.17_{\pm 0.26}$
	GraphMAE	$13.18_{\pm 6.02}$	$10.87_{\pm 4.43}$	$46.43_{\pm 16.73}$	$23.74_{\pm 5.95}$	$21.36_{\pm 7.80}$	$13.60_{\pm 1.69}$	$0.27_{\pm 0.23}$
	EdgePreGPPT	$28.54_{\pm 3.87}$	$17.62_{\pm 4.24}$	$34.55_{\pm 8.09}$	$20.61_{\pm 5.49}$	$23.84_{\pm 5.22}$	$18.48_{\pm 0.57}$	$9.15_{\pm 1.18}$
	EdgePreGprompt	$23.46_{\pm 5.11}$	$19.00_{\pm 4.24}$	$45.52_{\pm 10.54}$	$23.65_{\pm 3.96}$	$19.82_{\pm 7.93}$	$19.39_{\pm 1.08}$	$1.35_{\pm 1.07}$
	GraphCL	$38.99_{\pm 8.32}$	$23.76_{\pm 3.97}$	$36.75_{\pm 12.92}$	$18.08_{\pm 6.69}$	$25.64_{\pm 8.12}$	$19.62_{\pm 0.56}$	$1.52_{\pm 0.68}$
	SimGRACE	$21.70_{\pm 2.66}$	$22.13_{\pm 4.48}$	$31.55_{\pm 10.44}$	$21.48_{\pm 3.16}$	$21.75_{\pm 5.75}$	$17.15_{\pm 1.75}$	$0.66_{\pm 0.39}$
Gprompt	DGI	$30.20_{\pm 4.21}$	$33.58_{\pm 9.40}$	$31.89_{\pm 5.43}$	$55.65_{\pm 4.81}$	$23.68_{\pm 28.78}$	$20.12_{\pm 3.96}$	$42.90_{\pm 1.91}$
	GraphMAE	$45.91_{\pm 6.10}$	$40.94_{\pm 10.71}$	$39.46_{\pm 15.97}$	$60.14_{\pm 10.06}$	$19.27_{\pm 23.52}$	$20.06_{\pm 4.69}$	$57.88_{\pm 3.32}$
	EdgePreGPPT	$44.15_{\pm 7.57}$	$37.31_{\pm 6.26}$	$29.87_{\pm 12.41}$	$61.52_{\pm 6.78}$	$24.14_{\pm 29.42}$	$20.14_{\pm 3.34}$	$69.86_{\pm 1.92}$
	EdgePreGprompt	$46.28_{\pm 8.46}$	$42.82_{\pm 6.05}$	$33.57_{\pm 16.07}$	$64.46_{\pm 10.07}$	$25.86_{\pm 31.46}$	$18.67_{\pm 1.74}$	$69.89_{\pm 5.15}$
	GraphCL	$49.86_{\pm 10.36}$	$40.41_{\pm 9.12}$	$38.04_{\pm 13.45}$	$61.79_{\pm 6.30}$	$21.15_{\pm 25.84}$	$22.00_{\pm 1.74}$	$45.32_{\pm 3.89}$
	SimGRACE	$38.55_{\pm 5.02}$	$49.65_{\pm 11.42}$	$31.13_{\pm 9.15}$	$60.40_{\pm 7.58}$	$29.20_{\pm 35.62}$	$21.39_{\pm 3.95}$	$46.73_{\pm 4.62}$
All-in-one	DGI	$34.76_{\pm 3.89}$	$26.67_{\pm 4.46}$	$22.68_{\pm 5.29}$	$46.19_{\pm 6.62}$	$31.14_{\pm 20.37}$	$14.37_{\pm 2.36}$	$0.24_{\pm 0.21}$
	GraphMAE	$10.82_{\pm 4.45}$	$7.04_{\pm 2.29}$	$25.46_{\pm 9.17}$	$52.55_{\pm 4.85}$	$37.08_{\pm 9.37}$	$12.68_{\pm 2.63}$	$0.19_{\pm 0.26}$
	EdgePreGPPT	$44.16_{\pm 4.02}$	$26.79_{\pm 3.27}$	$36.27_{\pm 12.89}$	$57.44_{\pm 10.67}$	$26.56_{\pm 10.81}$	$14.72_{\pm 3.29}$	$0.98_{\pm 0.47}$
	EdgePreGprompt	$24.93_{\pm 4.99}$	$14.58_{\pm 5.03}$	$30.99_{\pm 7.62}$	$51.42_{\pm 6.35}$	$28.81_{\pm 17.76}$	$11.85_{\pm 1.48}$	$0.56_{\pm 0.27}$
	GraphCL	$46.58_{\pm 8.42}$	$29.35_{\pm 3.66}$	$38.05_{\pm 6.24}$	$34.06_{\pm 7.00}$	$43.37_{\pm 16.01}$	$16.05_{\pm 3.88}$	$0.75_{\pm 1.07}$
	SimGRACE	$27.35_{\pm 1.31}$	$30.20_{\pm 4.44}$	$24.61_{\pm 6.29}$	$49.19_{\pm 8.83}$	$39.01_{\pm 18.76}$	$13.86_{\pm 1.76}$	$1.28_{\pm 0.35}$
GPF	DGI	$19.39_{\pm 17.53}$	$4.68_{\pm 1.17}$	$28.70_{\pm 9.13}$	$54.29_{\pm 8.67}$	$51.28_{\pm 7.80}$	$19.25_{\pm 10.07}$	$8.47_{\pm 11.46}$
	GraphMAE	$23.79_{\pm 5.49}$	$12.89_{\pm 1.99}$	$45.36_{\pm 15.88}$	$69.67_{\pm 7.97}$	$63.67_{\pm 13.52}$	$21.69_{\pm 5.47}$	$19.48_{\pm 2.04}$
	EdgePreGPPT	$3.76_{\pm 1.75}$	$3.59_{\pm 1.38}$	$31.20_{\pm 15.26}$	$66.57_{\pm 6.08}$	$60.39_{\pm 13.78}$	$18.04_{\pm 8.30}$	$47.05_{\pm 3.15}$
	EdgePreGprompt	$15.85_{\pm 13.63}$	$18.63_{\pm 7.34}$	$39.90_{\pm 9.00}$	$68.49_{\pm 10.01}$	$63.33_{\pm 14.58}$	$19.72_{\pm 8.08}$	$21.67_{\pm 2.10}$
	GraphCL	$9.76_{\pm 4.57}$	$4.78_{\pm 0.34}$	$38.27_{\pm 17.17}$	$35.33_{\pm 13.99}$	$63.11_{\pm 7.98}$	$11.69_{\pm 9.39}$	$19.44_{\pm 2.52}$
	SimGRACE	$18.73_{\pm 10.08}$	$5.90_{\pm 1.13}$	$22.65_{\pm 5.69}$	$55.15_{\pm 21.36}$	$60.62_{\pm 9.24}$	$17.94_{\pm 5.36}$	$13.14_{\pm 2.90}$
GPF-plus	DGI	$9.94_{\pm 8.40}$	$13.74_{\pm 16.38}$	$24.21_{\pm 9.44}$	$68.32_{\pm 9.85}$	$60.71_{\pm 8.30}$	$14.51_{\pm 10.72}$	$5.37_{\pm 4.20}$
	GraphMAE	$50.19_{\pm 10.64}$	$56.22_{\pm 13.99}$	$42.38_{\pm 19.01}$	$75.07_{\pm 10.10}$	$64.08_{\pm 12.43}$	$19.25_{\pm 5.53}$	$23.38_{\pm 1.12}$
	EdgePreGPPT	$19.52_{\pm 17.74}$	$15.32_{\pm 19.12}$	$40.64_{\pm 17.93}$	$64.82_{\pm 11.61}$	$64.09_{\pm 9.70}$	$24.56_{\pm 8.79}$	$60.35_{\pm 7.73}$
	EdgePreGprompt	$53.28_{\pm 11.46}$	$48.37_{\pm 9.28}$	$40.02_{\pm 18.18}$	$71.73_{\pm 9.24}$	$58.51_{\pm 6.37}$	$15.58_{\pm 4.14}$	$31.37_{\pm 2.50}$
	GraphCL	$26.14_{\pm 21.11}$	$18.90_{\pm 24.01}$	$33.64_{\pm 11.41}$	$39.25_{\pm 18.65}$	$70.08_{\pm 15.87}$	$16.08_{\pm 3.05}$	$15.78_{\pm 1.60}$
	SimGRACE	$14.35_{\pm 19.65}$	$15.23_{\pm 20.93}$	$23.74_{\pm 12.56}$	$52.54_{\pm 12.12}$	$68.07_{\pm 10.86}$	$11.83_{\pm 4.68}$	$21.81_{\pm 1.06}$
MTG (Ours)	DGI	$44.86_{\pm 3.11}$	$55.19_{\pm 16.42}$	$45.56_{\pm 7.25}$	$63.66_{\pm 4.96}$	$54.31_{\pm 10.00}$	$23.43_{\pm 6.22}$	$9.08_{\pm 2.89}$
	GraphMAE	$42.83_{\pm 5.19}$	$44.41_{\pm 12.94}$	$44.53_{\pm 10.81}$	$76.12_{\pm 9.17}$	$65.26_{\pm 13.22}$	$27.54_{\pm 7.26}$	$15.27_{\pm 2.27}$
	EdgePreGPPT	$44.73_{\pm 3.41}$	$30.95_{\pm 12.96}$	$42.03_{\pm 11.12}$	$64.28_{\pm 9.95}$	$65.11_{\pm 11.93}$	$19.88_{\pm 2.86}$	$62.60_{\pm 2.61}$
	EdgePreGprompt	$44.91_{\pm 3.76}$	$41.03_{\pm 15.51}$	$47.73_{\pm 8.47}$	$66.38_{\pm 7.19}$	$71.33_{\pm 13.75}$	$18.47_{\pm 1.76}$	$11.31_{\pm 1.49}$
	GraphCL	$55.75_{\pm 3.45}$	$46.41_{\pm 16.57}$	$38.30_{\pm 8.83}$	$38.52_{\pm 8.45}$	$61.44_{\pm 7.07}$	$23.92_{\pm 8.32}$	$24.83_{\pm 3.25}$
	SimGRACE	$40.42_{\pm 4.74}$	$54.55_{\pm 8.97}$	$36.31_{\pm 6.58}$	$65.38_{\pm 8.40}$	$63.36_{\pm 6.90}$	$21.77_{\pm 3.92}$	$23.22_{\pm 4.69}$

Table 16: AUROC on 1-shot node classification.

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	76.71$_{\pm1.76}$	63.02$_{\pm4.09}$	62.01$_{\pm18.06}$	61.53$_{\pm4.96}$	62.86$_{\pm10.22}$	51.16$_{\pm1.01}$	72.48$_{\pm1.34}$
Fine-tuning	DGI	79.33$_{\pm2.82}$	65.16$_{\pm4.85}$	63.47$_{\pm13.36}$	59.72$_{\pm2.84}$	65.16$_{\pm9.96}$	50.13$_{\pm0.96}$	64.22$_{\pm2.71}$
	GraphMAE	78.93$_{\pm0.67}$	63.92$_{\pm3.46}$	66.78$_{\pm16.67}$	60.83$_{\pm5.18}$	61.79$_{\pm9.63}$	51.01$_{\pm0.55}$	72.86$_{\pm1.61}$
	EdgePreGPPT	78.44$_{\pm1.58}$	60.74$_{\pm3.98}$	63.49$_{\pm15.75}$	58.91$_{\pm3.16}$	65.05$_{\pm9.01}$	50.79$_{\pm0.65}$	77.26$_{\pm2.75}$
	EdgePreGprompt	80.12$_{\pm2.93}$	65.05$_{\pm4.65}$	60.89$_{\pm13.56}$	61.08$_{\pm5.22}$	62.77$_{\pm10.98}$	51.47$_{\pm1.28}$	75.81$_{\pm2.26}$
	GraphCL	86.34$_{\pm1.24}$	68.32$_{\pm1.24}$	57.56$_{\pm15.85}$	43.79$_{\pm4.25}$	63.09$_{\pm10.20}$	51.19$_{\pm0.63}$	63.36$_{\pm1.49}$
	SimGRACE	76.63$_{\pm1.67}$	71.91$_{\pm2.47}$	58.10$_{\pm9.13}$	57.76$_{\pm2.12}$	65.06$_{\pm9.74}$	50.81$_{\pm0.66}$	56.33$_{\pm0.77}$
GPPTPrompt	DGI	63.97$_{\pm5.40}$	71.52$_{\pm5.15}$	51.57$_{\pm5.86}$	59.89$_{\pm2.95}$	57.22$_{\pm9.90}$	50.71$_{\pm0.86}$	50.07$_{\pm0.08}$
	GraphMAE	78.20$_{\pm3.89}$	61.64$_{\pm4.31}$	64.12$_{\pm15.44}$	55.48$_{\pm6.44}$	53.95$_{\pm3.70}$	51.71$_{\pm1.74}$	59.28$_{\pm2.39}$
	EdgePreGPPT	65.75$_{\pm4.04}$	54.28$_{\pm5.62}$	55.12$_{\pm11.88}$	54.59$_{\pm4.65}$	50.55$_{\pm3.83}$	49.94$_{\pm0.53}$	74.38$_{\pm1.59}$
	EdgePreGprompt	62.60$_{\pm2.50}$	54.23$_{\pm6.14}$	61.41$_{\pm11.42}$	57.58$_{\pm2.23}$	53.16$_{\pm6.75}$	50.32$_{\pm0.29}$	63.71$_{\pm2.13}$
	GraphCL	72.09$_{\pm7.61}$	59.06$_{\pm3.51}$	55.95$_{\pm13.50}$	54.51$_{\pm4.05}$	56.59$_{\pm7.31}$	51.76$_{\pm0.95}$	54.45$_{\pm1.08}$
	SimGRACE	64.83$_{\pm2.80}$	59.61$_{\pm4.19}$	51.84$_{\pm12.98}$	56.88$_{\pm4.39}$	53.42$_{\pm6.67}$	50.48$_{\pm0.77}$	50.72$_{\pm0.43}$
Gprompt	DGI	70.70$_{\pm2.26}$	70.93$_{\pm4.80}$	52.96$_{\pm2.53}$	74.12$_{\pm8.50}$	57.93$_{\pm10.19}$	52.82$_{\pm3.88}$	90.60$_{\pm1.08}$
	GraphMAE	80.67$_{\pm5.24}$	70.72$_{\pm5.34}$	60.66$_{\pm19.63}$	89.64$_{\pm5.51}$	59.57$_{\pm11.81}$	52.87$_{\pm4.78}$	94.04$_{\pm1.32}$
	EdgePreGPPT	84.03$_{\pm2.26}$	67.95$_{\pm2.69}$	44.60$_{\pm13.08}$	88.97$_{\pm5.38}$	60.65$_{\pm12.95}$	53.55$_{\pm3.61}$	96.40$_{\pm0.74}$
	EdgePreGprompt	81.90$_{\pm4.04}$	74.85$_{\pm2.68}$	58.34$_{\pm22.51}$	91.99$_{\pm5.59}$	62.03$_{\pm14.44}$	50.16$_{\pm1.75}$	94.39$_{\pm1.12}$
	GraphCL	83.03$_{\pm4.16}$	78.33$_{\pm5.28}$	58.24$_{\pm13.76}$	87.80$_{\pm5.54}$	58.33$_{\pm10.63}$	54.45$_{\pm2.93}$	92.72$_{\pm0.72}$
	SimGRACE	76.99$_{\pm3.17}$	84.09$_{\pm2.88}$	51.91$_{\pm12.26}$	87.53$_{\pm2.42}$	61.60$_{\pm14.00}$	53.63$_{\pm4.60}$	91.93$_{\pm1.25}$
All-in-one	DGI	75.29$_{\pm1.07}$	71.50$_{\pm1.46}$	55.69$_{\pm1.38}$	74.35$_{\pm4.22}$	64.75$_{\pm6.31}$	51.05$_{\pm0.96}$	55.21$_{\pm1.45}$
	GraphMAE	73.99$_{\pm5.63}$	49.98$_{\pm3.07}$	53.18$_{\pm13.11}$	78.16$_{\pm4.96}$	62.90$_{\pm3.67}$	50.96$_{\pm1.97}$	50.21$_{\pm1.01}$
	EdgePreGPPT	83.26$_{\pm1.22}$	69.82$_{\pm1.32}$	73.63$_{\pm3.06}$	86.56$_{\pm6.53}$	75.38$_{\pm10.13}$	49.99$_{\pm0.28}$	54.22$_{\pm5.18}$
	EdgePreGprompt	73.84$_{\pm3.01}$	62.35$_{\pm6.11}$	61.26$_{\pm7.66}$	86.03$_{\pm3.16}$	67.99$_{\pm6.33}$	51.48$_{\pm2.87}$	75.81$_{\pm2.83}$
	GraphCL	84.34$_{\pm4.32}$	72.36$_{\pm2.82}$	68.22$_{\pm6.38}$	63.00$_{\pm6.09}$	69.39$_{\pm2.56}$	47.63$_{\pm2.10}$	55.77$_{\pm4.78}$
	SimGRACE	72.30$_{\pm1.28}$	69.88$_{\pm0.83}$	52.94$_{\pm2.50}$	76.60$_{\pm6.08}$	70.92$_{\pm7.19}$	48.18$_{\pm0.97}$	65.43$_{\pm1.59}$
GPF	DGI	65.74$_{\pm10.40}$	50.00$_{\pm0.00}$	51.90$_{\pm7.57}$	77.18$_{\pm4.67}$	67.01$_{\pm10.83}$	58.84$_{\pm3.16}$	64.47$_{\pm2.93}$
	GraphMAE	73.69$_{\pm2.75}$	66.16$_{\pm0.22}$	73.35$_{\pm8.02}$	96.98$_{\pm0.89}$	81.43$_{\pm10.22}$	55.43$_{\pm3.58}$	80.52$_{\pm0.81}$
	EdgePreGPPT	62.03$_{\pm6.87}$	57.99$_{\pm6.03}$	68.25$_{\pm7.09}$	95.97$_{\pm1.75}$	78.79$_{\pm9.70}$	55.90$_{\pm6.20}$	92.37$_{\pm0.30}$
	EdgePreGprompt	70.83$_{\pm8.42}$	71.94$_{\pm3.31}$	71.40$_{\pm8.27}$	90.33$_{\pm4.63}$	71.85$_{\pm9.72}$	56.13$_{\pm4.33}$	83.64$_{\pm2.82}$
	GraphCL	74.17$_{\pm5.66}$	61.21$_{\pm7.77}$	70.39$_{\pm4.76}$	65.12$_{\pm13.92}$	84.28$_{\pm8.08}$	54.42$_{\pm3.73}$	78.19$_{\pm1.35}$
	SimGRACE	69.60$_{\pm10.54}$	62.47$_{\pm3.70}$	50.73$_{\pm3.64}$	84.37$_{\pm12.04}$	79.69$_{\pm14.05}$	60.48$_{\pm2.78}$	61.04$_{\pm5.33}$
GPF-plus	DGI	58.07$_{\pm8.25}$	55.15$_{\pm11.69}$	55.70$_{\pm11.04}$	85.93$_{\pm6.24}$	79.28$_{\pm11.50}$	57.02$_{\pm8.99}$	60.73$_{\pm3.11}$
	GraphMAE	86.94$_{\pm4.26}$	87.43$_{\pm3.46}$	66.74$_{\pm18.62}$	97.98$_{\pm1.67}$	82.16$_{\pm10.09}$	63.32$_{\pm6.43}$	80.32$_{\pm3.00}$
	EdgePreGPPT	68.27$_{\pm10.69}$	59.85$_{\pm17.18}$	71.27$_{\pm10.40}$	92.67$_{\pm2.31}$	84.16$_{\pm10.73}$	61.72$_{\pm5.16}$	93.52$_{\pm1.09}$
	EdgePreGprompt	86.33$_{\pm4.94}$	88.60$_{\pm5.18}$	63.12$_{\pm18.73}$	96.27$_{\pm1.06}$	77.71$_{\pm2.31}$	61.97$_{\pm5.10}$	84.36$_{\pm0.87}$
	GraphCL	80.06$_{\pm6.24}$	61.58$_{\pm14.77}$	55.63$_{\pm15.68}$	75.26$_{\pm13.27}$	78.78$_{\pm1.50}$	58.86$_{\pm3.80}$	70.39$_{\pm3.22}$
	SimGRACE	64.39$_{\pm11.25}$	64.11$_{\pm10.53}$	54.79$_{\pm10.68}$	85.74$_{\pm6.67}$	85.07$_{\pm9.21}$	55.75$_{\pm4.51}$	69.09$_{\pm2.34}$
MTG (Ours)	DGI	82.30$_{\pm2.14}$	80.39$_{\pm9.20}$	62.64$_{\pm7.04}$	85.68$_{\pm4.00}$	73.62$_{\pm15.07}$	54.59$_{\pm5.10}$	74.20$_{\pm0.84}$
	GraphMAE	81.64$_{\pm4.86}$	71.84$_{\pm7.49}$	62.95$_{\pm11.20}$	98.24$_{\pm1.37}$	82.90$_{\pm11.67}$	61.70$_{\pm4.93}$	80.62$_{\pm2.62}$
	EdgePreGPPT	81.32$_{\pm3.49}$	65.29$_{\pm11.10}$	63.16$_{\pm10.25}$	93.74$_{\pm2.95}$	79.15$_{\pm6.07}$	50.47$_{\pm2.34}$	97.04$_{\pm0.59}$
	EdgePreGprompt	80.68$_{\pm2.31}$	73.35$_{\pm11.93}$	69.20$_{\pm13.88}$	90.19$_{\pm5.78}$	85.59$_{\pm9.47}$	51.89$_{\pm2.08}$	76.91$_{\pm3.05}$
	GraphCL	89.02$_{\pm2.73}$	78.61$_{\pm10.46}$	54.43$_{\pm11.25}$	75.82$_{\pm4.79}$	76.98$_{\pm9.39}$	55.56$_{\pm7.35}$	77.30$_{\pm2.70}$
	SimGRACE	78.41$_{\pm3.42}$	80.86$_{\pm8.79}$	57.61$_{\pm6.59}$	91.06$_{\pm1.60}$	82.24$_{\pm11.74}$	53.99$_{\pm2.42}$	63.13$_{\pm3.17}$

Table 17: Accuracy (%) on 3-shot node classification.

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	37.79$_{\pm9.16}$	35.18$_{\pm6.86}$	57.33$_{\pm4.64}$	41.03$_{\pm6.40}$	40.78$_{\pm12.55}$	18.62$_{\pm3.46}$	19.03$_{\pm5.08}$
Fine-tuning	DGI	45.84$_{\pm3.29}$	34.91$_{\pm10.07}$	63.00$_{\pm6.83}$	39.43$_{\pm6.93}$	43.13$_{\pm13.79}$	20.27$_{\pm1.70}$	21.85$_{\pm5.37}$
	GraphMAE	45.15$_{\pm4.77}$	28.59$_{\pm6.41}$	65.40$_{\pm3.00}$	41.49$_{\pm5.19}$	40.94$_{\pm13.91}$	19.03$_{\pm3.17}$	19.63$_{\pm7.64}$
	EdgePreGPPT	51.97$_{\pm2.84}$	33.19$_{\pm6.79}$	64.38$_{\pm3.59}$	42.40$_{\pm7.77}$	34.69$_{\pm12.07}$	20.83$_{\pm2.24}$	20.51$_{\pm4.02}$
	EdgePredGprompt	40.33$_{\pm6.62}$	33.39$_{\pm5.51}$	63.49$_{\pm3.17}$	41.37$_{\pm7.28}$	34.53$_{\pm10.09}$	19.29$_{\pm1.95}$	27.34$_{\pm6.61}$
	GraphCL	49.39$_{\pm9.15}$	38.40$_{\pm3.06}$	62.79$_{\pm3.21}$	41.26$_{\pm6.26}$	40.31$_{\pm13.36}$	21.06$_{\pm1.27}$	17.37$_{\pm7.34}$
	SimGRACE	43.61$_{\pm5.41}$	45.08$_{\pm2.09}$	62.66$_{\pm3.21}$	40.11$_{\pm8.04}$	40.94$_{\pm13.73}$	22.11$_{\pm1.97}$	16.06$_{\pm2.05}$
GPPTPrompt	DGI	37.46$_{\pm6.27}$	42.34$_{\pm8.31}$	44.36$_{\pm3.67}$	34.29$_{\pm4.71}$	37.80$_{\pm8.98}$	21.42$_{\pm3.13}$	14.34$_{\pm14.10}$
	GraphMAE	30.93$_{\pm3.64}$	20.76$_{\pm2.64}$	67.43$_{\pm2.96}$	33.60$_{\pm2.30}$	37.64$_{\pm5.88}$	21.65$_{\pm3.39}$	17.24$_{\pm13.25}$
	EdgePreGPPT	35.05$_{\pm2.95}$	24.26$_{\pm2.55}$	58.66$_{\pm5.93}$	29.37$_{\pm6.22}$	38.90$_{\pm7.56}$	20.35$_{\pm0.43}$	22.46$_{\pm4.05}$
	EdgePreGprompt	27.94$_{\pm5.07}$	23.21$_{\pm2.95}$	64.98$_{\pm5.35}$	33.49$_{\pm5.49}$	36.06$_{\pm12.01}$	19.85$_{\pm0.31}$	18.73$_{\pm4.77}$
	GraphCL	43.84$_{\pm6.11}$	27.09$_{\pm4.57}$	51.30$_{\pm6.35}$	23.89$_{\pm5.35}$	38.90$_{\pm8.86}$	20.60$_{\pm1.10}$	13.04$_{\pm6.44}$
	SimGRACE	29.66$_{\pm5.49}$	29.63$_{\pm1.87}$	43.95$_{\pm3.55}$	31.66$_{\pm6.21}$	34.65$_{\pm6.97}$	21.08$_{\pm1.16}$	14.13$_{\pm13.89}$
Gprompt	DGI	37.42$_{\pm7.07}$	41.19$_{\pm5.30}$	38.69$_{\pm1.54}$	71.54$_{\pm10.15}$	18.25$_{\pm21.68}$	26.64$_{\pm4.45}$	50.87$_{\pm3.51}$
	GraphMAE	63.78$_{\pm5.77}$	57.15$_{\pm3.90}$	62.47$_{\pm4.29}$	86.48$_{\pm7.48}$	36.99$_{\pm44.62}$	29.21$_{\pm3.27}$	73.92$_{\pm2.75}$
	EdgePreGPPT	51.29$_{\pm7.07}$	34.79$_{\pm4.07}$	53.63$_{\pm7.62}$	89.10$_{\pm4.09}$	37.49$_{\pm45.24}$	26.44$_{\pm3.36}$	64.52$_{\pm1.76}$
	EdgePreGprompt	53.37$_{\pm6.30}$	56.19$_{\pm2.95}$	66.68$_{\pm3.53}$	92.52$_{\pm5.38}$	38.66$_{\pm46.67}$	21.78$_{\pm1.85}$	68.77$_{\pm3.72}$
	GraphCL	56.61$_{\pm8.02}$	56.28$_{\pm5.89}$	50.33$_{\pm5.18}$	85.83$_{\pm4.53}$	36.16$_{\pm43.62}$	29.67$_{\pm2.53}$	58.48$_{\pm7.86}$
	SimGRACE	45.22$_{\pm5.69}$	60.00$_{\pm6.18}$	43.58$_{\pm5.78}$	82.84$_{\pm1.52}$	39.00$_{\pm47.08}$	25.07$_{\pm1.10}$	57.42$_{\pm4.67}$
All-in-one	DGI	33.80$_{\pm6.36}$	31.08$_{\pm2.95}$	48.09$_{\pm2.81}$	74.71$_{\pm2.35}$	31.94$_{\pm7.00}$	24.23$_{\pm1.39}$	29.92$_{\pm1.52}$
	GraphMAE	19.25$_{\pm3.11}$	19.02$_{\pm5.19}$	64.03$_{\pm4.70}$	76.90$_{\pm2.98}$	65.93$_{\pm17.42}$	19.56$_{\pm2.48}$	13.15$_{\pm4.63}$
	EdgePreGPPT	48.09$_{\pm4.83}$	34.29$_{\pm1.85}$	58.36$_{\pm4.63}$	89.62$_{\pm4.38}$	88.69$_{\pm1.08}$	20.91$_{\pm3.37}$	14.41$_{\pm5.57}$
	EdgePreGprompt	29.54$_{\pm6.25}$	24.63$_{\pm3.57}$	65.79$_{\pm5.79}$	84.29$_{\pm2.32}$	73.89$_{\pm16.92}$	20.70$_{\pm2.58}$	12.91$_{\pm2.08}$
	GraphCL	27.89$_{\pm18.34}$	43.96$_{\pm6.69}$	51.44$_{\pm4.46}$	67.06$_{\pm9.70}$	78.86$_{\pm4.06}$	22.14$_{\pm2.57}$	31.15$_{\pm2.25}$
	SimGRACE	25.16$_{\pm11.78}$	48.09$_{\pm8.18}$	47.87$_{\pm3.53}$	76.78$_{\pm1.28}$	82.93$_{\pm5.04}$	21.99$_{\pm1.17}$	15.47$_{\pm4.05}$
GPF	DGI	25.14$_{\pm19.07}$	25.92$_{\pm12.30}$	38.49$_{\pm14.50}$	64.36$_{\pm12.75}$	59.16$_{\pm14.47}$	25.03$_{\pm1.66}$	31.60$_{\pm9.57}$
	GraphMAE	33.93$_{\pm8.01}$	23.91$_{\pm5.11}$	71.20$_{\pm2.82}$	93.85$_{\pm3.71}$	93.66$_{\pm3.95}$	37.44$_{\pm3.43}$	38.70$_{\pm11.10}$
	EdgePreGPPT	23.05$_{\pm9.03}$	18.81$_{\pm4.35}$	65.06$_{\pm2.22}$	92.64$_{\pm2.97}$	92.15$_{\pm4.10}$	30.47$_{\pm3.57}$	59.67$_{\pm12.69}$
	EdgePreGprompt	32.57$_{\pm9.57}$	21.75$_{\pm2.88}$	67.57$_{\pm14.07}$	91.41$_{\pm2.76}$	95.34$_{\pm4.50}$	25.64$_{\pm8.58}$	35.13$_{\pm8.73}$
	GraphCL	34.84$_{\pm19.83}$	23.02$_{\pm5.56}$	49.15$_{\pm6.46}$	67.65$_{\pm15.44}$	95.47$_{\pm2.75}$	21.95$_{\pm6.54}$	36.09$_{\pm8.10}$
	SimGRACE	27.80$_{\pm18.09}$	21.20$_{\pm11.25}$	31.75$_{\pm13.86}$	86.96$_{\pm5.27}$	94.83$_{\pm4.07}$	22.82$_{\pm3.16}$	28.70$_{\pm5.98}$
GPF-plus	DGI	20.80$_{\pm13.69}$	24.46$_{\pm17.23}$	51.70$_{\pm11.29}$	77.55$_{\pm17.33}$	77.51$_{\pm22.66}$	23.99$_{\pm6.98}$	32.34$_{\pm11.81}$
	GraphMAE	56.38$_{\pm5.37}$	72.48$_{\pm5.63}$	70.85$_{\pm4.03}$	98.15$_{\pm0.73}$	96.50$_{\pm0.67}$	43.59$_{\pm4.52}$	44.44$_{\pm9.85}$
	EdgePreGPPT	32.62$_{\pm23.21}$	22.06$_{\pm21.01}$	67.20$_{\pm1.67}$	98.15$_{\pm0.83}$	97.49$_{\pm2.18}$	34.38$_{\pm5.23}$	64.51$_{\pm12.93}$
	EdgePreGprompt	52.74$_{\pm7.08}$	71.56$_{\pm8.03}$	68.84$_{\pm4.33}$	96.54$_{\pm1.57}$	97.66$_{\pm0.41}$	30.87$_{\pm6.61}$	64.63$_{\pm10.05}$
	GraphCL	30.34$_{\pm21.41}$	27.17$_{\pm24.65}$	45.62$_{\pm19.04}$	62.76$_{\pm14.68}$	97.50$_{\pm1.42}$	36.87$_{\pm2.45}$	43.66$_{\pm7.48}$
	SimGRACE	30.95$_{\pm19.32}$	28.97$_{\pm20.66}$	41.96$_{\pm15.64}$	80.92$_{\pm10.40}$	97.66$_{\pm1.41}$	34.89$_{\pm6.03}$	33.74$_{\pm5.22}$
MTG (Ours)	DGI	53.53$_{\pm6.60}$	73.81$_{\pm8.56}$	60.24$_{\pm7.33}$	74.61$_{\pm13.40}$	66.50$_{\pm14.18}$	32.94$_{\pm4.50}$	33.81$_{\pm8.32}$
	GraphMAE	52.27$_{\pm8.15}$	54.45$_{\pm5.07}$	65.01$_{\pm9.42}$	98.58$_{\pm0.93}$	93.66$_{\pm4.21}$	37.62$_{\pm4.72}$	37.97$_{\pm7.54}$
	EdgePreGPPT	51.87$_{\pm8.74}$	33.15$_{\pm15.11}$	63.32$_{\pm12.38}$	96.64$_{\pm2.19}$	93.15$_{\pm4.02}$	33.65$_{\pm3.39}$	76.01$_{\pm5.39}$
	EdgePreGprompt	52.78$_{\pm6.80}$	58.10$_{\pm4.89}$	71.38$_{\pm3.21}$	96.75$_{\pm3.53}$	98.17$_{\pm1.40}$	31.06$_{\pm1.82}$	28.45$_{\pm3.77}$
	GraphCL	66.11$_{\pm6.37}$	60.13$_{\pm4.49}$	55.87$_{\pm7.62}$	65.61$_{\pm8.65}$	93.49$_{\pm3.07}$	32.32$_{\pm3.27}$	42.53$_{\pm6.06}$
	SimGRACE	42.91$_{\pm9.14}$	64.66$_{\pm7.20}$	51.37$_{\pm10.41}$	78.42$_{\pm13.92}$	94.16$_{\pm3.70}$	27.73$_{\pm2.97}$	38.09$_{\pm13.67}$

Table 18: F1-score on 3-shot node classification.

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	$37.64_{\pm 6.32}$	$28.68_{\pm 8.09}$	$53.04_{\pm 6.01}$	$34.90_{\pm 5.93}$	$29.16_{\pm 7.77}$	$13.69_{\pm 2.87}$	$15.21_{\pm 2.66}$
Fine-tuning	DGI	$41.76_{\pm 1.48}$	$27.23_{\pm 9.82}$	$61.18_{\pm 10.18}$	$35.01_{\pm 5.55}$	$31.12_{\pm 8.67}$	$16.05_{\pm 0.75}$	$16.45_{\pm 1.97}$
	GraphMAE	$45.70_{\pm 6.35}$	$20.60_{\pm 5.72}$	$64.33_{\pm 3.42}$	$34.08_{\pm 4.01}$	$28.29_{\pm 8.05}$	$16.30_{\pm 2.37}$	$14.26_{\pm 4.46}$
	EdgePreGPPT	$49.76_{\pm 2.52}$	$28.14_{\pm 6.04}$	$63.62_{\pm 4.81}$	$35.03_{\pm 5.65}$	$25.55_{\pm 6.58}$	$17.79_{\pm 1.62}$	$16.68_{\pm 2.65}$
	EdgePredGprompt	$41.69_{\pm 5.58}$	$23.81_{\pm 4.86}$	$62.48_{\pm 4.02}$	$35.42_{\pm 6.02}$	$26.05_{\pm 6.95}$	$12.98_{\pm 2.02}$	$19.33_{\pm 2.46}$
	GraphCL	$47.79_{\pm 9.36}$	$34.47_{\pm 3.44}$	$63.10_{\pm 2.93}$	$22.42_{\pm 6.61}$	$29.45_{\pm 8.86}$	$19.91_{\pm 1.08}$	$7.46_{\pm 3.10}$
	SimGRACE	$41.71_{\pm 4.09}$	$42.01_{\pm 1.56}$	$63.14_{\pm 3.43}$	$34.28_{\pm 5.18}$	$29.55_{\pm 7.96}$	$17.92_{\pm 1.61}$	$11.68_{\pm 1.37}$
GPPTPrompt	DGI	$33.94_{\pm 7.56}$	$36.13_{\pm 8.65}$	$44.13_{\pm 3.80}$	$30.91_{\pm 2.41}$	$34.81_{\pm 7.09}$	$19.30_{\pm 2.98}$	$3.50_{\pm 2.51}$
	GraphMAE	$25.31_{\pm 6.18}$	$6.91_{\pm 2.93}$	$67.61_{\pm 2.62}$	$29.51_{\pm 3.03}$	$33.87_{\pm 3.38}$	$18.03_{\pm 1.97}$	$6.06_{\pm 5.21}$
	EdgePreGPPT	$32.77_{\pm 2.69}$	$19.54_{\pm 1.23}$	$58.07_{\pm 6.41}$	$27.20_{\pm 3.72}$	$33.65_{\pm 5.16}$	$19.32_{\pm 0.53}$	$17.24_{\pm 1.62}$
	EdgePreGprompt	$20.10_{\pm 4.36}$	$20.60_{\pm 1.98}$	$62.73_{\pm 4.96}$	$29.19_{\pm 4.29}$	$30.98_{\pm 7.36}$	$18.99_{\pm 0.70}$	$15.47_{\pm 2.95}$
	GraphCL	$41.25_{\pm 4.50}$	$25.71_{\pm 4.25}$	$50.27_{\pm 6.65}$	$15.65_{\pm 5.40}$	$34.61_{\pm 6.65}$	$19.81_{\pm 1.02}$	$5.96_{\pm 3.82}$
	SimGRACE	$27.19_{\pm 4.01}$	$27.09_{\pm 2.43}$	$42.79_{\pm 3.17}$	$26.79_{\pm 4.55}$	$32.38_{\pm 5.73}$	$19.02_{\pm 0.50}$	$4.01_{\pm 2.59}$
Gprompt	DGI	$32.92_{\pm 6.82}$	$37.71_{\pm 5.52}$	$37.90_{\pm 1.53}$	$60.12_{\pm 10.43}$	$23.89_{\pm 29.00}$	$26.39_{\pm 4.24}$	$46.53_{\pm 2.46}$
	GraphMAE	$59.30_{\pm 7.36}$	$53.60_{\pm 3.06}$	$61.33_{\pm 4.81}$	$79.06_{\pm 11.48}$	$30.56_{\pm 37.45}$	$29.02_{\pm 3.53}$	$65.63_{\pm 1.42}$
	EdgePreGPPT	$46.21_{\pm 6.52}$	$29.81_{\pm 3.69}$	$48.94_{\pm 10.06}$	$80.60_{\pm 7.17}$	$32.74_{\pm 40.02}$	$24.73_{\pm 3.32}$	$58.82_{\pm 1.46}$
	EdgePreGprompt	$52.43_{\pm 6.41}$	$53.88_{\pm 3.50}$	$65.02_{\pm 2.80}$	$91.03_{\pm 5.30}$	$29.77_{\pm 36.20}$	$20.40_{\pm 1.65}$	$57.08_{\pm 1.62}$
	GraphCL	$51.54_{\pm 8.49}$	$52.46_{\pm 5.08}$	$49.16_{\pm 5.73}$	$65.08_{\pm 11.12}$	$30.89_{\pm 37.90}$	$28.21_{\pm 2.34}$	$50.55_{\pm 3.75}$
	SimGRACE	$40.83_{\pm 3.88}$	$56.25_{\pm 7.01}$	$42.59_{\pm 5.51}$	$76.68_{\pm 3.65}$	$37.94_{\pm 46.21}$	$25.41_{\pm 1.35}$	$52.22_{\pm 1.38}$
All-in-one	DGI	$22.05_{\pm 4.15}$	$25.31_{\pm 3.49}$	$42.82_{\pm 4.66}$	$63.22_{\pm 7.20}$	$37.71_{\pm 11.00}$	$11.04_{\pm 2.36}$	$15.79_{\pm 0.87}$
	GraphMAE	$8.10_{\pm 2.43}$	$7.36_{\pm 2.81}$	$62.24_{\pm 5.80}$	$51.62_{\pm 2.63}$	$58.91_{\pm 14.27}$	$11.52_{\pm 0.97}$	$9.72_{\pm 3.20}$
	EdgePreGPPT	$41.54_{\pm 1.81}$	$27.24_{\pm 3.58}$	$56.20_{\pm 5.02}$	$80.48_{\pm 6.96}$	$82.89_{\pm 2.60}$	$9.51_{\pm 3.89}$	$9.61_{\pm 4.38}$
	EdgePreGprompt	$20.33_{\pm 9.09}$	$12.09_{\pm 3.36}$	$63.18_{\pm 7.52}$	$79.59_{\pm 6.58}$	$68.71_{\pm 14.50}$	$7.89_{\pm 2.38}$	$9.96_{\pm 1.24}$
	GraphCL	$14.85_{\pm 18.09}$	$33.38_{\pm 6.65}$	$50.52_{\pm 3.48}$	$61.79_{\pm 11.38}$	$69.09_{\pm 6.96}$	$16.40_{\pm 1.47}$	$14.33_{\pm 0.41}$
	SimGRACE	$12.71_{\pm 11.42}$	$40.22_{\pm 8.61}$	$44.79_{\pm 2.93}$	$64.88_{\pm 4.35}$	$79.46_{\pm 5.10}$	$13.70_{\pm 2.29}$	$10.07_{\pm 1.80}$
GPF	DGI	$12.93_{\pm 16.42}$	$12.92_{\pm 13.22}$	$32.76_{\pm 17.33}$	$57.67_{\pm 6.47}$	$54.60_{\pm 7.43}$	$15.68_{\pm 9.94}$	$12.83_{\pm 1.39}$
	GraphMAE	$19.59_{\pm 5.22}$	$14.48_{\pm 5.04}$	$70.15_{\pm 2.68}$	$87.58_{\pm 7.51}$	$86.43_{\pm 11.19}$	$31.56_{\pm 8.62}$	$14.17_{\pm 3.86}$
	EdgePreGPPT	$9.90_{\pm 9.04}$	$7.40_{\pm 3.59}$	$63.04_{\pm 3.09}$	$87.06_{\pm 6.90}$	$81.91_{\pm 9.67}$	$22.38_{\pm 8.76}$	$45.75_{\pm 5.33}$
	EdgePreGprompt	$21.59_{\pm 5.95}$	$11.79_{\pm 3.50}$	$62.02_{\pm 21.60}$	$86.44_{\pm 7.15}$	$86.41_{\pm 10.56}$	$22.04_{\pm 11.41}$	$19.34_{\pm 3.58}$
	GraphCL	$18.90_{\pm 20.40}$	$8.10_{\pm 4.95}$	$47.10_{\pm 8.36}$	$44.52_{\pm 22.48}$	$81.72_{\pm 11.88}$	$14.24_{\pm 8.36}$	$21.89_{\pm 6.08}$
	SimGRACE	$13.21_{\pm 14.54}$	$10.26_{\pm 8.85}$	$25.30_{\pm 17.26}$	$66.51_{\pm 11.13}$	$86.16_{\pm 10.16}$	$15.37_{\pm 7.26}$	$13.97_{\pm 3.27}$
GPF-plus	DGI	$10.87_{\pm 13.23}$	$14.35_{\pm 19.57}$	$43.25_{\pm 20.47}$	$75.32_{\pm 18.48}$	$61.47_{\pm 21.05}$	$14.70_{\pm 3.64}$	$14.05_{\pm 3.80}$
	GraphMAE	$55.76_{\pm 3.81}$	$70.45_{\pm 6.54}$	$69.78_{\pm 3.58}$	$94.11_{\pm 2.69}$	$81.45_{\pm 10.47}$	$45.02_{\pm 4.47}$	$20.13_{\pm 2.80}$
	EdgePreGPPT	$20.90_{\pm 24.90}$	$15.30_{\pm 23.70}$	$65.43_{\pm 1.28}$	$92.59_{\pm 4.18}$	$86.57_{\pm 9.96}$	$32.74_{\pm 6.14}$	$52.12_{\pm 8.51}$
	EdgePreGprompt	$52.13_{\pm 7.14}$	$68.79_{\pm 9.94}$	$67.98_{\pm 3.69}$	$94.33_{\pm 1.77}$	$85.78_{\pm 10.07}$	$30.59_{\pm 7.43}$	$44.07_{\pm 7.58}$
	GraphCL	$21.84_{\pm 23.15}$	$18.50_{\pm 28.28}$	$59.42_{\pm 13.43}$	$26.69_{\pm 14.43}$	$83.37_{\pm 9.14}$	$30.93_{\pm 5.62}$	$23.49_{\pm 1.47}$
	SimGRACE	$20.11_{\pm 22.26}$	$18.43_{\pm 24.95}$	$27.58_{\pm 21.15}$	$61.07_{\pm 14.66}$	$84.64_{\pm 10.24}$	$30.11_{\pm 6.73}$	$15.20_{\pm 0.24}$
MTG (Ours)	DGI	$46.54_{\pm 12.69}$	$72.30_{\pm 4.71}$	$59.76_{\pm 7.94}$	$67.09_{\pm 8.14}$	$58.29_{\pm 7.26}$	$30.76_{\pm 0.61}$	$17.44_{\pm 1.45}$
	GraphMAE	$49.48_{\pm 5.32}$	$49.16_{\pm 7.50}$	$64.33_{\pm 9.10}$	$94.48_{\pm 4.09}$	$86.91_{\pm 10.55}$	$34.46_{\pm 6.85}$	$14.52_{\pm 1.48}$
	EdgePreGPPT	$48.25_{\pm 7.70}$	$27.87_{\pm 15.20}$	$62.60_{\pm 12.00}$	$92.21_{\pm 5.04}$	$83.99_{\pm 10.75}$	$32.95_{\pm 4.15}$	$63.84_{\pm 5.11}$
	EdgePreGprompt	$50.76_{\pm 7.86}$	$52.42_{\pm 7.90}$	$67.80_{\pm 4.99}$	$93.83_{\pm 4.43}$	$87.36_{\pm 8.55}$	$29.64_{\pm 1.63}$	$16.90_{\pm 2.39}$
	GraphCL	$65.93_{\pm 6.51}$	$52.05_{\pm 5.93}$	$54.72_{\pm 7.26}$	$44.48_{\pm 11.68}$	$85.01_{\pm 10.01}$	$29.25_{\pm 5.15}$	$31.70_{\pm 3.22}$
	SimGRACE	$37.18_{\pm 6.46}$	$61.59_{\pm 10.38}$	$49.48_{\pm 9.46}$	$60.86_{\pm 11.75}$	$86.13_{\pm 10.46}$	$24.62_{\pm 2.91}$	$16.98_{\pm 6.64}$

Table 19: AUROC on 3-shot node classification.

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	$86.20_{\pm 2.47}$	$76.31_{\pm 3.45}$	$80.24_{\pm 2.82}$	$66.44_{\pm 6.12}$	$59.85_{\pm 12.05}$	$51.48_{\pm 0.74}$	$59.17_{\pm 1.26}$
Fine-tuning	DGI	$84.23_{\pm 2.75}$	$74.32_{\pm 4.21}$	$82.28_{\pm 5.43}$	$67.49_{\pm 3.80}$	$55.28_{\pm 10.97}$	$51.93_{\pm 1.20}$	$61.28_{\pm 1.40}$
	GraphMAE	$86.80_{\pm 2.35}$	$75.07_{\pm 2.48}$	$84.70_{\pm 1.28}$	$66.58_{\pm 6.51}$	$58.48_{\pm 11.37}$	$51.78_{\pm 1.01}$	$59.12_{\pm 1.66}$
	EdgePreGPPT	$85.62_{\pm 1.41}$	$72.39_{\pm 3.60}$	$83.03_{\pm 2.89}$	$66.70_{\pm 4.04}$	$54.49_{\pm 12.28}$	$52.25_{\pm 0.89}$	$60.55_{\pm 1.73}$
	EdgePredGprompt	$85.85_{\pm 0.79}$	$76.73_{\pm 2.26}$	$83.40_{\pm 2.74}$	$67.27_{\pm 4.41}$	$58.43_{\pm 10.76}$	$51.24_{\pm 0.96}$	$61.30_{\pm 1.07}$
	GraphCL	$87.71_{\pm 1.59}$	$75.10_{\pm 2.66}$	$81.83_{\pm 2.15}$	$61.56_{\pm 7.86}$	$57.10_{\pm 12.53}$	$52.03_{\pm 1.09}$	$54.66_{\pm 2.69}$
	SimGRACE	$80.39_{\pm 2.84}$	$76.94_{\pm 1.30}$	$79.56_{\pm 3.24}$	$67.34_{\pm 5.82}$	$59.02_{\pm 11.26}$	$51.60_{\pm 0.77}$	$56.28_{\pm 1.14}$
GPPTPrompt	DGI	$70.88_{\pm 2.54}$	$73.97_{\pm 5.46}$	$61.07_{\pm 2.93}$	$66.79_{\pm 2.66}$	$63.59_{\pm 7.91}$	$51.34_{\pm 1.54}$	$50.34_{\pm 0.68}$
	GraphMAE	$80.91_{\pm 2.38}$	$69.11_{\pm 3.46}$	$83.10_{\pm 1.03}$	$63.77_{\pm 3.61}$	$60.95_{\pm 3.13}$	$51.22_{\pm 1.38}$	$51.84_{\pm 3.86}$
	EdgePreGPPT	$68.71_{\pm 2.02}$	$59.01_{\pm 2.73}$	$74.22_{\pm 4.90}$	$57.60_{\pm 1.78}$	$58.20_{\pm 4.72}$	$50.23_{\pm 0.31}$	$58.95_{\pm 1.29}$
	EdgePreGprompt	$68.29_{\pm 3.52}$	$58.07_{\pm 2.40}$	$81.15_{\pm 4.81}$	$61.93_{\pm 5.66}$	$61.91_{\pm 5.28}$	$50.27_{\pm 0.63}$	$56.16_{\pm 2.47}$
	GraphCL	$75.26_{\pm 3.86}$	$60.58_{\pm 4.04}$	$71.43_{\pm 2.47}$	$53.79_{\pm 2.21}$	$61.73_{\pm 5.50}$	$50.96_{\pm 0.66}$	$50.69_{\pm 1.68}$
	SimGRACE	$66.90_{\pm 4.18}$	$65.72_{\pm 2.09}$	$62.58_{\pm 2.08}$	$66.87_{\pm 2.38}$	$60.35_{\pm 7.62}$	$50.53_{\pm 0.57}$	$50.07_{\pm 0.24}$
Gprompt	DGI	$62.24_{\pm 2.99}$	$75.87_{\pm 3.70}$	$55.46_{\pm 1.89}$	$87.72_{\pm 5.97}$	$60.44_{\pm 13.45}$	$55.78_{\pm 3.19}$	$92.08_{\pm 0.52}$
	GraphMAE	$80.27_{\pm 5.61}$	$74.38_{\pm 3.36}$	$84.37_{\pm 1.75}$	$96.34_{\pm 2.81}$	$60.67_{\pm 13.69}$	$58.52_{\pm 2.66}$	$93.17_{\pm 0.44}$
	EdgePreGPPT	$81.41_{\pm 5.40}$	$65.99_{\pm 4.06}$	$72.84_{\pm 0.48}$	$98.59_{\pm 1.16}$	$61.51_{\pm 14.67}$	$56.86_{\pm 2.69}$	$94.31_{\pm 0.11}$
	EdgePreGprompt	$83.15_{\pm 3.91}$	$82.01_{\pm 1.78}$	$84.45_{\pm 1.80}$	$99.32_{\pm 0.66}$	$62.34_{\pm 15.53}$	$51.33_{\pm 2.32}$	$95.60_{\pm 0.06}$
	GraphCL	$83.25_{\pm 3.22}$	$83.20_{\pm 0.97}$	$71.35_{\pm 3.39}$	$91.20_{\pm 4.59}$	$61.92_{\pm 15.06}$	$59.78_{\pm 3.41}$	$94.76_{\pm 0.71}$
	SimGRACE	$79.02_{\pm 2.55}$	$86.57_{\pm 1.86}$	$61.89_{\pm 6.64}$	$94.09_{\pm 3.00}$	$62.38_{\pm 15.57}$	$56.13_{\pm 1.59}$	$94.71_{\pm 0.39}$
All-in-one	DGI	$71.15_{\pm 3.33}$	$65.99_{\pm 0.90}$	$66.83_{\pm 1.02}$	$81.86_{\pm 0.82}$	$46.05_{\pm 2.66}$	$52.80_{\pm 2.23}$	$55.31_{\pm 0.33}$
	GraphMAE	$77.89_{\pm 3.94}$	$63.31_{\pm 2.05}$	$84.13_{\pm 1.99}$	$81.57_{\pm 1.86}$	$66.49_{\pm 5.26}$	$50.76_{\pm 1.17}$	$55.43_{\pm 0.07}$
	EdgePreGPPT	$80.90_{\pm 2.72}$	$68.74_{\pm 3.28}$	$76.23_{\pm 0.81}$	$96.17_{\pm 3.82}$	$75.39_{\pm 1.30}$	$52.98_{\pm 1.21}$	$56.37_{\pm 0.72}$
	EdgePreGprompt	$77.50_{\pm 2.96}$	$69.84_{\pm 1.80}$	$85.03_{\pm 1.73}$	$91.35_{\pm 2.44}$	$72.04_{\pm 6.05}$	$49.89_{\pm 1.95}$	$54.08_{\pm 0.15}$
	GraphCL	$65.18_{\pm 11.80}$	$77.64_{\pm 2.06}$	$73.05_{\pm 1.97}$	$82.74_{\pm 3.64}$	$72.60_{\pm 1.20}$	$51.67_{\pm 1.31}$	$53.03_{\pm 0.27}$
	SimGRACE	$59.44_{\pm 9.08}$	$78.11_{\pm 1.17}$	$64.76_{\pm 2.62}$	$84.99_{\pm 2.40}$	$92.99_{\pm 1.95}$	$51.64_{\pm 2.25}$	$51.35_{\pm 0.09}$
GPF	DGI	$60.20_{\pm 13.06}$	$59.79_{\pm 13.56}$	$59.52_{\pm 12.99}$	$77.84_{\pm 2.81}$	$63.85_{\pm 10.27}$	$54.39_{\pm 2.39}$	$73.91_{\pm 2.08}$
	GraphMAE	$82.70_{\pm 2.62}$	$72.25_{\pm 4.60}$	$87.91_{\pm 1.01}$	$98.45_{\pm 1.93}$	$87.47_{\pm 9.36}$	$67.24_{\pm 4.73}$	$78.12_{\pm 2.53}$
	EdgePreGPPT	$65.44_{\pm 9.07}$	$67.13_{\pm 1.95}$	$81.07_{\pm 0.78}$	$99.15_{\pm 0.36}$	$78.69_{\pm 1.83}$	$58.43_{\pm 3.84}$	$92.57_{\pm 1.23}$
	EdgePreGprompt	$83.18_{\pm 3.22}$	$69.63_{\pm 7.04}$	$85.38_{\pm 6.75}$	$98.35_{\pm 2.86}$	$83.30_{\pm 5.10}$	$56.19_{\pm 3.18}$	$83.95_{\pm 0.73}$
	GraphCL	$73.76_{\pm 9.73}$	$68.34_{\pm 10.91}$	$75.10_{\pm 2.04}$	$57.42_{\pm 18.02}$	$80.86_{\pm 1.38}$	$56.96_{\pm 4.69}$	$77.45_{\pm 3.21}$
	SimGRACE	$63.96_{\pm 7.80}$	$63.26_{\pm 11.65}$	$58.71_{\pm 6.49}$	$95.32_{\pm 3.52}$	$86.97_{\pm 9.21}$	$55.06_{\pm 3.04}$	$63.69_{\pm 5.22}$
GPF-plus	DGI	$55.30_{\pm 13.20}$	$55.81_{\pm 15.74}$	$66.54_{\pm 15.31}$	$91.04_{\pm 3.86}$	$78.85_{\pm 4.55}$	$57.09_{\pm 3.17}$	$72.14_{\pm 5.71}$
	GraphMAE	$87.26_{\pm 3.14}$	$91.53_{\pm 1.55}$	$87.12_{\pm 1.71}$	$99.03_{\pm 0.13}$	$82.17_{\pm 4.17}$	$72.39_{\pm 1.82}$	$82.95_{\pm 1.11}$
	EdgePreGPPT	$73.87_{\pm 10.93}$	$56.16_{\pm 16.39}$	$82.80_{\pm 1.90}$	$99.24_{\pm 0.09}$	$86.47_{\pm 7.95}$	$67.15_{\pm 3.76}$	$93.54_{\pm 1.76}$
	EdgePreGprompt	$85.81_{\pm 3.03}$	$91.47_{\pm 1.96}$	$85.13_{\pm 2.57}$	$98.99_{\pm 0.19}$	$84.00_{\pm 7.50}$	$65.17_{\pm 4.06}$	$89.16_{\pm 2.12}$
	GraphCL	$80.93_{\pm 6.55}$	$61.66_{\pm 15.01}$	$80.00_{\pm 7.05}$	$65.42_{\pm 19.34}$	$80.57_{\pm 1.41}$	$69.40_{\pm 1.84}$	$81.66_{\pm 2.30}$
	SimGRACE	$66.36_{\pm 16.56}$	$60.91_{\pm 15.05}$	$61.66_{\pm 12.28}$	$90.57_{\pm 6.03}$	$82.68_{\pm 5.45}$	$66.44_{\pm 4.39}$	$67.35_{\pm 2.71}$
MTG (Ours)	DGI	$84.42_{\pm 8.91}$	$91.92_{\pm 2.82}$	$84.10_{\pm 4.41}$	$89.88_{\pm 4.35}$	$67.37_{\pm 11.67}$	$59.80_{\pm 5.38}$	$73.10_{\pm 1.13}$
	GraphMAE	$85.04_{\pm 3.04}$	$78.35_{\pm 5.16}$	$87.65_{\pm 3.13}$	$99.19_{\pm 0.87}$	$87.51_{\pm 9.57}$	$69.62_{\pm 4.72}$	$80.01_{\pm 1.66}$
	EdgePreGPPT	$82.73_{\pm 4.81}$	$65.55_{\pm 10.41}$	$82.66_{\pm 9.56}$	$99.76_{\pm 0.29}$	$79.08_{\pm 0.73}$	$64.07_{\pm 2.31}$	$95.04_{\pm 1.27}$
	EdgePreGprompt	$83.39_{\pm 4.54}$	$81.04_{\pm 5.11}$	$86.16_{\pm 3.20}$	$99.20_{\pm 1.45}$	$84.19_{\pm 7.49}$	$61.34_{\pm 1.54}$	$77.30_{\pm 2.92}$
	GraphCL	$89.07_{\pm 3.69}$	$83.12_{\pm 6.58}$	$77.04_{\pm 4.99}$	$67.06_{\pm 9.98}$	$80.35_{\pm 0.79}$	$59.76_{\pm 3.77}$	$78.37_{\pm 2.86}$
	SimGRACE	$73.87_{\pm 5.96}$	$89.93_{\pm 1.91}$	$68.11_{\pm 11.04}$	$87.06_{\pm 12.66}$	$86.82_{\pm 8.78}$	$55.80_{\pm 2.89}$	$59.04_{\pm 7.05}$

Table 20: Accuracy (%) on 5-shot node classification

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	$50.25_{\pm 8.37}$	$41.22_{\pm 6.30}$	$67.88_{\pm 2.18}$	$39.43_{\pm 5.86}$	$43.91_{\pm 6.47}$	$21.92_{\pm 1.86}$	$22.38_{\pm 3.05}$
Fine-tuning	DGI	$48.79_{\pm 8.51}$	$35.91_{\pm 4.94}$	$61.44_{\pm 7.93}$	$40.00_{\pm 6.31}$	$47.19_{\pm 7.37}$	$21.48_{\pm 1.71}$	$12.48_{\pm 6.15}$
	GraphMAE	$54.09_{\pm 4.21}$	$32.75_{\pm 6.35}$	$70.04_{\pm 4.57}$	$40.69_{\pm 6.73}$	$43.44_{\pm 8.48}$	$22.92_{\pm 1.22}$	$24.08_{\pm 1.45}$
	EdgePreGPPT	$54.34_{\pm 3.78}$	$38.97_{\pm 7.35}$	$70.91_{\pm 4.87}$	$40.69_{\pm 6.77}$	$44.53_{\pm 11.05}$	$21.46_{\pm 1.00}$	$28.84_{\pm 3.11}$
	EdgePredGprompt	$49.04_{\pm 8.99}$	$32.08_{\pm 8.24}$	$70.44_{\pm 5.04}$	$41.03_{\pm 6.58}$	$45.78_{\pm 8.07}$	$22.46_{\pm 1.99}$	$25.94_{\pm 2.63}$
	GraphCL	$62.66_{\pm 3.55}$	$39.54_{\pm 3.54}$	$66.84_{\pm 5.78}$	$42.97_{\pm 8.99}$	$46.41_{\pm 5.47}$	$21.99_{\pm 1.61}$	$13.42_{\pm 3.73}$
	SimGRACE	$45.13_{\pm 7.81}$	$38.90_{\pm 6.03}$	$62.65_{\pm 6.29}$	$41.49_{\pm 5.77}$	$46.09_{\pm 8.71}$	$22.77_{\pm 1.56}$	$9.45_{\pm 1.62}$
GPPTPrompt	DGI	$43.68_{\pm 7.12}$	$45.77_{\pm 7.41}$	$47.39_{\pm 10.22}$	$36.29_{\pm 3.97}$	$48.82_{\pm 5.15}$	$20.91_{\pm 1.36}$	$8.85_{\pm 6.21}$
	GraphMAE	$31.50_{\pm 11.89}$	$19.93_{\pm 6.60}$	$66.97_{\pm 3.70}$	$37.00_{\pm 3.19}$	$42.68_{\pm 6.91}$	$22.10_{\pm 0.86}$	$8.90_{\pm 6.25}$
	EdgePreGPPT	$32.94_{\pm 2.82}$	$26.93_{\pm 1.56}$	$63.45_{\pm 3.63}$	$34.00_{\pm 5.69}$	$36.85_{\pm 3.24}$	$20.42_{\pm 0.81}$	$28.90_{\pm 1.64}$
	EdgePreGprompt	$30.55_{\pm 4.59}$	$28.46_{\pm 4.25}$	$63.33_{\pm 4.18}$	$32.57_{\pm 3.31}$	$39.21_{\pm 6.09}$	$20.07_{\pm 0.45}$	$4.01_{\pm 0.69}$
	GraphCL	$51.98_{\pm 3.43}$	$26.68_{\pm 2.95}$	$57.59_{\pm 2.89}$	$30.43_{\pm 8.83}$	$45.20_{\pm 2.20}$	$21.25_{\pm 0.78}$	$14.15_{\pm 5.26}$
	SimGRACE	$34.93_{\pm 2.55}$	$29.51_{\pm 2.60}$	$42.87_{\pm 1.57}$	$33.00_{\pm 5.27}$	$41.42_{\pm 6.80}$	$21.58_{\pm 0.84}$	$9.11_{\pm 3.50}$
Gprompt	DGI	$27.81_{\pm 10.28}$	$52.84_{\pm 2.36}$	$37.83_{\pm 5.71}$	$57.24_{\pm 27.27}$	$34.47_{\pm 41.15}$	$26.07_{\pm 4.20}$	$56.59_{\pm 4.05}$
	GraphMAE	$66.82_{\pm 3.98}$	$60.07_{\pm 3.98}$	$66.66_{\pm 2.25}$	$74.89_{\pm 35.83}$	$36.73_{\pm 43.95}$	$30.71_{\pm 2.12}$	$76.98_{\pm 3.88}$
	EdgePreGPPT	$53.15_{\pm 2.85}$	$36.15_{\pm 5.00}$	$62.18_{\pm 4.88}$	$78.22_{\pm 37.33}$	$37.59_{\pm 45.03}$	$28.76_{\pm 3.95}$	$85.40_{\pm 0.79}$
	EdgePreGprompt	$62.63_{\pm 3.26}$	$65.34_{\pm 4.60}$	$67.87_{\pm 2.08}$	$78.01_{\pm 37.38}$	$38.45_{\pm 46.03}$	$21.86_{\pm 2.13}$	$81.85_{\pm 3.16}$
	GraphCL	$69.03_{\pm 3.61}$	$61.27_{\pm 5.37}$	$57.21_{\pm 2.78}$	$65.72_{\pm 31.30}$	$36.38_{\pm 43.47}$	$34.67_{\pm 1.28}$	$57.52_{\pm 4.19}$
	SimGRACE	$51.27_{\pm 6.05}$	$66.13_{\pm 1.64}$	$44.78_{\pm 6.52}$	$68.34_{\pm 32.42}$	$39.32_{\pm 47.08}$	$30.06_{\pm 3.36}$	$57.36_{\pm 3.68}$
All-in-one	DGI	$30.45_{\pm 0.19}$	$21.82_{\pm 4.73}$	$43.60_{\pm 5.24}$	$79.95_{\pm 4.01}$	$62.50_{\pm 12.39}$	$19.83_{\pm 3.14}$	$2.65_{\pm 0.09}$
	GraphMAE	$25.02_{\pm 7.58}$	$19.38_{\pm 4.38}$	$46.16_{\pm 15.83}$	$82.43_{\pm 2.84}$	$66.47_{\pm 17.32}$	$15.03_{\pm 3.13}$	$0.24_{\pm 0.00}$
	EdgePreGPPT	$30.36_{\pm 13.48}$	$25.83_{\pm 9.32}$	$37.61_{\pm 20.58}$	$88.55_{\pm 3.52}$	$47.98_{\pm 27.78}$	$21.49_{\pm 3.02}$	$1.59_{\pm 1.30}$
	EdgePreGprompt	$21.59_{\pm 7.30}$	$18.75_{\pm 1.86}$	$39.36_{\pm 11.45}$	$87.16_{\pm 3.02}$	$46.32_{\pm 29.89}$	$21.23_{\pm 5.64}$	$13.01_{\pm 6.29}$
	GraphCL	$25.63_{\pm 17.68}$	$27.93_{\pm 10.59}$	$42.86_{\pm 7.21}$	$84.75_{\pm 8.86}$	$73.28_{\pm 9.91}$	$20.83_{\pm 8.04}$	$6.70_{\pm 6.01}$
	SimGRACE	$15.20_{\pm 9.51}$	$25.19_{\pm 12.95}$	$21.59_{\pm 2.58}$	$78.90_{\pm 2.26}$	$64.63_{\pm 25.59}$	$21.26_{\pm 6.24}$	$4.80_{\pm 5.17}$
GPF	DGI	$29.57_{\pm 20.89}$	$24.55_{\pm 12.61}$	$52.43_{\pm 6.94}$	$83.21_{\pm 5.19}$	$89.30_{\pm 1.58}$	$35.40_{\pm 8.78}$	$20.72_{\pm 0.26}$
	GraphMAE	$35.43_{\pm 1.02}$	$25.12_{\pm 3.01}$	$68.96_{\pm 3.99}$	$96.30_{\pm 5.12}$	$92.44_{\pm 3.04}$	$44.07_{\pm 3.94}$	$45.20_{\pm 2.03}$
	EdgePreGPPT	$15.39_{\pm 8.19}$	$18.11_{\pm 2.78}$	$58.67_{\pm 3.21}$	$98.26_{\pm 1.19}$	$89.47_{\pm 1.91}$	$30.99_{\pm 4.68}$	$71.83_{\pm 9.37}$
	EdgePreGprompt	$31.58_{\pm 18.16}$	$24.55_{\pm 8.28}$	$66.25_{\pm 8.53}$	$96.30_{\pm 5.12}$	$90.18_{\pm 1.57}$	$31.92_{\pm 5.75}$	$28.60_{\pm 3.11}$
	GraphCL	$28.60_{\pm 11.19}$	$17.69_{\pm 1.35}$	$52.47_{\pm 6.73}$	$69.15_{\pm 21.44}$	$98.42_{\pm 0.36}$	$29.13_{\pm 2.63}$	$23.90_{\pm 0.12}$
	SimGRACE	$18.94_{\pm 12.60}$	$22.45_{\pm 3.45}$	$40.35_{\pm 1.71}$	$92.29_{\pm 3.36}$	$91.73_{\pm 3.84}$	$30.28_{\pm 2.64}$	$33.77_{\pm 8.48}$
GPF-plus	DGI	$27.23_{\pm 14.61}$	$26.31_{\pm 11.48}$	$47.02_{\pm 14.51}$	$83.86_{\pm 17.24}$	$96.18_{\pm 4.12}$	$36.03_{\pm 7.49}$	$16.86_{\pm 3.30}$
	GraphMAE	$63.28_{\pm 4.69}$	$75.73_{\pm 2.19}$	$69.59_{\pm 4.33}$	$99.01_{\pm 1.43}$	$99.12_{\pm 0.95}$	$44.58_{\pm 5.95}$	$47.79_{\pm 1.09}$
	EdgePreGPPT	$22.44_{\pm 23.88}$	$13.63_{\pm 5.27}$	$66.43_{\pm 3.28}$	$98.52_{\pm 2.07}$	$96.18_{\pm 4.12}$	$37.15_{\pm 8.48}$	$66.88_{\pm 6.14}$
	EdgePreGprompt	$66.22_{\pm 6.20}$	$64.49_{\pm 14.12}$	$68.10_{\pm 4.56}$	$98.64_{\pm 2.14}$	$97.74_{\pm 2.47}$	$41.98_{\pm 4.70}$	$51.25_{\pm 1.31}$
	GraphCL	$47.71_{\pm 22.44}$	$29.16_{\pm 16.66}$	$64.53_{\pm 4.15}$	$70.98_{\pm 19.27}$	$99.12_{\pm 0.95}$	$36.99_{\pm 6.59}$	$25.74_{\pm 1.70}$
	SimGRACE	$27.79_{\pm 21.51}$	$28.37_{\pm 22.54}$	$50.25_{\pm 10.04}$	$89.57_{\pm 4.22}$	$95.83_{\pm 3.85}$	$38.23_{\pm 2.34}$	$44.70_{\pm 2.69}$
MTG (Ours)	DGI	$53.47_{\pm 11.79}$	$76.34_{\pm 6.18}$	$61.52_{\pm 4.74}$	$82.69_{\pm 8.67}$	$89.77_{\pm 3.04}$	$35.42_{\pm 3.92}$	$13.38_{\pm 7.85}$
	GraphMAE	$57.16_{\pm 8.68}$	$61.37_{\pm 7.73}$	$65.71_{\pm 3.94}$	$99.23_{\pm 0.62}$	$93.61_{\pm 5.01}$	$45.09_{\pm 3.26}$	$42.98_{\pm 10.64}$
	EdgePreGPPT	$56.96_{\pm 6.84}$	$44.49_{\pm 7.80}$	$63.34_{\pm 6.29}$	$98.48_{\pm 0.70}$	$97.17_{\pm 6.08}$	$35.36_{\pm 1.85}$	$85.94_{\pm 1.93}$
	EdgePreGprompt	$53.61_{\pm 7.57}$	$61.26_{\pm 9.76}$	$70.84_{\pm 3.28}$	$97.51_{\pm 3.88}$	$98.76_{\pm 2.36}$	$35.06_{\pm 2.02}$	$38.04_{\pm 2.24}$
	GraphCL	$71.81_{\pm 3.59}$	$66.59_{\pm 8.39}$	$58.18_{\pm 3.34}$	$72.02_{\pm 6.21}$	$94.65_{\pm 4.67}$	$33.94_{\pm 3.77}$	$50.04_{\pm 7.84}$
	SimGRACE	$48.98_{\pm 6.01}$	$68.33_{\pm 5.69}$	$52.74_{\pm 1.25}$	$87.24_{\pm 5.85}$	$94.81_{\pm 5.06}$	$32.76_{\pm 0.93}$	$38.67_{\pm 3.86}$

Table 21: F1-score on 5-shot node classification.

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	$51.42_{\pm 8.70}$	$34.20_{\pm 6.12}$	$67.45_{\pm 2.25}$	$34.31_{\pm 4.41}$	$35.02_{\pm 8.10}$	$16.89_{\pm 1.20}$	$14.76_{\pm 1.13}$
Fine-tuning	DGI	$47.53_{\pm 9.44}$	$28.71_{\pm 2.76}$	$59.33_{\pm 10.23}$	$34.71_{\pm 5.28}$	$38.80_{\pm 7.04}$	$16.49_{\pm 1.42}$	$4.78_{\pm 2.53}$
	GraphMAE	$55.17_{\pm 5.27}$	$29.67_{\pm 6.94}$	$69.58_{\pm 4.63}$	$34.69_{\pm 5.63}$	$34.57_{\pm 9.23}$	$16.92_{\pm 1.76}$	$12.97_{\pm 0.32}$
	EdgePreGPPT	$55.50_{\pm 3.62}$	$31.31_{\pm 8.86}$	$70.31_{\pm 4.65}$	$34.86_{\pm 4.83}$	$36.63_{\pm 7.69}$	$19.42_{\pm 1.02}$	$20.19_{\pm 1.27}$
	EdgePredGprompt	$49.18_{\pm 8.50}$	$27.04_{\pm 6.83}$	$69.46_{\pm 4.79}$	$34.90_{\pm 5.77}$	$36.38_{\pm 7.04}$	$18.28_{\pm 1.91}$	$18.51_{\pm 1.18}$
	GraphCL	$61.32_{\pm 3.43}$	$35.95_{\pm 3.18}$	$66.16_{\pm 5.43}$	$26.66_{\pm 5.78}$	$33.81_{\pm 3.28}$	$20.44_{\pm 1.45}$	$6.96_{\pm 1.07}$
	SimGRACE	$45.33_{\pm 4.88}$	$34.74_{\pm 7.40}$	$62.66_{\pm 6.39}$	$35.86_{\pm 4.86}$	$36.86_{\pm 7.72}$	$20.63_{\pm 1.13}$	$4.74_{\pm 0.48}$
GPPTPrompt	DGI	$40.79_{\pm 7.28}$	$39.86_{\pm 7.94}$	$44.67_{\pm 12.25}$	$32.76_{\pm 3.66}$	$39.86_{\pm 4.07}$	$19.28_{\pm 0.85}$	$0.39_{\pm 0.27}$
	GraphMAE	$24.55_{\pm 7.97}$	$7.49_{\pm 4.08}$	$66.55_{\pm 3.62}$	$33.86_{\pm 3.88}$	$36.66_{\pm 2.42}$	$19.92_{\pm 0.84}$	$0.42_{\pm 0.29}$
	EdgePreGPPT	$32.50_{\pm 2.42}$	$19.95_{\pm 1.98}$	$62.42_{\pm 4.78}$	$30.57_{\pm 4.59}$	$33.93_{\pm 2.76}$	$19.89_{\pm 0.56}$	$18.11_{\pm 1.03}$
	EdgePreGprompt	$22.20_{\pm 3.24}$	$25.96_{\pm 4.33}$	$61.22_{\pm 5.46}$	$30.39_{\pm 2.77}$	$35.84_{\pm 3.35}$	$19.22_{\pm 0.33}$	$2.56_{\pm 1.55}$
	GraphCL	$50.60_{\pm 2.51}$	$25.00_{\pm 2.05}$	$57.27_{\pm 2.84}$	$18.78_{\pm 5.00}$	$35.83_{\pm 1.96}$	$20.34_{\pm 0.74}$	$3.03_{\pm 1.03}$
	SimGRACE	$31.56_{\pm 2.75}$	$28.27_{\pm 1.68}$	$40.87_{\pm 2.90}$	$27.06_{\pm 3.81}$	$36.11_{\pm 2.81}$	$20.45_{\pm 0.51}$	$1.63_{\pm 0.40}$
Gprompt	DGI	$19.96_{\pm 9.68}$	$47.84_{\pm 2.39}$	$37.26_{\pm 5.45}$	$48.86_{\pm 23.99}$	$29.50_{\pm 36.22}$	$24.02_{\pm 4.65}$	$49.69_{\pm 1.68}$
	GraphMAE	$61.57_{\pm 5.38}$	$56.39_{\pm 3.60}$	$65.31_{\pm 2.36}$	$70.18_{\pm 34.72}$	$36.05_{\pm 43.75}$	$30.08_{\pm 1.88}$	$66.82_{\pm 2.68}$
	EdgePreGPPT	$51.03_{\pm 3.08}$	$31.15_{\pm 3.86}$	$60.74_{\pm 4.30}$	$73.65_{\pm 36.66}$	$34.07_{\pm 41.83}$	$27.29_{\pm 3.43}$	$79.73_{\pm 1.29}$
	EdgePreGprompt	$60.77_{\pm 3.39}$	$63.42_{\pm 5.11}$	$66.46_{\pm 2.30}$	$77.08_{\pm 37.97}$	$34.45_{\pm 41.99}$	$21.03_{\pm 2.04}$	$79.20_{\pm 1.95}$
	GraphCL	$64.22_{\pm 4.08}$	$57.65_{\pm 5.74}$	$56.66_{\pm 2.35}$	$51.82_{\pm 25.62}$	$31.38_{\pm 38.22}$	$32.99_{\pm 1.00}$	$49.82_{\pm 2.81}$
	SimGRACE	$45.98_{\pm 5.22}$	$63.84_{\pm 1.37}$	$42.20_{\pm 5.64}$	$62.12_{\pm 30.55}$	$38.47_{\pm 46.70}$	$29.63_{\pm 3.89}$	$53.41_{\pm 2.08}$
All-in-one	DGI	$6.67_{\pm 0.03}$	$10.68_{\pm 7.84}$	$26.81_{\pm 9.85}$	$60.73_{\pm 6.34}$	$31.59_{\pm 18.97}$	$10.66_{\pm 3.88}$	$0.21_{\pm 0.04}$
	GraphMAE	$6.77_{\pm 3.23}$	$8.66_{\pm 4.25}$	$33.97_{\pm 19.42}$	$58.02_{\pm 3.64}$	$49.27_{\pm 9.03}$	$8.31_{\pm 2.80}$	$0.01_{\pm 0.00}$
	EdgePreGPPT	$15.19_{\pm 16.25}$	$13.46_{\pm 10.23}$	$31.47_{\pm 24.42}$	$78.97_{\pm 11.06}$	$29.95_{\pm 20.92}$	$10.10_{\pm 3.63}$	$0.53_{\pm 0.63}$
	EdgePreGprompt	$4.99_{\pm 1.38}$	$5.28_{\pm 0.47}$	$27.49_{\pm 15.02}$	$68.69_{\pm 3.29}$	$25.44_{\pm 17.47}$	$8.08_{\pm 2.12}$	$0.56_{\pm 0.27}$
	GraphCL	$14.88_{\pm 18.86}$	$14.98_{\pm 12.03}$	$31.71_{\pm 13.56}$	$74.33_{\pm 11.49}$	$49.70_{\pm 21.87}$	$10.10_{\pm 6.01}$	$0.75_{\pm 1.07}$
	SimGRACE	$7.21_{\pm 7.14}$	$13.71_{\pm 14.05}$	$13.42_{\pm 3.26}$	$66.26_{\pm 6.68}$	$45.47_{\pm 21.83}$	$11.42_{\pm 5.14}$	$0.52_{\pm 0.70}$
GPF	DGI	$18.44_{\pm 18.83}$	$13.25_{\pm 11.04}$	$47.23_{\pm 9.28}$	$75.78_{\pm 14.28}$	$77.01_{\pm 10.57}$	$28.58_{\pm 14.86}$	$11.02_{\pm 0.99}$
	GraphMAE	$23.17_{\pm 2.53}$	$12.78_{\pm 1.23}$	$67.88_{\pm 4.07}$	$87.75_{\pm 13.24}$	$82.81_{\pm 10.55}$	$43.09_{\pm 4.07}$	$16.76_{\pm 2.76}$
	EdgePreGPPT	$3.69_{\pm 1.67}$	$5.09_{\pm 0.66}$	$53.91_{\pm 7.41}$	$94.37_{\pm 8.31}$	$80.86_{\pm 9.83}$	$23.04_{\pm 9.22}$	$55.01_{\pm 8.05}$
	EdgePreGprompt	$20.17_{\pm 13.79}$	$12.79_{\pm 6.88}$	$62.47_{\pm 11.25}$	$89.10_{\pm 11.52}$	$79.88_{\pm 10.50}$	$30.44_{\pm 9.87}$	$18.77_{\pm 3.29}$
	GraphCL	$11.74_{\pm 6.74}$	$5.99_{\pm 2.19}$	$46.46_{\pm 14.28}$	$51.01_{\pm 22.81}$	$86.64_{\pm 9.95}$	$22.59_{\pm 5.87}$	$18.74_{\pm 0.78}$
	SimGRACE	$10.17_{\pm 12.35}$	$9.91_{\pm 2.56}$	$29.36_{\pm 8.75}$	$71.38_{\pm 2.97}$	$81.14_{\pm 11.49}$	$22.30_{\pm 8.60}$	$12.31_{\pm 4.13}$
GPF-plus	DGI	$13.92_{\pm 14.65}$	$12.41_{\pm 13.45}$	$40.73_{\pm 16.83}$	$71.22_{\pm 13.69}$	$84.85_{\pm 10.28}$	$32.05_{\pm 9.40}$	$6.42_{\pm 2.47}$
	GraphMAE	$61.55_{\pm 4.90}$	$74.86_{\pm 2.27}$	$68.34_{\pm 4.25}$	$94.48_{\pm 8.79}$	$86.93_{\pm 10.70}$	$44.28_{\pm 6.46}$	$24.14_{\pm 2.16}$
	EdgePreGPPT	$15.91_{\pm 26.40}$	$4.26_{\pm 1.11}$	$61.79_{\pm 8.54}$	$93.18_{\pm 10.98}$	$82.10_{\pm 9.29}$	$35.11_{\pm 10.76}$	$56.15_{\pm 4.28}$
	EdgePreGprompt	$65.96_{\pm 5.07}$	$57.16_{\pm 18.91}$	$66.11_{\pm 4.85}$	$93.46_{\pm 10.79}$	$85.56_{\pm 9.53}$	$39.84_{\pm 6.02}$	$32.07_{\pm 2.05}$
	GraphCL	$38.84_{\pm 26.82}$	$18.53_{\pm 21.72}$	$63.06_{\pm 3.88}$	$65.13_{\pm 17.26}$	$86.59_{\pm 11.04}$	$33.23_{\pm 7.80}$	$19.46_{\pm 1.47}$
	SimGRACE	$19.60_{\pm 23.99}$	$18.27_{\pm 26.81}$	$36.93_{\pm 16.65}$	$69.27_{\pm 9.83}$	$84.63_{\pm 10.23}$	$35.78_{\pm 3.00}$	$24.02_{\pm 2.29}$
MTG (Ours)	DGI	$51.91_{\pm 11.08}$	$73.99_{\pm 3.61}$	$60.73_{\pm 4.22}$	$69.16_{\pm 7.94}$	$73.14_{\pm 7.79}$	$33.52_{\pm 5.07}$	$12.71_{\pm 0.55}$
	GraphMAE	$53.10_{\pm 8.79}$	$57.30_{\pm 6.85}$	$64.65_{\pm 3.69}$	$95.64_{\pm 4.65}$	$90.65_{\pm 10.95}$	$45.18_{\pm 3.35}$	$25.08_{\pm 2.22}$
	EdgePreGPPT	$53.29_{\pm 7.72}$	$37.19_{\pm 7.94}$	$62.22_{\pm 5.54}$	$93.76_{\pm 2.08}$	$90.38_{\pm 14.30}$	$32.06_{\pm 1.70}$	$76.02_{\pm 2.67}$
	EdgePreGprompt	$56.54_{\pm 7.85}$	$55.55_{\pm 11.40}$	$69.94_{\pm 5.00}$	$91.81_{\pm 7.61}$	$92.49_{\pm 12.19}$	$33.62_{\pm 2.80}$	$34.91_{\pm 1.87}$
	GraphCL	$67.85_{\pm 4.39}$	$59.66_{\pm 8.92}$	$56.94_{\pm 2.93}$	$56.56_{\pm 13.47}$	$91.24_{\pm 10.50}$	$30.42_{\pm 3.41}$	$29.39_{\pm 3.96}$
	SimGRACE	$41.64_{\pm 4.50}$	$64.90_{\pm 6.41}$	$50.14_{\pm 3.88}$	$65.15_{\pm 10.07}$	$90.44_{\pm 10.87}$	$27.20_{\pm 3.37}$	$17.00_{\pm 2.72}$

Table 22: AUROC on 5-shot node classification.

Adaptation	Pre-training	Cora	Citeseer	Pubmed	Wisconsin	Texas	Actor	ogbn-arxiv
Supervised	-	$88.39_{\pm 2.56}$	$78.16_{\pm 2.95}$	$84.93_{\pm 2.77}$	$65.96_{\pm 5.03}$	$66.32_{\pm 8.89}$	$51.44_{\pm 0.62}$	$78.83_{\pm 1.37}$
Fine-tuning	DGI	$84.60_{\pm 5.40}$	$74.29_{\pm 3.78}$	$81.26_{\pm 3.74}$	$67.40_{\pm 4.97}$	$66.62_{\pm 8.12}$	$51.78_{\pm 0.48}$	$62.86_{\pm 4.35}$
	GraphMAE	$90.30_{\pm 1.62}$	$77.04_{\pm 3.75}$	$86.09_{\pm 2.54}$	$65.32_{\pm 5.26}$	$63.95_{\pm 9.63}$	$51.73_{\pm 0.46}$	$75.65_{\pm 0.78}$
	EdgePreGPPT	$86.88_{\pm 1.90}$	$75.17_{\pm 2.72}$	$86.23_{\pm 2.32}$	$67.71_{\pm 3.54}$	$69.36_{\pm 8.52}$	$51.34_{\pm 0.65}$	$83.86_{\pm 1.25}$
	EdgePredGprompt	$87.41_{\pm 2.15}$	$73.79_{\pm 5.07}$	$85.66_{\pm 3.26}$	$66.04_{\pm 5.12}$	$66.70_{\pm 8.92}$	$52.40_{\pm 0.91}$	$82.34_{\pm 1.45}$
	GraphCL	$90.49_{\pm 1.49}$	$76.54_{\pm 1.56}$	$83.07_{\pm 3.73}$	$57.71_{\pm 8.14}$	$65.57_{\pm 7.78}$	$52.68_{\pm 0.69}$	$66.70_{\pm 0.95}$
	SimGRACE	$85.45_{\pm 2.29}$	$76.95_{\pm 2.79}$	$80.86_{\pm 3.17}$	$67.73_{\pm 4.47}$	$66.11_{\pm 7.69}$	$52.52_{\pm 0.92}$	$60.36_{\pm 0.72}$
GPPTPrompt	DGI	$76.38_{\pm 6.02}$	$76.59_{\pm 3.18}$	$63.76_{\pm 10.00}$	$69.27_{\pm 1.49}$	$66.52_{\pm 3.33}$	$51.27_{\pm 1.04}$	$50.09_{\pm 0.25}$
	GraphMAE	$84.77_{\pm 3.18}$	$69.69_{\pm 1.62}$	$82.56_{\pm 1.84}$	$66.53_{\pm 4.44}$	$63.43_{\pm 2.99}$	$50.92_{\pm 0.44}$	$63.55_{\pm 2.93}$
	EdgePreGPPT	$69.44_{\pm 2.15}$	$60.14_{\pm 1.47}$	$77.28_{\pm 3.58}$	$62.04_{\pm 3.27}$	$61.22_{\pm 2.13}$	$50.85_{\pm 0.37}$	$85.67_{\pm 0.67}$
	EdgePreGprompt	$69.60_{\pm 0.83}$	$62.44_{\pm 2.17}$	$81.98_{\pm 4.11}$	$63.83_{\pm 2.92}$	$65.86_{\pm 2.34}$	$50.32_{\pm 0.54}$	$71.01_{\pm 5.38}$
	GraphCL	$82.06_{\pm 2.24}$	$61.54_{\pm 2.42}$	$76.26_{\pm 1.36}$	$55.23_{\pm 2.54}$	$62.83_{\pm 4.62}$	$50.87_{\pm 0.69}$	$58.53_{\pm 1.53}$
	SimGRACE	$70.54_{\pm 2.07}$	$64.49_{\pm 1.54}$	$60.67_{\pm 1.80}$	$66.38_{\pm 0.42}$	$62.69_{\pm 1.92}$	$51.00_{\pm 0.54}$	$51.46_{\pm 0.26}$
Gprompt	DGI	$70.76_{\pm 4.67}$	$80.01_{\pm 2.22}$	$54.50_{\pm 2.51}$	$84.53_{\pm 17.37}$	$65.41_{\pm 9.24}$	$54.02_{\pm 2.61}$	$92.66_{\pm 0.73}$
	GraphMAE	$72.59_{\pm 7.56}$	$76.04_{\pm 5.23}$	$82.88_{\pm 2.18}$	$89.40_{\pm 19.62}$	$66.34_{\pm 10.38}$	$56.23_{\pm 2.46}$	$96.45_{\pm 0.14}$
	EdgePreGPPT	$85.57_{\pm 0.39}$	$68.97_{\pm 4.56}$	$76.57_{\pm 1.08}$	$89.57_{\pm 19.71}$	$66.48_{\pm 10.55}$	$58.85_{\pm 2.98}$	$98.40_{\pm 0.08}$
	EdgePreGprompt	$88.49_{\pm 0.88}$	$84.82_{\pm 2.89}$	$83.58_{\pm 1.29}$	$89.71_{\pm 19.78}$	$66.52_{\pm 10.61}$	$52.51_{\pm 2.39}$	$97.79_{\pm 0.15}$
	GraphCL	$88.70_{\pm 2.05}$	$81.00_{\pm 6.61}$	$74.92_{\pm 2.53}$	$83.64_{\pm 16.83}$	$66.33_{\pm 10.37}$	$63.03_{\pm 3.07}$	$94.54_{\pm 0.37}$
	SimGRACE	$78.28_{\pm 3.96}$	$89.06_{\pm 0.86}$	$59.22_{\pm 5.76}$	$86.22_{\pm 18.17}$	$66.68_{\pm 10.79}$	$58.88_{\pm 3.13}$	$95.27_{\pm 0.28}$
All-in-one	DGI	$42.65_{\pm 0.41}$	$57.45_{\pm 7.01}$	$57.38_{\pm 8.87}$	$85.53_{\pm 2.93}$	$68.87_{\pm 6.32}$	$51.24_{\pm 2.32}$	$54.76_{\pm 0.75}$
	GraphMAE	$80.75_{\pm 0.64}$	$57.95_{\pm 1.98}$	$81.32_{\pm 2.25}$	$83.91_{\pm 5.14}$	$72.30_{\pm 3.24}$	$53.43_{\pm 0.71}$	$53.46_{\pm 2.41}$
	EdgePreGPPT	$62.42_{\pm 11.37}$	$62.25_{\pm 7.23}$	$63.45_{\pm 13.43}$	$95.30_{\pm 4.46}$	$78.68_{\pm 10.06}$	$51.90_{\pm 0.91}$	$54.22_{\pm 1.57}$
	EdgePreGprompt	$77.24_{\pm 3.51}$	$64.14_{\pm 1.01}$	$75.97_{\pm 3.33}$	$91.72_{\pm 1.62}$	$80.10_{\pm 10.86}$	$50.40_{\pm 3.16}$	$75.81_{\pm 2.83}$
	GraphCL	$65.99_{\pm 10.83}$	$65.27_{\pm 11.57}$	$69.76_{\pm 4.24}$	$88.87_{\pm 4.34}$	$76.09_{\pm 6.05}$	$54.21_{\pm 2.36}$	$55.77_{\pm 4.78}$
	SimGRACE	$59.53_{\pm 3.53}$	$64.55_{\pm 7.97}$	$47.71_{\pm 3.86}$	$86.55_{\pm 2.81}$	$83.01_{\pm 10.84}$	$52.44_{\pm 1.68}$	$53.56_{\pm 7.86}$
GPF	DGI	$62.20_{\pm 15.24}$	$62.32_{\pm 17.75}$	$67.14_{\pm 11.86}$	$89.86_{\pm 8.25}$	$78.39_{\pm 1.66}$	$65.97_{\pm 7.26}$	$74.59_{\pm 0.97}$
	GraphMAE	$80.34_{\pm 2.03}$	$67.33_{\pm 3.16}$	$86.64_{\pm 2.02}$	$97.72_{\pm 3.33}$	$89.25_{\pm 8.10}$	$73.06_{\pm 2.24}$	$74.39_{\pm 0.76}$
	EdgePreGPPT	$49.95_{\pm 0.10}$	$52.76_{\pm 6.92}$	$74.91_{\pm 5.59}$	$99.98_{\pm 0.03}$	$83.65_{\pm 7.64}$	$59.31_{\pm 5.01}$	$90.96_{\pm 1.32}$
	EdgePreGprompt	$73.56_{\pm 5.48}$	$76.82_{\pm 4.28}$	$85.50_{\pm 1.95}$	$97.15_{\pm 4.98}$	$86.81_{\pm 8.42}$	$60.43_{\pm 6.30}$	$75.36_{\pm 2.61}$
	GraphCL	$79.04_{\pm 5.41}$	$64.25_{\pm 7.12}$	$70.90_{\pm 6.31}$	$73.84_{\pm 18.96}$	$83.62_{\pm 4.75}$	$57.39_{\pm 3.99}$	$75.19_{\pm 0.54}$
	SimGRACE	$60.98_{\pm 9.22}$	$64.93_{\pm 3.80}$	$58.21_{\pm 3.99}$	$90.38_{\pm 1.52}$	$90.43_{\pm 8.89}$	$61.68_{\pm 2.02}$	$60.89_{\pm 3.56}$
GPF-plus	DGI	$60.30_{\pm 14.39}$	$54.40_{\pm 11.09}$	$73.47_{\pm 9.02}$	$95.86_{\pm 3.95}$	$89.08_{\pm 8.43}$	$68.06_{\pm 2.70}$	$67.37_{\pm 3.03}$
	GraphMAE	$89.99_{\pm 1.39}$	$92.09_{\pm 1.11}$	$85.69_{\pm 3.12}$	$99.93_{\pm 0.14}$	$87.29_{\pm 7.92}$	$73.10_{\pm 2.68}$	$81.99_{\pm 0.56}$
	EdgePreGPPT	$65.26_{\pm 14.38}$	$55.10_{\pm 2.95}$	$82.94_{\pm 2.55}$	$99.96_{\pm 0.05}$	$86.89_{\pm 8.72}$	$70.17_{\pm 4.50}$	$93.73_{\pm 1.05}$
	EdgePreGprompt	$91.09_{\pm 1.03}$	$82.22_{\pm 12.12}$	$83.84_{\pm 3.91}$	$99.96_{\pm 0.08}$	$83.88_{\pm 7.87}$	$72.25_{\pm 1.48}$	$81.54_{\pm 1.05}$
	GraphCL	$84.11_{\pm 7.92}$	$61.55_{\pm 15.26}$	$81.38_{\pm 3.75}$	$91.95_{\pm 9.63}$	$86.24_{\pm 7.72}$	$69.74_{\pm 1.87}$	$77.42_{\pm 2.14}$
	SimGRACE	$66.55_{\pm 14.49}$	$63.11_{\pm 14.51}$	$65.07_{\pm 10.97}$	$88.17_{\pm 7.54}$	$89.58_{\pm 9.03}$	$68.81_{\pm 1.60}$	$69.60_{\pm 2.37}$
MTG (Ours)	DGI	$79.10_{\pm 6.06}$	$91.10_{\pm 3.30}$	$82.26_{\pm 3.91}$	$92.82_{\pm 5.80}$	$75.08_{\pm 7.59}$	$65.51_{\pm 1.26}$	$71.53_{\pm 2.45}$
	GraphMAE	$84.04_{\pm 3.89}$	$79.50_{\pm 5.64}$	$81.29_{\pm 3.88}$	$97.38_{\pm 2.96}$	$83.27_{\pm 7.89}$	$72.52_{\pm 1.86}$	$88.91_{\pm 0.39}$
	EdgePreGPPT	$86.80_{\pm 2.58}$	$73.62_{\pm 5.31}$	$84.55_{\pm 3.01}$	$99.96_{\pm 0.04}$	$79.55_{\pm 0.64}$	$63.35_{\pm 0.82}$	$97.59_{\pm 0.21}$
	EdgePreGprompt	$88.07_{\pm 2.30}$	$83.98_{\pm 3.25}$	$87.81_{\pm 2.54}$	$99.31_{\pm 0.75}$	$85.44_{\pm 4.87}$	$63.89_{\pm 1.70}$	$92.11_{\pm 1.84}$
	GraphCL	$91.33_{\pm 3.92}$	$87.10_{\pm 3.73}$	$79.18_{\pm 2.11}$	$78.90_{\pm 8.28}$	$80.32_{\pm 0.89}$	$60.14_{\pm 4.10}$	$84.05_{\pm 0.99}$
	SimGRACE	$78.67_{\pm 3.11}$	$90.21_{\pm 2.24}$	$70.49_{\pm 6.82}$	$86.53_{\pm 9.52}$	$82.97_{\pm 6.07}$	$63.82_{\pm 0.80}$	$64.57_{\pm 5.00}$

Table 23: Accuracy (%) on 1-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	$57.30_{\pm 0.98}$	$47.23_{\pm 0.61}$	$56.36_{\pm 7.97}$	$65.20_{\pm 6.70}$	$20.58_{\pm 2.00}$	$27.08_{\pm 11.94}$	$25.80_{\pm 6.53}$	$55.33_{\pm 6.22}$
Fine-tuning	DGI	$57.32_{\pm 0.90}$	$42.22_{\pm 0.73}$	$60.00_{\pm 4.48}$	$64.13_{\pm 7.90}$	$17.83_{\pm 1.88}$	$29.44_{\pm 9.68}$	$26.48_{\pm 7.61}$	$57.15_{\pm 4.32}$
	GraphMAE	$57.70_{\pm 1.13}$	$48.10_{\pm 0.23}$	$62.40_{\pm 1.94}$	$65.20_{\pm 5.00}$	$22.21_{\pm 2.79}$	$28.47_{\pm 14.72}$	$25.80_{\pm 6.53}$	$53.59_{\pm 6.93}$
	EdgePreGPPT	$57.20_{\pm 0.85}$	$47.14_{\pm 0.55}$	$58.27_{\pm 10.66}$	$64.27_{\pm 4.73}$	$19.79_{\pm 2.17}$	$27.83_{\pm 13.44}$	$34.69_{\pm 8.50}$	$52.82_{\pm 9.38}$
	EdgePreGprompt	$57.35_{\pm 0.92}$	$47.20_{\pm 0.53}$	$61.84_{\pm 2.59}$	$62.67_{\pm 2.67}$	$19.75_{\pm 2.33}$	$27.13_{\pm 12.05}$	$29.44_{\pm 11.20}$	$56.16_{\pm 5.10}$
	GraphCL	$57.75_{\pm 1.02}$	$39.62_{\pm 0.63}$	$63.44_{\pm 3.64}$	$65.07_{\pm 8.38}$	$21.21_{\pm 0.87}$	$53.14_{\pm 21.32}$	$29.07_{\pm 7.00}$	$55.50_{\pm 5.83}$
	SimGRACE	$57.33_{\pm 0.96}$	$46.89_{\pm 0.42}$	$60.07_{\pm 3.21}$	$65.47_{\pm 5.89}$	$19.71_{\pm 1.76}$	$76.19_{\pm 5.41}$	$28.46_{\pm 6.49}$	$53.23_{\pm 9.71}$
GPPTPrompt	DGI	$49.07_{\pm 10.36}$	$39.34_{\pm 9.11}$	$60.81_{\pm 1.55}$	$51.33_{\pm 15.87}$	$20.29_{\pm 1.40}$	$78.23_{\pm 1.38}$	$44.07_{\pm 22.42}$	$53.65_{\pm 10.00}$
	GraphMAE	$50.15_{\pm 0.75}$	$29.46_{\pm 13.65}$	$60.72_{\pm 1.70}$	$44.80_{\pm 15.52}$	$20.37_{\pm 1.96}$	$68.63_{\pm 20.51}$	$55.99_{\pm 19.28}$	$57.69_{\pm 6.89}$
	EdgePreGPPT	$49.38_{\pm 10.29}$	$36.47_{\pm 7.88}$	$60.92_{\pm 2.47}$	$42.80_{\pm 12.98}$	$20.87_{\pm 2.42}$	$73.99_{\pm 9.79}$	$49.81_{\pm 20.17}$	$53.69_{\pm 6.88}$
	EdgePreGprompt	$50.15_{\pm 0.75}$	$40.22_{\pm 9.56}$	$57.03_{\pm 4.55}$	$37.87_{\pm 10.43}$	$22.08_{\pm 3.42}$	$72.28_{\pm 13.22}$	$50.06_{\pm 18.97}$	$55.33_{\pm 8.51}$
	GraphCL	$45.70_{\pm 8.20}$	$47.18_{\pm 5.93}$	$59.24_{\pm 1.01}$	$60.40_{\pm 15.43}$	$21.29_{\pm 3.79}$	$68.36_{\pm 21.05}$	$59.32_{\pm 11.22}$	$56.26_{\pm 8.20}$
	SimGRACE	$46.03_{\pm 10.29}$	$41.11_{\pm 8.47}$	$55.42_{\pm 8.81}$	$52.67_{\pm 17.12}$	$20.83_{\pm 3.47}$	$62.31_{\pm 19.42}$	$59.20_{\pm 15.06}$	$55.88_{\pm 7.81}$
Gprompt	DGI	$50.47_{\pm 10.10}$	$47.29_{\pm 7.78}$	$56.61_{\pm 7.93}$	$63.33_{\pm 14.36}$	$20.50_{\pm 1.79}$	$45.52_{\pm 16.98}$	$55.43_{\pm 13.69}$	$56.18_{\pm 6.13}$
	GraphMAE	$54.75_{\pm 12.43}$	$36.39_{\pm 7.72}$	$57.66_{\pm 12.56}$	$68.80_{\pm 4.76}$	$19.54_{\pm 1.99}$	$43.91_{\pm 6.64}$	$47.16_{\pm 4.72}$	$55.22_{\pm 6.40}$
	EdgePreGPPT	$51.18_{\pm 11.11}$	$46.70_{\pm 5.74}$	$59.17_{\pm 11.26}$	$52.13_{\pm 5.80}$	$19.71_{\pm 4.46}$	$50.08_{\pm 8.00}$	$45.06_{\pm 15.93}$	$51.04_{\pm 4.82}$
	EdgePreGprompt	$51.57_{\pm 11.87}$	$40.53_{\pm 12.02}$	$55.55_{\pm 8.17}$	$73.60_{\pm 4.76}$	$19.67_{\pm 3.08}$	$54.64_{\pm 9.94}$	$51.36_{\pm 15.55}$	$57.20_{\pm 5.54}$
	GraphCL	$50.50_{\pm 10.42}$	$45.54_{\pm 9.05}$	$55.51_{\pm 10.73}$	$56.00_{\pm 13.79}$	$19.83_{\pm 2.19}$	$44.40_{\pm 5.74}$	$46.42_{\pm 20.67}$	$52.65_{\pm 9.17}$
	SimGRACE	$50.40_{\pm 10.54}$	$48.25_{\pm 13.64}$	$57.53_{\pm 11.05}$	$64.67_{\pm 7.92}$	$22.29_{\pm 3.50}$	$47.02_{\pm 5.59}$	$52.90_{\pm 11.76}$	$57.81_{\pm 2.68}$
All-in-one	DGI	$60.07_{\pm 4.81}$	$39.56_{\pm 5.00}$	$62.58_{\pm 7.07}$	$73.87_{\pm 6.13}$	$23.96_{\pm 1.45}$	$50.72_{\pm 9.93}$	$64.38_{\pm 9.32}$	$55.97_{\pm 6.52}$
	GraphMAE	$52.62_{\pm 3.04}$	$40.82_{\pm 14.63}$	$66.49_{\pm 6.26}$	$69.67_{\pm 9.13}$	$23.21_{\pm 1.72}$	$56.68_{\pm 7.38}$	$58.64_{\pm 19.59}$	$58.77_{\pm 1.05}$
	EdgePreGPPT	$59.12_{\pm 0.77}$	$42.74_{\pm 4.65}$	$65.71_{\pm 5.49}$	$75.20_{\pm 6.33}$	$20.92_{\pm 2.04}$	$60.27_{\pm 16.97}$	$59.69_{\pm 9.90}$	$56.24_{\pm 2.46}$
	EdgePreGprompt	$53.78_{\pm 2.82}$	$42.87_{\pm 6.19}$	$61.82_{\pm 7.53}$	$68.27_{\pm 3.88}$	$21.88_{\pm 0.56}$	$49.06_{\pm 5.53}$	$32.65_{\pm 10.08}$	$57.60_{\pm 4.37}$
	GraphCL	$58.75_{\pm 0.80}$	$51.66_{\pm 0.26}$	$64.36_{\pm 7.30}$	$66.00_{\pm 8.79}$	$19.46_{\pm 2.85}$	$52.55_{\pm 13.51}$	$42.65_{\pm 14.43}$	$59.72_{\pm 1.52}$
	SimGRACE	$58.83_{\pm 0.85}$	$47.60_{\pm 0.39}$	$61.17_{\pm 1.73}$	$66.67_{\pm 5.73}$	$22.50_{\pm 1.56}$	$76.14_{\pm 5.51}$	$59.01_{\pm 12.34}$	$58.26_{\pm 1.18}$
GPF	DGI	$52.85_{\pm 8.91}$	$42.75_{\pm 7.10}$	$59.17_{\pm 3.63}$	$63.07_{\pm 7.22}$	$22.00_{\pm 1.25}$	$27.94_{\pm 13.65}$	$70.56_{\pm 15.46}$	$59.36_{\pm 1.18}$
	GraphMAE	$49.27_{\pm 7.77}$	$37.23_{\pm 17.95}$	$58.65_{\pm 8.49}$	$65.73_{\pm 6.91}$	$20.71_{\pm 1.92}$	$40.43_{\pm 10.43}$	$67.84_{\pm 20.28}$	$55.80_{\pm 6.34}$
	EdgePreGPPT	$59.35_{\pm 1.02}$	$37.53_{\pm 5.19}$	$62.54_{\pm 2.55}$	$66.40_{\pm 5.93}$	$22.04_{\pm 1.48}$	$27.40_{\pm 12.58}$	$24.75_{\pm 8.01}$	$43.86_{\pm 5.37}$
	EdgePreGprompt	$59.65_{\pm 5.06}$	$41.44_{\pm 0.52}$	$61.82_{\pm 2.61}$	$68.40_{\pm 5.09}$	$22.17_{\pm 1.48}$	$65.79_{\pm 17.72}$	$60.49_{\pm 24.20}$	$55.56_{\pm 5.07}$
	GraphCL	$57.73_{\pm 0.79}$	$47.42_{\pm 1.22}$	$63.91_{\pm 3.26}$	$59.20_{\pm 7.01}$	$21.13_{\pm 2.11}$	$37.05_{\pm 9.39}$	$71.67_{\pm 14.71}$	$58.43_{\pm 1.17}$
	SimGRACE	$58.30_{\pm 0.77}$	$41.04_{\pm 0.24}$	$63.35_{\pm 3.69}$	$66.93_{\pm 5.05}$	$21.79_{\pm 2.40}$	$33.99_{\pm 11.05}$	$27.47_{\pm 5.92}$	$58.03_{\pm 1.63}$
GPF-plus	DGI	$57.87_{\pm 6.19}$	$43.98_{\pm 7.17}$	$61.26_{\pm 3.06}$	$62.53_{\pm 6.86}$	$18.71_{\pm 1.19}$	$26.70_{\pm 11.19}$	$67.90_{\pm 14.60}$	$48.79_{\pm 9.14}$
	GraphMAE	$55.00_{\pm 5.81}$	$40.32_{\pm 9.49}$	$62.49_{\pm 2.05}$	$62.13_{\pm 8.71}$	$22.92_{\pm 1.64}$	$33.78_{\pm 11.61}$	$68.40_{\pm 20.02}$	$56.31_{\pm 4.93}$
	EdgePreGPPT	$56.65_{\pm 4.08}$	$41.41_{\pm 0.73}$	$63.06_{\pm 2.55}$	$65.07_{\pm 4.41}$	$21.50_{\pm 2.37}$	$27.83_{\pm 11.54}$	$29.57_{\pm 7.69}$	$57.62_{\pm 2.42}$
	EdgePreGprompt	$55.35_{\pm 3.91}$	$40.08_{\pm 2.95}$	$61.33_{\pm 2.81}$	$65.20_{\pm 6.04}$	$19.42_{\pm 1.88}$	$28.85_{\pm 15.48}$	$57.22_{\pm 18.17}$	$55.90_{\pm 5.56}$
	GraphCL	$57.10_{\pm 1.34}$	$46.89_{\pm 0.24}$	$59.75_{\pm 7.95}$	$64.00_{\pm 7.89}$	$18.79_{\pm 1.46}$	$25.90_{\pm 9.58}$	$71.17_{\pm 14.92}$	$57.56_{\pm 2.54}$
	SimGRACE	$57.93_{\pm 1.62}$	$47.24_{\pm 0.29}$	$62.92_{\pm 2.78}$	$61.33_{\pm 3.84}$	$20.29_{\pm 1.62}$	$27.08_{\pm 7.86}$	$28.46_{\pm 6.49}$	$57.11_{\pm 3.42}$
MTG (Ours)	DGI	$59.15_{\pm 5.44}$	$43.46_{\pm 6.83}$	$62.78_{\pm 2.36}$	$65.60_{\pm 7.29}$	$24.71_{\pm 1.88}$	$51.74_{\pm 13.90}$	$74.81_{\pm 13.96}$	$56.39_{\pm 3.27}$
	GraphMAE	$58.10_{\pm 5.72}$	$48.24_{\pm 9.56}$	$59.62_{\pm 6.41}$	$66.93_{\pm 7.03}$	$22.71_{\pm 2.58}$	$58.93_{\pm 12.05}$	$54.07_{\pm 18.34}$	$58.01_{\pm 5.85}$
	EdgePreGPPT	$62.25_{\pm 3.72}$	$45.15_{\pm 6.00}$	$62.71_{\pm 2.30}$	$67.20_{\pm 6.36}$	$26.08_{\pm 4.31}$	$60.16_{\pm 10.63}$	$62.28_{\pm 10.13}$	$56.37_{\pm 8.33}$
	EdgePreGprompt	$59.45_{\pm 5.45}$	$47.72_{\pm 8.45}$	$65.66_{\pm 1.56}$	$75.80_{\pm 5.49}$	$22.29_{\pm 1.94}$	$57.75_{\pm 10.76}$	$49.94_{\pm 9.08}$	$60.68_{\pm 2.42}$
	GraphCL	$57.65_{\pm 7.05}$	$47.81_{\pm 3.73}$	$63.70_{\pm 2.87}$	$66.20_{\pm 7.52}$	$20.96_{\pm 1.97}$	$50.36_{\pm 12.97}$	$51.05_{\pm 15.50}$	$55.46_{\pm 4.77}$
	SimGRACE	$61.82_{\pm 3.49}$	$52.25_{\pm 0.56}$	$66.98_{\pm 2.17}$	$68.87_{\pm 5.01}$	$21.33_{\pm 1.92}$	$78.27_{\pm 2.01}$	$65.68_{\pm 16.41}$	$57.26_{\pm 2.01}$

Table 24: F1-score on 1-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	54.62$_{\pm1.12}$	41.10$_{\pm0.39}$	46.69$_{\pm10.82}$	63.47$_{\pm6.36}$	15.25$_{\pm3.96}$	22.78$_{\pm10.69}$	23.71$_{\pm8.23}$	44.74$_{\pm4.23}$
Fine-tuning	DGI	54.60$_{\pm1.00}$	38.53$_{\pm0.34}$	54.82$_{\pm3.34}$	61.97$_{\pm7.76}$	10.76$_{\pm4.28}$	27.09$_{\pm11.48}$	24.34$_{\pm9.21}$	46.15$_{\pm5.41}$
	GraphMAE	55.20$_{\pm1.24}$	41.71$_{\pm0.17}$	52.05$_{\pm7.26}$	63.41$_{\pm4.44}$	19.17$_{\pm3.42}$	23.63$_{\pm12.40}$	23.71$_{\pm8.23}$	46.25$_{\pm7.84}$
	EdgePreGPPT	54.39$_{\pm0.95}$	41.10$_{\pm0.37}$	55.82$_{\pm10.61}$	60.94$_{\pm3.46}$	12.89$_{\pm3.54}$	23.08$_{\pm11.30}$	33.12$_{\pm7.45}$	36.31$_{\pm5.78}$
	EdgePreGprompt	54.62$_{\pm1.07}$	41.14$_{\pm0.41}$	59.73$_{\pm1.34}$	59.05$_{\pm1.33}$	13.72$_{\pm4.13}$	22.91$_{\pm10.95}$	27.09$_{\pm12.51}$	45.49$_{\pm4.58}$
	GraphCL	55.24$_{\pm1.07}$	36.27$_{\pm0.63}$	56.25$_{\pm7.55}$	63.37$_{\pm8.64}$	16.78$_{\pm1.91}$	39.11$_{\pm4.29}$	27.67$_{\pm8.70}$	48.68$_{\pm6.42}$
	SimGRACE	54.69$_{\pm1.09}$	40.92$_{\pm0.37}$	52.67$_{\pm7.14}$	63.70$_{\pm5.32}$	14.15$_{\pm2.49}$	45.06$_{\pm1.93}$	27.05$_{\pm8.20}$	37.84$_{\pm7.14}$
GPPTPrompt	DGI	41.17$_{\pm12.71}$	27.05$_{\pm13.23}$	46.05$_{\pm10.61}$	41.76$_{\pm15.84}$	17.26$_{\pm2.39}$	44.68$_{\pm1.17}$	33.93$_{\pm14.06}$	43.61$_{\pm5.55}$
	GraphMAE	33.40$_{\pm0.33}$	14.61$_{\pm5.30}$	46.64$_{\pm11.32}$	39.44$_{\pm18.49}$	17.61$_{\pm2.22}$	40.48$_{\pm7.24}$	44.38$_{\pm9.91}$	50.34$_{\pm5.80}$
	EdgePreGPPT	44.16$_{\pm6.70}$	21.35$_{\pm9.96}$	47.07$_{\pm11.95}$	38.03$_{\pm16.75}$	17.02$_{\pm2.90}$	43.67$_{\pm0.88}$	40.73$_{\pm9.69}$	51.50$_{\pm6.54}$
	EdgePreGprompt	33.40$_{\pm0.33}$	18.92$_{\pm3.19}$	43.34$_{\pm8.00}$	31.42$_{\pm13.60}$	19.87$_{\pm2.99}$	43.73$_{\pm0.75}$	41.78$_{\pm10.08}$	45.12$_{\pm7.86}$
	GraphCL	39.08$_{\pm10.25}$	42.87$_{\pm7.70}$	41.15$_{\pm7.80}$	53.15$_{\pm16.82}$	19.62$_{\pm3.92}$	40.11$_{\pm7.97}$	45.58$_{\pm3.38}$	50.02$_{\pm8.57}$
	SimGRACE	43.18$_{\pm8.95}$	33.88$_{\pm13.05}$	40.87$_{\pm7.11}$	46.54$_{\pm18.13}$	18.87$_{\pm3.37}$	41.86$_{\pm9.35}$	49.40$_{\pm8.41}$	46.82$_{\pm6.89}$
Gprompt	DGI	48.68$_{\pm9.78}$	42.80$_{\pm9.19}$	55.95$_{\pm7.78}$	61.15$_{\pm13.98}$	18.68$_{\pm2.94}$	38.30$_{\pm12.89}$	44.61$_{\pm5.71}$	49.81$_{\pm1.61}$
	GraphMAE	52.10$_{\pm13.61}$	17.64$_{\pm2.56}$	55.24$_{\pm12.01}$	64.58$_{\pm3.26}$	18.36$_{\pm2.20}$	42.68$_{\pm5.98}$	43.38$_{\pm3.73}$	50.47$_{\pm3.41}$
	EdgePreGPPT	49.33$_{\pm10.58}$	43.20$_{\pm8.14}$	58.30$_{\pm10.88}$	50.70$_{\pm6.00}$	18.20$_{\pm5.07}$	44.54$_{\pm3.28}$	39.06$_{\pm9.23}$	50.78$_{\pm5.00}$
	EdgePreGprompt	50.43$_{\pm11.93}$	36.62$_{\pm12.55}$	54.29$_{\pm7.32}$	71.38$_{\pm3.64}$	17.17$_{\pm4.25}$	46.26$_{\pm5.14}$	43.73$_{\pm9.27}$	48.18$_{\pm4.55}$
	GraphCL	48.91$_{\pm10.12}$	40.78$_{\pm10.09}$	53.98$_{\pm9.93}$	53.39$_{\pm14.36}$	18.26$_{\pm2.77}$	42.26$_{\pm4.15}$	38.58$_{\pm11.82}$	50.85$_{\pm8.14}$
	SimGRACE	48.78$_{\pm10.20}$	43.35$_{\pm10.75}$	55.51$_{\pm10.10}$	60.58$_{\pm6.08}$	19.52$_{\pm3.36}$	44.68$_{\pm4.01}$	44.81$_{\pm6.73}$	52.80$_{\pm3.60}$
All-in-one	DGI	56.82$_{\pm6.07}$	35.40$_{\pm5.55}$	60.66$_{\pm6.94}$	67.26$_{\pm6.79}$	14.48$_{\pm3.58}$	44.46$_{\pm4.45}$	54.86$_{\pm6.67}$	48.28$_{\pm7.29}$
	GraphMAE	45.83$_{\pm5.38}$	18.76$_{\pm5.47}$	64.27$_{\pm4.78}$	69.07$_{\pm9.55}$	19.66$_{\pm3.11}$	49.40$_{\pm3.96}$	40.11$_{\pm6.22}$	56.70$_{\pm1.89}$
	EdgePreGPPT	57.29$_{\pm0.74}$	37.07$_{\pm5.56}$	64.68$_{\pm5.35}$	70.35$_{\pm6.20}$	12.95$_{\pm3.18}$	49.62$_{\pm10.42}$	49.67$_{\pm4.66}$	55.10$_{\pm1.49}$
	EdgePreGprompt	48.44$_{\pm4.51}$	34.64$_{\pm5.55}$	60.04$_{\pm8.57}$	63.74$_{\pm6.21}$	12.50$_{\pm3.12}$	45.57$_{\pm5.70}$	30.69$_{\pm11.20}$	48.13$_{\pm4.31}$
	GraphCL	56.83$_{\pm0.76}$	47.78$_{\pm0.10}$	62.99$_{\pm7.19}$	60.07$_{\pm12.25}$	12.01$_{\pm5.16}$	46.65$_{\pm6.50}$	39.12$_{\pm10.09}$	43.55$_{\pm8.21}$
	SimGRACE	56.88$_{\pm0.80}$	41.64$_{\pm0.13}$	53.18$_{\pm7.57}$	59.95$_{\pm11.21}$	12.23$_{\pm2.42}$	45.03$_{\pm1.86}$	46.98$_{\pm5.91}$	39.55$_{\pm5.05}$
GPF	DGI	50.50$_{\pm7.88}$	34.56$_{\pm6.02}$	49.27$_{\pm10.07}$	62.02$_{\pm6.87}$	15.08$_{\pm1.44}$	23.42$_{\pm11.99}$	44.77$_{\pm3.37}$	39.53$_{\pm5.01}$
	GraphMAE	42.98$_{\pm8.29}$	17.19$_{\pm6.84}$	52.62$_{\pm9.40}$	59.14$_{\pm10.41}$	13.10$_{\pm3.28}$	37.75$_{\pm10.47}$	41.84$_{\pm7.65}$	48.52$_{\pm7.11}$
	EdgePreGPPT	55.67$_{\pm0.84}$	34.09$_{\pm5.32}$	57.01$_{\pm5.79}$	58.18$_{\pm3.05}$	16.57$_{\pm1.64}$	22.68$_{\pm10.50}$	21.77$_{\pm9.16}$	34.22$_{\pm10.09}$
	EdgePreGprompt	56.22$_{\pm6.17}$	38.14$_{\pm0.44}$	56.91$_{\pm6.21}$	63.90$_{\pm4.05}$	17.34$_{\pm2.45}$	43.08$_{\pm4.88}$	39.86$_{\pm11.54}$	47.44$_{\pm4.83}$
	GraphCL	55.23$_{\pm0.77}$	38.04$_{\pm0.46}$	56.08$_{\pm7.40}$	57.99$_{\pm6.96}$	15.97$_{\pm3.75}$	35.89$_{\pm9.97}$	48.83$_{\pm5.30}$	40.86$_{\pm4.89}$
	SimGRACE	56.19$_{\pm0.68}$	37.69$_{\pm0.21}$	55.50$_{\pm9.14}$	58.38$_{\pm2.44}$	14.39$_{\pm3.45}$	31.82$_{\pm12.08}$	26.02$_{\pm7.56}$	39.13$_{\pm4.20}$
GPF-plus	DGI	53.13$_{\pm10.49}$	37.59$_{\pm1.42}$	54.74$_{\pm7.01}$	61.19$_{\pm6.31}$	13.03$_{\pm1.21}$	22.69$_{\pm10.52}$	46.57$_{\pm4.62}$	33.21$_{\pm4.87}$
	GraphMAE	48.23$_{\pm10.59}$	18.94$_{\pm3.14}$	52.88$_{\pm6.59}$	61.01$_{\pm8.89}$	18.39$_{\pm2.76}$	30.90$_{\pm11.56}$	44.87$_{\pm9.19}$	46.24$_{\pm4.86}$
	EdgePreGPPT	50.88$_{\pm7.65}$	37.29$_{\pm2.27}$	57.58$_{\pm7.28}$	62.03$_{\pm2.92}$	17.40$_{\pm2.32}$	24.17$_{\pm10.38}$	28.06$_{\pm9.23}$	40.06$_{\pm6.06}$
	EdgePreGprompt	50.07$_{\pm6.94}$	37.43$_{\pm1.84}$	54.79$_{\pm2.74}$	63.20$_{\pm5.31}$	14.44$_{\pm1.44}$	24.60$_{\pm14.35}$	42.77$_{\pm1.25}$	45.59$_{\pm4.64}$
	GraphCL	54.24$_{\pm1.69}$	38.53$_{\pm0.20}$	57.54$_{\pm6.94}$	62.31$_{\pm7.93}$	13.66$_{\pm2.70}$	22.64$_{\pm10.42}$	48.71$_{\pm5.51}$	39.37$_{\pm4.68}$
	SimGRACE	55.55$_{\pm2.03}$	41.24$_{\pm0.31}$	54.80$_{\pm6.88}$	59.38$_{\pm2.64}$	14.74$_{\pm2.94}$	24.79$_{\pm9.43}$	27.05$_{\pm8.20}$	39.51$_{\pm4.96}$
MTG (Ours)	DGI	55.05$_{\pm9.42}$	37.61$_{\pm6.03}$	59.71$_{\pm3.13}$	62.66$_{\pm4.88}$	17.66$_{\pm1.52}$	45.38$_{\pm7.25}$	55.35$_{\pm12.83}$	46.44$_{\pm6.77}$
	GraphMAE	50.89$_{\pm8.77}$	23.24$_{\pm3.16}$	53.89$_{\pm8.92}$	64.49$_{\pm5.74}$	16.32$_{\pm2.69}$	52.25$_{\pm6.37}$	47.19$_{\pm12.10}$	48.40$_{\pm6.40}$
	EdgePreGPPT	59.17$_{\pm2.24}$	36.30$_{\pm5.45}$	58.00$_{\pm3.48}$	65.60$_{\pm5.84}$	21.49$_{\pm5.92}$	51.43$_{\pm4.26}$	52.99$_{\pm3.80}$	52.20$_{\pm10.61}$
	EdgePreGprompt	57.08$_{\pm1.85}$	35.13$_{\pm8.96}$	59.63$_{\pm0.53}$	71.06$_{\pm6.36}$	15.13$_{\pm1.72}$	49.33$_{\pm6.03}$	45.04$_{\pm5.12}$	55.42$_{\pm6.62}$
	GraphCL	52.24$_{\pm8.19}$	38.90$_{\pm3.82}$	60.53$_{\pm1.86}$	63.70$_{\pm7.49}$	13.30$_{\pm1.86}$	42.60$_{\pm10.31}$	45.14$_{\pm10.05}$	50.07$_{\pm3.75}$
	SimGRACE	55.50$_{\pm4.79}$	45.12$_{\pm0.20}$	64.12$_{\pm5.77}$	61.23$_{\pm3.71}$	12.03$_{\pm1.20}$	55.16$_{\pm7.14}$	50.63$_{\pm8.77}$	45.36$_{\pm8.02}$

Table 25: AUROC on 1-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	67.05$_{\pm1.01}$	54.23$_{\pm0.34}$	57.88$_{\pm1.72}$	71.68$_{\pm1.25}$	53.49$_{\pm1.11}$	48.39$_{\pm1.89}$	51.13$_{\pm1.38}$	49.60$_{\pm2.94}$
Fine-tuning	DGI	67.06$_{\pm1.00}$	74.13$_{\pm1.27}$	56.87$_{\pm2.74}$	71.45$_{\pm1.55}$	51.76$_{\pm2.18}$	52.35$_{\pm7.49}$	53.79$_{\pm1.12}$	49.32$_{\pm3.05}$
	GraphMAE	66.91$_{\pm1.18}$	55.16$_{\pm0.20}$	59.87$_{\pm0.78}$	69.81$_{\pm1.39}$	53.43$_{\pm0.98}$	49.46$_{\pm1.08}$	51.01$_{\pm0.87}$	50.66$_{\pm4.90}$
	EdgePreGPPT	67.50$_{\pm1.06}$	54.41$_{\pm0.31}$	58.29$_{\pm3.43}$	72.42$_{\pm1.64}$	53.62$_{\pm1.61}$	46.88$_{\pm1.16}$	48.35$_{\pm3.28}$	52.27$_{\pm1.89}$
	EdgePreGprompt	66.92$_{\pm0.96}$	54.36$_{\pm0.30}$	58.24$_{\pm1.27}$	71.58$_{\pm0.75}$	53.30$_{\pm1.75}$	49.62$_{\pm0.75}$	55.48$_{\pm2.09}$	49.97$_{\pm3.16}$
	GraphCL	67.11$_{\pm1.09}$	68.95$_{\pm1.48}$	59.68$_{\pm1.80}$	71.07$_{\pm1.87}$	54.35$_{\pm0.92}$	51.84$_{\pm2.24}$	53.34$_{\pm2.75}$	53.14$_{\pm5.70}$
	SimGRACE	66.95$_{\pm0.99}$	53.79$_{\pm0.47}$	59.80$_{\pm1.06}$	71.49$_{\pm1.82}$	54.20$_{\pm1.36}$	49.38$_{\pm1.24}$	54.01$_{\pm2.14}$	49.20$_{\pm0.84}$
GPPTPrompt	DGI	48.58$_{\pm10.92}$	58.79$_{\pm4.61}$	71.14$_{\pm3.06}$	51.11$_{\pm31.50}$	53.09$_{\pm1.46}$	53.40$_{\pm0.84}$	48.98$_{\pm7.81}$	53.82$_{\pm11.83}$
	GraphMAE	50.27$_{\pm0.74}$	51.39$_{\pm4.26}$	70.25$_{\pm3.45}$	36.15$_{\pm27.67}$	52.89$_{\pm1.24}$	53.91$_{\pm1.86}$	54.53$_{\pm11.66}$	58.36$_{\pm8.51}$
	EdgePreGPPT	48.81$_{\pm11.41}$	56.39$_{\pm0.58}$	70.36$_{\pm3.79}$	33.54$_{\pm24.51}$	52.98$_{\pm1.23}$	50.31$_{\pm5.34}$	51.48$_{\pm6.83}$	54.80$_{\pm8.30}$
	EdgePreGprompt	50.27$_{\pm0.74}$	54.48$_{\pm1.95}$	65.03$_{\pm10.25}$	24.49$_{\pm23.32}$	53.83$_{\pm2.13}$	51.51$_{\pm2.95}$	53.38$_{\pm9.22}$	56.26$_{\pm10.88}$
	GraphCL	45.42$_{\pm8.63}$	67.97$_{\pm6.02}$	69.82$_{\pm6.05}$	65.58$_{\pm22.89}$	53.36$_{\pm2.88}$	54.61$_{\pm3.27}$	51.68$_{\pm3.28}$	56.31$_{\pm9.89}$
	SimGRACE	45.58$_{\pm10.97}$	61.78$_{\pm7.34}$	62.48$_{\pm15.87}$	48.86$_{\pm29.41}$	53.23$_{\pm2.42}$	55.27$_{\pm7.83}$	58.65$_{\pm6.19}$	56.15$_{\pm9.67}$
Gprompt	DGI	53.74$_{\pm19.12}$	73.04$_{\pm8.22}$	60.36$_{\pm9.11}$	69.64$_{\pm13.88}$	55.68$_{\pm2.50}$	49.63$_{\pm0.56}$	49.08$_{\pm3.10}$	52.16$_{\pm4.08}$
	GraphMAE	51.48$_{\pm9.49}$	44.86$_{\pm0.14}$	55.65$_{\pm7.44}$	70.27$_{\pm4.19}$	55.28$_{\pm3.35}$	56.97$_{\pm3.76}$	51.24$_{\pm5.39}$	52.47$_{\pm6.09}$
	EdgePreGPPT	55.21$_{\pm19.21}$	69.85$_{\pm10.59}$	60.31$_{\pm14.12}$	59.50$_{\pm10.08}$	53.75$_{\pm2.67}$	53.86$_{\pm6.62}$	49.52$_{\pm5.86}$	52.73$_{\pm6.44}$
	EdgePreGprompt	53.58$_{\pm19.20}$	73.09$_{\pm5.29}$	57.72$_{\pm7.86}$	79.17$_{\pm2.18}$	55.53$_{\pm3.27}$	48.74$_{\pm2.54}$	50.81$_{\pm7.50}$	54.96$_{\pm5.21}$
	GraphCL	53.54$_{\pm18.02}$	66.59$_{\pm11.21}$	57.80$_{\pm12.67}$	71.02$_{\pm3.35}$	55.06$_{\pm3.09}$	54.08$_{\pm4.37}$	47.73$_{\pm6.12}$	52.92$_{\pm6.67}$
	SimGRACE	53.21$_{\pm18.90}$	66.59$_{\pm8.84}$	59.42$_{\pm13.20}$	66.20$_{\pm4.91}$	55.23$_{\pm3.64}$	53.95$_{\pm3.70}$	49.34$_{\pm3.63}$	52.15$_{\pm6.51}$
All-in-one	DGI	69.12$_{\pm1.04}$	66.27$_{\pm11.54}$	75.07$_{\pm0.76}$	83.44$_{\pm0.78}$	57.42$_{\pm1.15}$	50.21$_{\pm1.87}$	60.32$_{\pm16.19}$	60.56$_{\pm0.91}$
	GraphMAE	65.65$_{\pm1.22}$	50.00$_{\pm0.00}$	73.97$_{\pm0.80}$	83.95$_{\pm0.75}$	54.66$_{\pm0.90}$	57.09$_{\pm6.45}$	44.91$_{\pm18.23}$	59.65$_{\pm2.74}$
	EdgePreGPPT	65.44$_{\pm0.92}$	64.21$_{\pm9.84}$	75.51$_{\pm0.71}$	83.67$_{\pm0.86}$	55.67$_{\pm1.65}$	61.50$_{\pm1.13}$	57.89$_{\pm11.17}$	57.44$_{\pm0.59}$
	EdgePreGprompt	68.48$_{\pm1.11}$	51.35$_{\pm0.30}$	73.73$_{\pm0.94}$	77.05$_{\pm0.84}$	55.15$_{\pm0.59}$	57.02$_{\pm10.92}$	51.53$_{\pm1.13}$	56.95$_{\pm4.42}$
	GraphCL	65.20$_{\pm1.06}$	63.82$_{\pm0.26}$	77.76$_{\pm0.74}$	76.57$_{\pm1.58}$	53.58$_{\pm1.38}$	66.43$_{\pm2.88}$	49.20$_{\pm7.86}$	55.04$_{\pm1.51}$
	SimGRACE	66.33$_{\pm0.99}$	53.09$_{\pm0.21}$	57.32$_{\pm1.26}$	73.89$_{\pm0.92}$	54.75$_{\pm0.64}$	45.81$_{\pm0.74}$	56.09$_{\pm10.92}$	50.64$_{\pm1.32}$
GPF	DGI	55.96$_{\pm12.77}$	50.91$_{\pm1.05}$	59.34$_{\pm0.55}$	71.40$_{\pm1.24}$	53.13$_{\pm1.13}$	48.97$_{\pm0.73}$	47.84$_{\pm2.93}$	49.47$_{\pm0.63}$
	GraphMAE	61.68$_{\pm14.66}$	50.00$_{\pm0.00}$	59.21$_{\pm1.29}$	73.95$_{\pm2.06}$	51.73$_{\pm1.39}$	52.21$_{\pm1.49}$	46.87$_{\pm1.80}$	49.36$_{\pm4.14}$
	EdgePreGPPT	69.73$_{\pm3.23}$	64.68$_{\pm11.19}$	61.63$_{\pm2.41}$	68.41$_{\pm1.95}$	54.48$_{\pm1.10}$	48.80$_{\pm2.39}$	51.30$_{\pm2.60}$	49.08$_{\pm3.86}$
	EdgePreGprompt	69.06$_{\pm1.34}$	71.37$_{\pm1.86}$	58.36$_{\pm1.20}$	74.51$_{\pm1.18}$	54.01$_{\pm0.86}$	50.42$_{\pm0.62}$	58.34$_{\pm0.87}$	52.14$_{\pm2.73}$
	GraphCL	67.60$_{\pm1.02}$	54.30$_{\pm0.39}$	59.15$_{\pm1.73}$	68.50$_{\pm2.13}$	53.76$_{\pm1.38}$	59.21$_{\pm1.87}$	56.57$_{\pm1.65}$	49.30$_{\pm1.63}$
	SimGRACE	66.95$_{\pm1.01}$	73.55$_{\pm0.37}$	60.00$_{\pm1.43}$	68.25$_{\pm1.28}$	55.25$_{\pm1.53}$	55.80$_{\pm0.69}$	52.30$_{\pm1.91}$	46.17$_{\pm0.78}$
GPF-plus	DGI	68.77$_{\pm1.50}$	72.06$_{\pm4.49}$	58.37$_{\pm2.32}$	70.12$_{\pm1.06}$	53.83$_{\pm0.61}$	48.71$_{\pm0.50}$	51.49$_{\pm1.31}$	48.76$_{\pm1.56}$
	GraphMAE	68.98$_{\pm1.49}$	50.00$_{\pm0.00}$	59.58$_{\pm1.13}$	72.03$_{\pm2.16}$	52.68$_{\pm1.28}$	47.21$_{\pm1.67}$	53.26$_{\pm1.91}$	49.42$_{\pm2.18}$
	EdgePreGPPT	69.57$_{\pm2.53}$	69.92$_{\pm1.65}$	63.58$_{\pm3.31}$	72.02$_{\pm0.86}$	55.66$_{\pm1.62}$	45.11$_{\pm0.89}$	55.06$_{\pm4.29}$	44.29$_{\pm3.43}$
	EdgePreGprompt	68.39$_{\pm1.07}$	69.85$_{\pm1.36}$	57.57$_{\pm1.65}$	73.37$_{\pm1.78}$	52.19$_{\pm1.18}$	51.14$_{\pm2.27}$	56.22$_{\pm5.10}$	49.74$_{\pm1.32}$
	GraphCL	67.01$_{\pm1.22}$	54.03$_{\pm0.40}$	59.78$_{\pm1.47}$	71.25$_{\pm1.88}$	52.59$_{\pm2.06}$	51.25$_{\pm2.24}$	56.70$_{\pm1.44}$	48.94$_{\pm2.26}$
	SimGRACE	67.19$_{\pm1.54}$	53.04$_{\pm0.17}$	59.66$_{\pm1.32}$	68.33$_{\pm0.76}$	52.47$_{\pm1.76}$	51.79$_{\pm1.73}$	53.39$_{\pm2.81}$	49.75$_{\pm1.52}$
MTG (Ours)	DGI	69.97$_{\pm1.05}$	61.26$_{\pm2.80}$	61.95$_{\pm1.09}$	73.70$_{\pm1.07}$	59.37$_{\pm2.54}$	57.52$_{\pm2.30}$	68.76$_{\pm7.78}$	52.23$_{\pm1.11}$
	GraphMAE	69.21$_{\pm1.04}$	50.00$_{\pm0.00}$	60.01$_{\pm2.22}$	72.03$_{\pm2.16}$	55.09$_{\pm2.51}$	63.01$_{\pm2.02}$	53.68$_{\pm11.92}$	53.04$_{\pm3.88}$
	EdgePreGPPT	71.26$_{\pm1.96}$	60.93$_{\pm1.55}$	63.48$_{\pm0.87}$	75.61$_{\pm1.11}$	62.17$_{\pm2.68}$	59.42$_{\pm3.73}$	64.84$_{\pm5.57}$	54.39$_{\pm8.08}$
	EdgePreGprompt	70.01$_{\pm1.09}$	61.59$_{\pm1.56}$	65.34$_{\pm0.71}$	83.72$_{\pm1.39}$	54.59$_{\pm1.13}$	52.65$_{\pm2.64}$	53.32$_{\pm7.33}$	60.34$_{\pm2.66}$
	GraphCL	64.55$_{\pm4.07}$	54.07$_{\pm3.48}$	63.22$_{\pm1.10}$	75.16$_{\pm1.93}$	54.42$_{\pm1.74}$	52.51$_{\pm5.30}$	51.56$_{\pm10.43}$	53.42$_{\pm8.08}$
	SimGRACE	68.50$_{\pm3.09}$	72.18$_{\pm2.16}$	64.54$_{\pm4.41}$	69.65$_{\pm1.12}$	52.36$_{\pm0.99}$	65.39$_{\pm2.50}$	55.74$_{\pm9.61}$	51.53$_{\pm4.48}$

Table 26: Accuracy (%) on 3-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	$53.33_{\pm 6.61}$	$50.77_{\pm 2.44}$	$61.33_{\pm 2.89}$	$59.47_{\pm 8.34}$	$15.96_{\pm 1.64}$	$65.15_{\pm 18.61}$	$52.35_{\pm 8.12}$	$59.77_{\pm 1.10}$
Fine-tuning	DGI	$53.33_{\pm 6.61}$	$56.10_{\pm 3.46}$	$61.33_{\pm 2.75}$	$59.87_{\pm 8.78}$	$21.71_{\pm 0.81}$	$51.96_{\pm 13.00}$	$52.22_{\pm 10.64}$	$59.70_{\pm 0.98}$
	GraphMAE	$53.33_{\pm 6.61}$	$49.11_{\pm 16.81}$	$61.19_{\pm 1.57}$	$44.00_{\pm 13.56}$	$15.04_{\pm 1.86}$	$60.11_{\pm 18.73}$	$38.52_{\pm 17.15}$	$57.90_{\pm 2.69}$
	EdgePreGPPT	$63.43_{\pm 2.65}$	$47.40_{\pm 15.87}$	$62.72_{\pm 2.39}$	$43.47_{\pm 13.44}$	$21.96_{\pm 2.45}$	$49.81_{\pm 9.44}$	$43.70_{\pm 17.89}$	$58.94_{\pm 0.66}$
	EdgePreGprompt	$53.33_{\pm 6.61}$	$54.67_{\pm 1.34}$	$60.67_{\pm 2.32}$	$59.20_{\pm 7.05}$	$17.58_{\pm 1.45}$	$35.82_{\pm 14.38}$	$29.07_{\pm 9.83}$	$59.45_{\pm 9.10}$
	GraphCL	$62.22_{\pm 1.38}$	$55.27_{\pm 2.61}$	$62.07_{\pm 2.39}$	$54.80_{\pm 5.97}$	$22.00_{\pm 1.71}$	$31.90_{\pm 8.27}$	$51.23_{\pm 11.67}$	$55.54_{\pm 6.26}$
	SimGRACE	$66.10_{\pm 0.70}$	$55.38_{\pm 3.58}$	$60.09_{\pm 0.63}$	$54.00_{\pm 7.03}$	$22.71_{\pm 0.86}$	$69.97_{\pm 13.89}$	$36.67_{\pm 1.80}$	$58.17_{\pm 2.79}$
GPPTPrompt	DGI	$50.33_{\pm 0.92}$	$38.40_{\pm 8.13}$	$60.36_{\pm 8.99}$	$64.13_{\pm 18.31}$	$17.67_{\pm 2.05}$	$56.84_{\pm 28.02}$	$69.57_{\pm 19.07}$	$56.11_{\pm 7.51}$
	GraphMAE	$50.23_{\pm 0.95}$	$36.37_{\pm 7.89}$	$56.94_{\pm 6.67}$	$47.87_{\pm 17.55}$	$17.63_{\pm 1.97}$	$69.38_{\pm 19.31}$	$65.19_{\pm 17.99}$	$56.50_{\pm 8.48}$
	EdgePreGPPT	$59.48_{\pm 5.42}$	$38.45_{\pm 9.13}$	$64.74_{\pm 1.99}$	$52.00_{\pm 16.93}$	$19.12_{\pm 2.43}$	$69.87_{\pm 18.34}$	$70.93_{\pm 16.35}$	$52.14_{\pm 6.23}$
	EdgePreGprompt	$51.85_{\pm 2.10}$	$36.11_{\pm 7.73}$	$60.76_{\pm 1.52}$	$62.00_{\pm 18.69}$	$18.71_{\pm 5.10}$	$52.17_{\pm 15.58}$	$67.47_{\pm 10.95}$	$57.94_{\pm 7.47}$
	GraphCL	$50.43_{\pm 11.80}$	$50.88_{\pm 6.31}$	$60.31_{\pm 0.57}$	$48.93_{\pm 18.86}$	$17.25_{\pm 1.19}$	$71.90_{\pm 14.28}$	$53.33_{\pm 16.88}$	$56.28_{\pm 7.21}$
	SimGRACE	$50.12_{\pm 12.86}$	$41.87_{\pm 8.73}$	$55.93_{\pm 6.16}$	$57.87_{\pm 20.52}$	$15.62_{\pm 2.32}$	$58.28_{\pm 20.18}$	$48.64_{\pm 11.03}$	$59.00_{\pm 6.34}$
Gprompt	DGI	$58.95_{\pm 9.88}$	$55.27_{\pm 8.86}$	$62.43_{\pm 4.09}$	$54.93_{\pm 17.15}$	$20.50_{\pm 2.36}$	$50.29_{\pm 7.71}$	$49.69_{\pm 6.79}$	$53.84_{\pm 5.72}$
	GraphMAE	$59.17_{\pm 10.00}$	$36.11_{\pm 7.73}$	$61.98_{\pm 4.45}$	$64.40_{\pm 16.46}$	$21.42_{\pm 2.71}$	$44.83_{\pm 9.16}$	$48.15_{\pm 8.09}$	$52.89_{\pm 5.27}$
	EdgePreGPPT	$64.35_{\pm 1.21}$	$53.20_{\pm 7.90}$	$64.94_{\pm 2.92}$	$53.60_{\pm 14.41}$	$18.50_{\pm 3.69}$	$45.47_{\pm 7.03}$	$54.63_{\pm 2.95}$	$53.61_{\pm 2.31}$
	EdgePreGprompt	$59.30_{\pm 10.17}$	$54.95_{\pm 9.47}$	$62.02_{\pm 3.15}$	$66.53_{\pm 14.84}$	$21.42_{\pm 3.01}$	$50.56_{\pm 9.27}$	$49.88_{\pm 12.32}$	$52.55_{\pm 6.16}$
	GraphCL	$59.85_{\pm 9.50}$	$52.52_{\pm 9.71}$	$58.49_{\pm 9.20}$	$52.40_{\pm 20.58}$	$21.42_{\pm 0.77}$	$48.15_{\pm 9.42}$	$54.26_{\pm 9.10}$	$55.61_{\pm 3.21}$
	SimGRACE	$60.00_{\pm 9.95}$	$53.45_{\pm 7.18}$	$60.27_{\pm 4.44}$	$56.40_{\pm 13.37}$	$22.08_{\pm 3.57}$	$51.53_{\pm 13.08}$	$43.21_{\pm 8.84}$	$55.99_{\pm 7.53}$
All-in-one	DGI	$64.28_{\pm 0.75}$	$52.63_{\pm 8.14}$	$69.84_{\pm 6.02}$	$75.73_{\pm 6.05}$	$22.87_{\pm 0.93}$	$52.17_{\pm 12.81}$	$59.81_{\pm 15.62}$	$54.95_{\pm 6.52}$
	GraphMAE	$63.88_{\pm 0.73}$	$52.09_{\pm 0.33}$	$65.69_{\pm 3.31}$	$72.00_{\pm 9.11}$	$21.04_{\pm 2.51}$	$53.83_{\pm 7.02}$	$61.98_{\pm 11.32}$	$56.56_{\pm 4.54}$
	EdgePreGPPT	$63.80_{\pm 1.07}$	$55.73_{\pm 3.59}$	$65.62_{\pm 7.36}$	$80.00_{\pm 5.67}$	$22.17_{\pm 2.17}$	$50.19_{\pm 10.89}$	$54.26_{\pm 13.30}$	$52.19_{\pm 6.17}$
	EdgePreGprompt	$63.90_{\pm 1.57}$	$51.69_{\pm 8.39}$	$61.30_{\pm 2.47}$	$76.40_{\pm 1.96}$	$23.25_{\pm 1.11}$	$60.21_{\pm 8.86}$	$54.44_{\pm 17.27}$	$58.96_{\pm 5.93}$
	GraphCL	$65.67_{\pm 0.58}$	$57.12_{\pm 1.99}$	$65.57_{\pm 2.24}$	$59.20_{\pm 12.95}$	$23.96_{\pm 0.62}$	$52.17_{\pm 14.65}$	$58.64_{\pm 4.86}$	$52.65_{\pm 5.87}$
	SimGRACE	$64.20_{\pm 1.29}$	$55.48_{\pm 3.48}$	$62.36_{\pm 1.86}$	$55.20_{\pm 11.93}$	$22.58_{\pm 1.18}$	$66.06_{\pm 18.23}$	$61.30_{\pm 16.21}$	$53.23_{\pm 6.95}$
GPF	DGI	$63.53_{\pm 2.47}$	$49.84_{\pm 7.48}$	$61.39_{\pm 2.63}$	$48.67_{\pm 15.53}$	$16.63_{\pm 3.49}$	$65.31_{\pm 19.45}$	$61.79_{\pm 21.19}$	$59.07_{\pm 0.65}$
	GraphMAE	$62.80_{\pm 2.90}$	$37.01_{\pm 13.81}$	$62.72_{\pm 3.07}$	$55.87_{\pm 12.48}$	$18.29_{\pm 2.39}$	$53.51_{\pm 13.09}$	$51.91_{\pm 8.73}$	$57.15_{\pm 5.68}$
	EdgePreGPPT	$65.25_{\pm 2.65}$	$51.91_{\pm 8.13}$	$63.35_{\pm 2.45}$	$74.27_{\pm 1.55}$	$19.92_{\pm 2.19}$	$44.50_{\pm 4.46}$	$54.63_{\pm 10.59}$	$51.59_{\pm 5.62}$
	EdgePreGprompt	$64.05_{\pm 1.03}$	$50.37_{\pm 7.25}$	$62.49_{\pm 2.18}$	$55.60_{\pm 13.42}$	$23.08_{\pm 3.11}$	$61.72_{\pm 11.67}$	$74.38_{\pm 11.62}$	$56.37_{\pm 6.77}$
	GraphCL	$63.25_{\pm 2.36}$	$53.87_{\pm 3.44}$	$62.90_{\pm 2.52}$	$54.00_{\pm 12.02}$	$22.38_{\pm 1.93}$	$49.33_{\pm 11.40}$	$50.19_{\pm 4.33}$	$52.34_{\pm 6.89}$
	SimGRACE	$65.97_{\pm 0.69}$	$53.23_{\pm 4.59}$	$60.92_{\pm 1.65}$	$50.13_{\pm 13.88}$	$23.87_{\pm 3.45}$	$62.31_{\pm 8.87}$	$25.62_{\pm 8.25}$	$57.54_{\pm 4.65}$
GPF-plus	DGI	$62.45_{\pm 2.52}$	$52.14_{\pm 7.67}$	$62.16_{\pm 2.14}$	$75.20_{\pm 3.64}$	$21.92_{\pm 0.74}$	$65.25_{\pm 18.07}$	$60.86_{\pm 16.47}$	$59.43_{\pm 0.52}$
	GraphMAE	$61.97_{\pm 2.88}$	$36.87_{\pm 13.90}$	$63.55_{\pm 1.85}$	$59.33_{\pm 7.66}$	$17.08_{\pm 1.68}$	$48.20_{\pm 19.15}$	$40.99_{\pm 10.64}$	$57.18_{\pm 5.31}$
	EdgePreGPPT	$64.38_{\pm 2.30}$	$54.63_{\pm 7.14}$	$62.99_{\pm 1.94}$	$72.40_{\pm 2.00}$	$23.62_{\pm 2.56}$	$49.44_{\pm 11.92}$	$36.85_{\pm 19.33}$	$58.85_{\pm 1.33}$
	EdgePreGprompt	$64.00_{\pm 3.54}$	$50.77_{\pm 9.01}$	$60.38_{\pm 2.47}$	$50.27_{\pm 18.19}$	$24.46_{\pm 2.27}$	$52.87_{\pm 12.00}$	$50.06_{\pm 16.36}$	$58.54_{\pm 2.38}$
	GraphCL	$63.25_{\pm 2.63}$	$56.50_{\pm 3.71}$	$60.56_{\pm 1.94}$	$74.27_{\pm 4.59}$	$19.00_{\pm 2.42}$	$51.58_{\pm 11.78}$	$71.67_{\pm 14.87}$	$53.99_{\pm 6.38}$
	SimGRACE	$63.55_{\pm 2.25}$	$52.72_{\pm 6.39}$	$60.07_{\pm 0.97}$	$50.67_{\pm 17.37}$	$22.17_{\pm 2.30}$	$63.86_{\pm 10.00}$	$25.62_{\pm 8.25}$	$59.51_{\pm 0.62}$
MTG (Ours)	DGI	$63.70_{\pm 2.84}$	$52.81_{\pm 7.35}$	$62.81_{\pm 1.99}$	$71.87_{\pm 7.79}$	$28.67_{\pm 1.72}$	$46.68_{\pm 19.61}$	$74.65_{\pm 12.14}$	$55.17_{\pm 6.35}$
	GraphMAE	$63.98_{\pm 2.71}$	$45.42_{\pm 9.71}$	$63.57_{\pm 2.03}$	$64.27_{\pm 8.79}$	$24.92_{\pm 1.55}$	$51.10_{\pm 16.87}$	$55.12_{\pm 15.59}$	$58.09_{\pm 1.73}$
	EdgePreGPPT	$65.88_{\pm 3.56}$	$48.77_{\pm 7.89}$	$61.08_{\pm 2.49}$	$72.93_{\pm 1.24}$	$29.71_{\pm 2.06}$	$48.36_{\pm 6.65}$	$62.47_{\pm 14.52}$	$57.50_{\pm 4.93}$
	EdgePreGprompt	$64.05_{\pm 2.42}$	$50.86_{\pm 8.15}$	$70.49_{\pm 0.68}$	$78.13_{\pm 6.36}$	$24.71_{\pm 1.24}$	$63.27_{\pm 8.85}$	$49.57_{\pm 10.27}$	$60.85_{\pm 6.39}$
	GraphCL	$64.05_{\pm 2.63}$	$57.49_{\pm 2.52}$	$63.21_{\pm 2.66}$	$64.13_{\pm 8.00}$	$21.00_{\pm 1.75}$	$73.86_{\pm 9.74}$	$50.99_{\pm 15.50}$	$55.74_{\pm 6.19}$
	SimGRACE	$66.95_{\pm 0.59}$	$53.47_{\pm 7.07}$	$65.25_{\pm 3.49}$	$58.87_{\pm 7.66}$	$20.71_{\pm 0.90}$	$66.85_{\pm 15.53}$	$60.31_{\pm 17.89}$	$58.09_{\pm 5.01}$

Table 27: F1-score on 3-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	$39.88_{\pm 13.07}$	$40.73_{\pm 1.00}$	$51.01_{\pm 8.18}$	$57.79_{\pm 8.44}$	$11.97_{\pm 0.92}$	$42.57_{\pm 5.89}$	$47.71_{\pm 5.27}$	$41.05_{\pm 7.70}$
Fine-tuning	DGI	$39.88_{\pm 13.07}$	$56.10_{\pm 3.90}$	$50.89_{\pm 8.01}$	$58.19_{\pm 8.98}$	$14.68_{\pm 2.46}$	$47.30_{\pm 8.88}$	$46.39_{\pm 5.60}$	$41.31_{\pm 8.21}$
	GraphMAE	$39.88_{\pm 13.08}$	$47.71_{\pm 19.43}$	$51.90_{\pm 7.98}$	$37.98_{\pm 16.20}$	$10.19_{\pm 2.54}$	$43.74_{\pm 6.72}$	$33.85_{\pm 14.41}$	$40.24_{\pm 6.08}$
	EdgePreGPPT	$60.66_{\pm 4.52}$	$35.28_{\pm 13.12}$	$54.53_{\pm 4.29}$	$37.20_{\pm 15.67}$	$15.47_{\pm 3.45}$	$45.43_{\pm 4.85}$	$39.01_{\pm 11.37}$	$40.78_{\pm 7.17}$
	EdgePreGprompt	$39.88_{\pm 13.07}$	$35.69_{\pm 7.57}$	$49.31_{\pm 7.52}$	$57.28_{\pm 6.35}$	$12.99_{\pm 2.98}$	$33.02_{\pm 14.02}$	$26.66_{\pm 11.36}$	$48.50_{\pm 10.32}$
	GraphCL	$60.72_{\pm 1.89}$	$54.45_{\pm 4.00}$	$52.75_{\pm 6.77}$	$53.13_{\pm 5.12}$	$17.00_{\pm 3.88}$	$30.17_{\pm 7.80}$	$45.33_{\pm 7.28}$	$51.94_{\pm 4.43}$
	SimGRACE	$65.91_{\pm 0.61}$	$54.77_{\pm 3.76}$	$48.85_{\pm 6.10}$	$52.10_{\pm 6.02}$	$16.13_{\pm 2.28}$	$45.75_{\pm 2.10}$	$36.07_{\pm 1.29}$	$50.54_{\pm 6.18}$
GPPTPrompt	DGI	$33.47_{\pm 0.41}$	$25.72_{\pm 11.59}$	$56.81_{\pm 8.37}$	$56.46_{\pm 18.74}$	$7.26_{\pm 4.98}$	$33.97_{\pm 12.73}$	$53.00_{\pm 6.62}$	$52.76_{\pm 6.80}$
	GraphMAE	$33.43_{\pm 0.42}$	$17.63_{\pm 2.61}$	$44.10_{\pm 8.25}$	$41.20_{\pm 19.70}$	$7.70_{\pm 5.86}$	$41.47_{\pm 5.35}$	$54.54_{\pm 3.82}$	$53.64_{\pm 7.98}$
	EdgePreGPPT	$54.15_{\pm 12.01}$	$28.60_{\pm 16.18}$	$61.46_{\pm 1.10}$	$49.73_{\pm 17.65}$	$12.26_{\pm 6.37}$	$41.87_{\pm 4.54}$	$54.17_{\pm 4.30}$	$50.03_{\pm 5.25}$
	EdgePreGprompt	$36.01_{\pm 4.61}$	$17.54_{\pm 2.57}$	$51.33_{\pm 7.17}$	$61.15_{\pm 18.28}$	$11.46_{\pm 8.27}$	$41.89_{\pm 5.55}$	$54.83_{\pm 5.76}$	$53.54_{\pm 7.77}$
	GraphCL	$46.73_{\pm 14.15}$	$49.62_{\pm 6.77}$	$41.87_{\pm 8.55}$	$41.46_{\pm 20.04}$	$9.53_{\pm 5.93}$	$43.94_{\pm 0.52}$	$48.82_{\pm 8.73}$	$53.74_{\pm 6.42}$
	SimGRACE	$47.67_{\pm 15.00}$	$38.51_{\pm 10.60}$	$54.02_{\pm 4.98}$	$55.79_{\pm 19.08}$	$6.44_{\pm 2.25}$	$42.49_{\pm 8.52}$	$53.66_{\pm 6.91}$	$56.28_{\pm 5.60}$
Gprompt	DGI	$58.08_{\pm 10.23}$	$55.32_{\pm 8.74}$	$58.05_{\pm 3.97}$	$52.54_{\pm 17.09}$	$18.81_{\pm 1.44}$	$44.13_{\pm 4.49}$	$45.80_{\pm 7.78}$	$50.97_{\pm 5.00}$
	GraphMAE	$58.01_{\pm 10.19}$	$17.54_{\pm 2.57}$	$56.27_{\pm 5.10}$	$60.26_{\pm 15.18}$	$19.39_{\pm 1.66}$	$42.42_{\pm 7.00}$	$43.14_{\pm 4.47}$	$49.67_{\pm 2.97}$
	EdgePreGPPT	$63.89_{\pm 1.11}$	$53.51_{\pm 7.87}$	$61.58_{\pm 3.09}$	$49.72_{\pm 12.38}$	$16.84_{\pm 3.73}$	$42.05_{\pm 5.59}$	$47.32_{\pm 2.37}$	$52.83_{\pm 1.88}$
	EdgePreGprompt	$58.51_{\pm 10.62}$	$55.39_{\pm 8.86}$	$56.27_{\pm 2.08}$	$64.64_{\pm 13.86}$	$18.56_{\pm 3.11}$	$43.87_{\pm 4.26}$	$44.57_{\pm 9.37}$	$49.71_{\pm 7.25}$
	GraphCL	$58.83_{\pm 10.22}$	$52.04_{\pm 10.22}$	$57.08_{\pm 8.81}$	$51.26_{\pm 19.96}$	$19.60_{\pm 1.34}$	$44.29_{\pm 5.52}$	$47.69_{\pm 4.92}$	$54.82_{\pm 2.66}$
	SimGRACE	$59.21_{\pm 10.45}$	$52.82_{\pm 8.34}$	$55.53_{\pm 7.37}$	$52.05_{\pm 11.07}$	$21.63_{\pm 3.19}$	$43.90_{\pm 7.83}$	$40.90_{\pm 6.25}$	$53.39_{\pm 6.72}$
All-in-one	DGI	$63.99_{\pm 0.86}$	$50.43_{\pm 9.08}$	$68.85_{\pm 5.72}$	$73.80_{\pm 4.55}$	$14.63_{\pm 1.31}$	$43.21_{\pm 7.33}$	$53.56_{\pm 10.74}$	$43.35_{\pm 5.95}$
	GraphMAE	$62.94_{\pm 1.54}$	$22.83_{\pm 0.10}$	$62.48_{\pm 1.62}$	$69.80_{\pm 7.77}$	$15.61_{\pm 1.28}$	$46.34_{\pm 2.82}$	$55.20_{\pm 8.12}$	$54.23_{\pm 4.13}$
	EdgePreGPPT	$62.99_{\pm 2.30}$	$55.56_{\pm 4.18}$	$64.50_{\pm 6.98}$	$76.24_{\pm 4.31}$	$17.11_{\pm 3.24}$	$42.10_{\pm 6.16}$	$44.01_{\pm 4.11}$	$49.69_{\pm 6.19}$
	EdgePreGprompt	$62.88_{\pm 3.31}$	$49.83_{\pm 10.02}$	$52.05_{\pm 8.59}$	$71.90_{\pm 2.68}$	$17.70_{\pm 1.81}$	$50.66_{\pm 4.25}$	$47.08_{\pm 8.47}$	$55.21_{\pm 4.94}$
	GraphCL	$65.60_{\pm 0.57}$	$56.55_{\pm 1.88}$	$62.62_{\pm 3.77}$	$56.86_{\pm 15.65}$	$22.37_{\pm 1.31}$	$45.27_{\pm 9.63}$	$49.39_{\pm 2.68}$	$50.28_{\pm 4.51}$
	SimGRACE	$63.50_{\pm 2.41}$	$55.34_{\pm 3.89}$	$60.36_{\pm 1.11}$	$52.08_{\pm 14.23}$	$18.20_{\pm 1.04}$	$41.93_{\pm 5.18}$	$49.00_{\pm 9.10}$	$49.06_{\pm 6.95}$
GPF	DGI	$62.80_{\pm 2.95}$	$47.97_{\pm 9.36}$	$51.68_{\pm 7.67}$	$44.82_{\pm 12.44}$	$11.70_{\pm 4.44}$	$41.60_{\pm 6.58}$	$40.70_{\pm 4.95}$	$40.29_{\pm 5.61}$
	GraphMAE	$61.83_{\pm 3.65}$	$17.49_{\pm 5.13}$	$55.21_{\pm 7.59}$	$52.99_{\pm 14.80}$	$15.77_{\pm 1.71}$	$45.77_{\pm 7.77}$	$47.24_{\pm 5.79}$	$54.79_{\pm 4.60}$
	EdgePreGPPT	$63.82_{\pm 3.68}$	$50.28_{\pm 10.09}$	$56.68_{\pm 3.32}$	$68.47_{\pm 1.51}$	$18.77_{\pm 2.43}$	$42.92_{\pm 3.39}$	$48.61_{\pm 4.66}$	$49.22_{\pm 6.62}$
	EdgePreGprompt	$63.59_{\pm 0.99}$	$48.60_{\pm 9.17}$	$53.75_{\pm 5.47}$	$51.63_{\pm 14.57}$	$16.62_{\pm 1.75}$	$53.47_{\pm 7.31}$	$48.42_{\pm 3.50}$	$53.86_{\pm 7.88}$
	GraphCL	$62.47_{\pm 2.84}$	$53.42_{\pm 3.86}$	$55.24_{\pm 8.26}$	$50.93_{\pm 14.25}$	$15.40_{\pm 2.55}$	$45.37_{\pm 9.89}$	$47.00_{\pm 2.24}$	$48.53_{\pm 5.59}$
	SimGRACE	$65.85_{\pm 0.61}$	$52.54_{\pm 5.59}$	$49.67_{\pm 6.21}$	$47.58_{\pm 13.28}$	$18.36_{\pm 4.95}$	$51.59_{\pm 4.59}$	$22.84_{\pm 9.35}$	$55.12_{\pm 4.27}$
GPF-plus	DGI	$60.90_{\pm 3.22}$	$50.49_{\pm 9.05}$	$53.28_{\pm 5.40}$	$73.17_{\pm 3.16}$	$17.51_{\pm 1.75}$	$43.17_{\pm 5.33}$	$46.21_{\pm 5.60}$	$40.98_{\pm 7.58}$
	GraphMAE	$60.30_{\pm 3.23}$	$17.43_{\pm 5.19}$	$59.82_{\pm 4.60}$	$57.10_{\pm 7.36}$	$12.55_{\pm 0.72}$	$37.20_{\pm 9.82}$	$38.17_{\pm 8.53}$	$52.72_{\pm 3.75}$
	EdgePreGPPT	$63.35_{\pm 3.33}$	$53.82_{\pm 8.44}$	$59.39_{\pm 1.65}$	$65.96_{\pm 1.23}$	$18.73_{\pm 2.85}$	$46.05_{\pm 10.35}$	$31.44_{\pm 17.20}$	$41.26_{\pm 6.70}$
	EdgePreGprompt	$63.65_{\pm 3.51}$	$48.46_{\pm 10.27}$	$49.63_{\pm 6.76}$	$48.24_{\pm 17.45}$	$19.80_{\pm 4.12}$	$47.82_{\pm 9.03}$	$43.31_{\pm 12.09}$	$44.08_{\pm 7.81}$
	GraphCL	$62.49_{\pm 3.11}$	$56.38_{\pm 3.66}$	$49.25_{\pm 7.67}$	$72.21_{\pm 3.78}$	$15.66_{\pm 2.56}$	$47.78_{\pm 9.88}$	$43.52_{\pm 1.30}$	$43.72_{\pm 6.72}$
	SimGRACE	$62.91_{\pm 2.78}$	$51.40_{\pm 8.33}$	$48.38_{\pm 6.37}$	$48.42_{\pm 16.33}$	$16.25_{\pm 3.30}$	$50.53_{\pm 3.93}$	$22.82_{\pm 9.32}$	$40.63_{\pm 6.86}$
MTG (Ours)	DGI	$62.69_{\pm 3.56}$	$51.19_{\pm 9.38}$	$55.88_{\pm 8.05}$	$60.64_{\pm 7.17}$	$22.08_{\pm 2.56}$	$39.27_{\pm 14.67}$	$53.96_{\pm 10.01}$	$47.74_{\pm 9.47}$
	GraphMAE	$63.10_{\pm 3.36}$	$23.97_{\pm 3.21}$	$56.93_{\pm 4.18}$	$60.75_{\pm 4.47}$	$17.00_{\pm 2.58}$	$43.25_{\pm 13.60}$	$47.56_{\pm 11.55}$	$52.06_{\pm 7.45}$
	EdgePreGPPT	$64.92_{\pm 5.54}$	$47.41_{\pm 9.79}$	$53.44_{\pm 6.46}$	$69.05_{\pm 2.04}$	$27.09_{\pm 2.83}$	$44.97_{\pm 4.60}$	$48.58_{\pm 11.51}$	$52.97_{\pm 4.61}$
	EdgePreGprompt	$63.14_{\pm 2.98}$	$48.89_{\pm 9.87}$	$68.63_{\pm 1.12}$	$76.72_{\pm 2.47}$	$21.12_{\pm 1.42}$	$50.03_{\pm 1.19}$	$44.13_{\pm 4.70}$	$56.04_{\pm 5.72}$
	GraphCL	$63.16_{\pm 3.26}$	$57.43_{\pm 2.82}$	$59.12_{\pm 4.32}$	$62.01_{\pm 9.97}$	$17.15_{\pm 2.94}$	$53.94_{\pm 2.89}$	$45.21_{\pm 9.86}$	$49.25_{\pm 5.27}$
	SimGRACE	$66.84_{\pm 0.53}$	$52.58_{\pm 8.90}$	$64.00_{\pm 5.68}$	$60.03_{\pm 7.32}$	$15.17_{\pm 3.35}$	$43.51_{\pm 13.49}$	$45.33_{\pm 5.84}$	$56.02_{\pm 4.07}$

Table 28: AUROC on 3-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	$53.85_{\pm 7.70}$	$54.29_{\pm 0.68}$	$58.83_{\pm 2.09}$	$73.36_{\pm 0.52}$	$51.34_{\pm 1.42}$	$53.03_{\pm 0.66}$	$55.36_{\pm 2.82}$	$50.97_{\pm 1.94}$
Fine-tuning	DGI	$53.85_{\pm 7.71}$	$79.16_{\pm 1.29}$	$59.21_{\pm 1.87}$	$72.49_{\pm 0.69}$	$55.06_{\pm 0.72}$	$59.34_{\pm 3.68}$	$55.52_{\pm 3.89}$	$50.91_{\pm 1.85}$
	GraphMAE	$53.86_{\pm 7.71}$	$72.58_{\pm 11.44}$	$59.12_{\pm 1.77}$	$73.70_{\pm 2.39}$	$51.64_{\pm 1.15}$	$56.52_{\pm 2.87}$	$57.30_{\pm 2.71}$	$50.87_{\pm 1.75}$
	EdgePreGPPT	$68.90_{\pm 1.00}$	$53.70_{\pm 2.76}$	$59.61_{\pm 1.05}$	$63.78_{\pm 11.40}$	$56.63_{\pm 1.88}$	$58.62_{\pm 4.95}$	$53.24_{\pm 1.06}$	$50.87_{\pm 1.74}$
	EdgePreGprompt	$53.88_{\pm 7.76}$	$54.66_{\pm 1.29}$	$57.43_{\pm 1.26}$	$72.67_{\pm 0.71}$	$52.81_{\pm 1.23}$	$56.43_{\pm 2.54}$	$52.85_{\pm 3.05}$	$55.71_{\pm 3.36}$
	GraphCL	$66.91_{\pm 0.96}$	$77.24_{\pm 5.27}$	$59.63_{\pm 1.25}$	$71.94_{\pm 1.69}$	$55.22_{\pm 0.62}$	$55.65_{\pm 1.39}$	$52.36_{\pm 4.70}$	$56.27_{\pm 3.63}$
	SimGRACE	$69.07_{\pm 0.74}$	$78.72_{\pm 0.81}$	$54.56_{\pm 4.50}$	$71.53_{\pm 1.53}$	$57.45_{\pm 1.25}$	$54.90_{\pm 1.47}$	$53.79_{\pm 5.86}$	$58.14_{\pm 3.45}$
GPPTPrompt	DGI	$50.70_{\pm 0.98}$	$59.06_{\pm 5.42}$	$63.21_{\pm 13.02}$	$69.64_{\pm 24.34}$	$51.97_{\pm 1.01}$	$51.15_{\pm 3.38}$	$53.46_{\pm 4.77}$	$58.16_{\pm 7.86}$
	GraphMAE	$50.02_{\pm 1.20}$	$55.37_{\pm 1.65}$	$64.79_{\pm 12.01}$	$36.64_{\pm 30.88}$	$51.88_{\pm 0.85}$	$51.90_{\pm 3.51}$	$57.33_{\pm 2.22}$	$57.48_{\pm 8.59}$
	EdgePreGPPT	$60.13_{\pm 5.46}$	$59.32_{\pm 8.93}$	$68.26_{\pm 2.20}$	$53.05_{\pm 23.65}$	$52.97_{\pm 1.80}$	$54.30_{\pm 1.55}$	$53.75_{\pm 1.34}$	$54.71_{\pm 6.13}$
	EdgePreGprompt	$51.09_{\pm 3.03}$	$55.18_{\pm 1.81}$	$68.92_{\pm 2.46}$	$66.15_{\pm 23.60}$	$51.81_{\pm 3.18}$	$50.12_{\pm 2.64}$	$56.67_{\pm 1.07}$	$59.80_{\pm 7.91}$
	GraphCL	$50.36_{\pm 13.14}$	$70.37_{\pm 8.98}$	$71.44_{\pm 3.56}$	$41.39_{\pm 32.92}$	$51.66_{\pm 0.70}$	$54.79_{\pm 2.46}$	$55.63_{\pm 1.51}$	$58.02_{\pm 8.00}$
	SimGRACE	$50.30_{\pm 13.72}$	$63.01_{\pm 7.00}$	$57.69_{\pm 9.30}$	$59.52_{\pm 26.16}$	$49.66_{\pm 2.26}$	$53.90_{\pm 4.39}$	$53.94_{\pm 3.10}$	$60.31_{\pm 6.86}$
Gprompt	DGI	$61.28_{\pm 6.92}$	$72.35_{\pm 6.43}$	$60.00_{\pm 4.15}$	$59.82_{\pm 8.05}$	$53.15_{\pm 3.14}$	$49.01_{\pm 1.79}$	$50.69_{\pm 8.24}$	$52.10_{\pm 4.06}$
	GraphMAE	$56.22_{\pm 7.41}$	$44.90_{\pm 0.12}$	$53.63_{\pm 3.55}$	$64.43_{\pm 15.28}$	$52.29_{\pm 3.02}$	$55.66_{\pm 5.33}$	$50.42_{\pm 1.52}$	$55.22_{\pm 1.63}$
	EdgePreGPPT	$63.01_{\pm 2.64}$	$74.04_{\pm 6.87}$	$65.82_{\pm 4.00}$	$56.24_{\pm 15.55}$	$51.97_{\pm 3.30}$	$49.66_{\pm 7.34}$	$50.37_{\pm 2.68}$	$57.12_{\pm 1.01}$
	EdgePreGprompt	$61.76_{\pm 6.92}$	$74.20_{\pm 7.18}$	$63.20_{\pm 4.24}$	$68.36_{\pm 15.60}$	$54.30_{\pm 3.58}$	$51.46_{\pm 3.88}$	$52.11_{\pm 4.43}$	$55.65_{\pm 4.85}$
	GraphCL	$62.10_{\pm 7.10}$	$72.58_{\pm 8.88}$	$55.52_{\pm 9.16}$	$54.21_{\pm 19.23}$	$52.15_{\pm 2.03}$	$57.07_{\pm 3.48}$	$52.27_{\pm 2.39}$	$57.24_{\pm 4.44}$
	SimGRACE	$63.49_{\pm 7.53}$	$73.85_{\pm 5.74}$	$58.07_{\pm 7.27}$	$56.06_{\pm 18.58}$	$53.72_{\pm 4.10}$	$53.97_{\pm 4.68}$	$49.80_{\pm 7.28}$	$56.78_{\pm 6.19}$
All-in-one	DGI	$68.29_{\pm 0.65}$	$81.02_{\pm 0.32}$	$78.78_{\pm 0.89}$	$79.29_{\pm 6.57}$	$55.50_{\pm 0.41}$	$50.11_{\pm 4.43}$	$64.10_{\pm 1.82}$	$47.72_{\pm 1.00}$
	GraphMAE	$67.69_{\pm 0.61}$	$50.00_{\pm 0.00}$	$71.38_{\pm 4.27}$	$78.41_{\pm 1.12}$	$54.06_{\pm 0.38}$	$51.31_{\pm 6.59}$	$62.34_{\pm 1.48}$	$59.05_{\pm 3.75}$
	EdgePreGPPT	$67.39_{\pm 0.61}$	$80.54_{\pm 1.13}$	$71.08_{\pm 8.07}$	$81.40_{\pm 0.45}$	$55.76_{\pm 1.18}$	$47.76_{\pm 1.24}$	$48.35_{\pm 3.93}$	$54.96_{\pm 0.53}$
	EdgePreGprompt	$68.01_{\pm 0.66}$	$77.37_{\pm 3.38}$	$59.05_{\pm 4.06}$	$80.89_{\pm 2.30}$	$57.14_{\pm 0.79}$	$56.46_{\pm 1.24}$	$64.23_{\pm 1.73}$	$60.92_{\pm 1.89}$
	GraphCL	$68.55_{\pm 0.68}$	$81.56_{\pm 1.31}$	$68.66_{\pm 3.13}$	$75.02_{\pm 2.18}$	$54.68_{\pm 0.82}$	$58.33_{\pm 4.46}$	$54.20_{\pm 4.83}$	$55.30_{\pm 3.48}$
	SimGRACE	$67.71_{\pm 0.63}$	$79.47_{\pm 0.36}$	$62.54_{\pm 0.80}$	$73.98_{\pm 2.36}$	$54.94_{\pm 0.85}$	$47.16_{\pm 4.60}$	$54.11_{\pm 1.26}$	$58.19_{\pm 0.80}$
GPF	DGI	$68.17_{\pm 1.10}$	$76.98_{\pm 3.82}$	$60.16_{\pm 1.20}$	$49.38_{\pm 22.29}$	$47.07_{\pm 1.84}$	$48.08_{\pm 2.39}$	$57.22_{\pm 3.09}$	$48.85_{\pm 1.22}$
	GraphMAE	$68.64_{\pm 0.80}$	$50.00_{\pm 0.00}$	$60.27_{\pm 1.58}$	$75.23_{\pm 2.86}$	$52.79_{\pm 1.96}$	$52.96_{\pm 5.14}$	$59.08_{\pm 2.59}$	$57.07_{\pm 5.22}$
	EdgePreGPPT	$72.39_{\pm 1.43}$	$78.68_{\pm 3.14}$	$59.70_{\pm 1.16}$	$71.74_{\pm 1.36}$	$53.23_{\pm 0.59}$	$55.95_{\pm 3.53}$	$57.16_{\pm 1.15}$	$55.12_{\pm 1.37}$
	EdgePreGprompt	$64.81_{\pm 4.48}$	$76.99_{\pm 3.95}$	$59.65_{\pm 1.05}$	$75.64_{\pm 1.21}$	$55.95_{\pm 1.04}$	$61.22_{\pm 3.30}$	$57.14_{\pm 3.83}$	$60.99_{\pm 3.42}$
	GraphCL	$68.50_{\pm 0.89}$	$79.36_{\pm 0.64}$	$58.20_{\pm 2.58}$	$75.42_{\pm 2.40}$	$55.17_{\pm 1.18}$	$57.66_{\pm 3.49}$	$56.26_{\pm 0.72}$	$54.38_{\pm 2.29}$
	SimGRACE	$68.88_{\pm 0.73}$	$79.18_{\pm 0.70}$	$59.03_{\pm 1.13}$	$54.34_{\pm 11.05}$	$56.49_{\pm 0.82}$	$59.26_{\pm 4.32}$	$53.33_{\pm 0.93}$	$58.78_{\pm 0.91}$
GPF-plus	DGI	$67.30_{\pm 2.04}$	$78.76_{\pm 3.06}$	$59.98_{\pm 1.03}$	$78.67_{\pm 3.24}$	$55.76_{\pm 0.99}$	$58.01_{\pm 1.54}$	$55.60_{\pm 4.64}$	$53.17_{\pm 0.34}$
	GraphMAE	$66.86_{\pm 1.34}$	$50.00_{\pm 0.00}$	$61.20_{\pm 1.57}$	$72.94_{\pm 0.86}$	$49.96_{\pm 0.80}$	$47.18_{\pm 2.64}$	$54.12_{\pm 12.01}$	$59.32_{\pm 3.11}$
	EdgePreGPPT	$68.32_{\pm 1.40}$	$79.72_{\pm 3.57}$	$61.31_{\pm 2.41}$	$68.83_{\pm 3.06}$	$55.99_{\pm 0.45}$	$61.66_{\pm 1.49}$	$51.94_{\pm 1.06}$	$49.50_{\pm 2.07}$
	EdgePreGprompt	$66.08_{\pm 4.10}$	$78.31_{\pm 2.86}$	$58.73_{\pm 2.17}$	$51.73_{\pm 17.01}$	$57.28_{\pm 2.04}$	$59.88_{\pm 2.24}$	$52.69_{\pm 2.74}$	$55.53_{\pm 2.48}$
	GraphCL	$67.76_{\pm 1.67}$	$78.53_{\pm 0.70}$	$57.90_{\pm 2.40}$	$77.86_{\pm 2.73}$	$53.24_{\pm 1.85}$	$61.30_{\pm 2.87}$	$50.97_{\pm 1.94}$	$52.80_{\pm 2.36}$
	SimGRACE	$67.38_{\pm 2.08}$	$79.41_{\pm 0.77}$	$53.05_{\pm 5.96}$	$49.04_{\pm 15.66}$	$55.30_{\pm 1.21}$	$59.92_{\pm 3.03}$	$52.68_{\pm 1.00}$	$51.35_{\pm 2.70}$
MTG (Ours)	DGI	$68.95_{\pm 0.76}$	$80.49_{\pm 3.04}$	$62.32_{\pm 3.10}$	$72.02_{\pm 8.27}$	$59.58_{\pm 1.34}$	$46.07_{\pm 1.73}$	$68.81_{\pm 2.08}$	$54.48_{\pm 4.11}$
	GraphMAE	$68.55_{\pm 0.82}$	$50.00_{\pm 0.00}$	$61.39_{\pm 1.59}$	$75.05_{\pm 2.89}$	$54.61_{\pm 0.71}$	$53.34_{\pm 4.63}$	$58.05_{\pm 5.94}$	$57.26_{\pm 4.95}$
	EdgePreGPPT	$72.87_{\pm 0.86}$	$77.24_{\pm 2.52}$	$57.91_{\pm 6.55}$	$72.79_{\pm 3.33}$	$63.07_{\pm 2.61}$	$52.85_{\pm 3.60}$	$66.10_{\pm 4.03}$	$56.70_{\pm 2.20}$
	EdgePreGprompt	$68.90_{\pm 0.80}$	$78.38_{\pm 2.51}$	$69.61_{\pm 0.42}$	$81.07_{\pm 3.97}$	$57.51_{\pm 1.67}$	$52.76_{\pm 2.84}$	$51.80_{\pm 4.14}$	$62.71_{\pm 4.47}$
	GraphCL	$68.58_{\pm 0.69}$	$81.55_{\pm 0.42}$	$59.80_{\pm 1.08}$	$76.23_{\pm 2.36}$	$53.27_{\pm 0.85}$	$58.85_{\pm 6.87}$	$57.23_{\pm 7.72}$	$54.20_{\pm 1.33}$
	SimGRACE	$69.00_{\pm 0.85}$	$78.44_{\pm 3.57}$	$65.16_{\pm 2.85}$	$67.17_{\pm 8.36}$	$51.30_{\pm 1.01}$	$54.05_{\pm 4.52}$	$56.49_{\pm 7.52}$	$59.78_{\pm 0.85}$

Table 29: Accuracy (%) on 5-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	$62.60_{\pm 4.01}$	$55.23_{\pm 4.26}$	$62.90_{\pm 5.03}$	$73.47_{\pm 3.92}$	$25.67_{\pm 0.48}$	$64.99_{\pm 10.42}$	$51.48_{\pm 2.29}$	$63.59_{\pm 2.86}$
Fine-tuning	DGI	$50.80_{\pm 0.88}$	$60.72_{\pm 2.09}$	$62.47_{\pm 4.25}$	$74.53_{\pm 3.71}$	$25.17_{\pm 1.67}$	$68.36_{\pm 11.83}$	$24.94_{\pm 7.80}$	$58.94_{\pm 9.66}$
	GraphMAE	$55.75_{\pm 5.48}$	$54.40_{\pm 5.15}$	$61.21_{\pm 6.92}$	$67.33_{\pm 5.61}$	$26.13_{\pm 1.09}$	$68.42_{\pm 11.71}$	$38.27_{\pm 18.11}$	$64.71_{\pm 3.22}$
	EdgePreGPPT	$60.97_{\pm 10.07}$	$57.44_{\pm 3.75}$	$62.85_{\pm 4.63}$	$70.80_{\pm 3.05}$	$27.04_{\pm 1.75}$	$69.38_{\pm 10.91}$	$58.40_{\pm 12.79}$	$64.52_{\pm 2.97}$
	EdgePreGprompt	$62.88_{\pm 4.02}$	$58.11_{\pm 3.76}$	$60.25_{\pm 4.96}$	$73.20_{\pm 2.44}$	$27.46_{\pm 1.29}$	$73.19_{\pm 9.53}$	$34.20_{\pm 16.57}$	$62.70_{\pm 3.37}$
	GraphCL	$65.40_{\pm 3.33}$	$48.28_{\pm 6.30}$	$63.33_{\pm 4.13}$	$75.33_{\pm 1.89}$	$24.63_{\pm 1.52}$	$55.44_{\pm 3.84}$	$72.96_{\pm 11.98}$	$63.84_{\pm 2.06}$
	SimGRACE	$60.65_{\pm 4.54}$	$48.40_{\pm 5.90}$	$63.08_{\pm 4.66}$	$71.73_{\pm 2.33}$	$24.79_{\pm 1.83}$	$61.50_{\pm 16.06}$	$70.80_{\pm 16.29}$	$62.25_{\pm 3.23}$
GPPTPrompt	DGI	$49.90_{\pm 0.84}$	$48.54_{\pm 9.22}$	$54.49_{\pm 9.48}$	$70.53_{\pm 3.90}$	$19.54_{\pm 3.17}$	$47.08_{\pm 25.71}$	$54.75_{\pm 19.07}$	$52.65_{\pm 9.39}$
	GraphMAE	$49.60_{\pm 0.75}$	$33.61_{\pm 16.35}$	$58.27_{\pm 4.63}$	$69.20_{\pm 3.30}$	$22.17_{\pm 2.34}$	$56.14_{\pm 25.93}$	$69.63_{\pm 14.96}$	$57.07_{\pm 5.16}$
	EdgePreGPPT	$66.37_{\pm 3.59}$	$50.16_{\pm 9.23}$	$57.37_{\pm 4.85}$	$65.33_{\pm 3.65}$	$21.33_{\pm 1.04}$	$60.05_{\pm 13.16}$	$45.37_{\pm 19.12}$	$51.36_{\pm 4.26}$
	EdgePreGprompt	$49.60_{\pm 0.75}$	$37.57_{\pm 10.07}$	$46.92_{\pm 11.86}$	$40.67_{\pm 14.47}$	$21.96_{\pm 2.12}$	$67.88_{\pm 17.34}$	$48.02_{\pm 12.75}$	$60.02_{\pm 3.24}$
	GraphCL	$59.58_{\pm 9.19}$	$55.21_{\pm 1.21}$	$58.25_{\pm 3.13}$	$70.40_{\pm 4.10}$	$20.92_{\pm 2.40}$	$64.50_{\pm 19.35}$	$56.11_{\pm 21.68}$	$57.05_{\pm 3.15}$
	SimGRACE	$60.35_{\pm 8.28}$	$54.05_{\pm 4.58}$	$56.85_{\pm 10.85}$	$68.67_{\pm 5.98}$	$20.96_{\pm 1.23}$	$60.80_{\pm 13.90}$	$60.80_{\pm 8.71}$	$56.71_{\pm 4.10}$
Gprompt	DGI	$53.05_{\pm 10.49}$	$60.62_{\pm 2.31}$	$61.30_{\pm 3.46}$	$62.93_{\pm 15.66}$	$21.38_{\pm 1.97}$	$46.54_{\pm 7.06}$	$59.38_{\pm 14.43}$	$54.12_{\pm 5.51}$
	GraphMAE	$55.15_{\pm 7.10}$	$32.55_{\pm 0.12}$	$58.00_{\pm 7.16}$	$65.87_{\pm 8.50}$	$19.87_{\pm 2.24}$	$52.23_{\pm 5.40}$	$53.77_{\pm 6.74}$	$50.68_{\pm 5.33}$
	EdgePreGPPT	$66.70_{\pm 3.87}$	$57.23_{\pm 2.43}$	$62.94_{\pm 1.38}$	$63.87_{\pm 12.45}$	$17.79_{\pm 1.27}$	$53.35_{\pm 4.42}$	$51.98_{\pm 8.17}$	$52.46_{\pm 4.48}$
	EdgePreGprompt	$52.75_{\pm 10.26}$	$57.71_{\pm 1.56}$	$55.55_{\pm 6.18}$	$73.07_{\pm 2.13}$	$20.79_{\pm 2.66}$	$52.12_{\pm 7.21}$	$54.32_{\pm 11.18}$	$56.24_{\pm 3.69}$
	GraphCL	$61.38_{\pm 10.05}$	$60.71_{\pm 4.54}$	$60.54_{\pm 2.22}$	$64.40_{\pm 8.95}$	$20.33_{\pm 2.27}$	$53.24_{\pm 10.24}$	$56.73_{\pm 7.80}$	$58.28_{\pm 2.18}$
	SimGRACE	$54.85_{\pm 10.71}$	$60.76_{\pm 5.08}$	$61.62_{\pm 2.86}$	$63.47_{\pm 7.90}$	$21.46_{\pm 2.27}$	$53.35_{\pm 7.75}$	$51.36_{\pm 9.14}$	$54.33_{\pm 4.04}$
All-in-one	DGI	$60.47_{\pm 7.38}$	$57.59_{\pm 5.73}$	$71.37_{\pm 4.89}$	$78.53_{\pm 1.36}$	$26.46_{\pm 2.24}$	$48.15_{\pm 4.32}$	$61.73_{\pm 8.94}$	$54.59_{\pm 2.82}$
	GraphMAE	$59.37_{\pm 9.45}$	$36.86_{\pm 13.96}$	$68.72_{\pm 3.98}$	$77.40_{\pm 2.88}$	$24.00_{\pm 0.87}$	$48.79_{\pm 7.42}$	$57.10_{\pm 17.82}$	$63.44_{\pm 1.35}$
	EdgePreGPPT	$63.62_{\pm 2.30}$	$57.86_{\pm 5.88}$	$70.56_{\pm 2.54}$	$80.93_{\pm 1.96}$	$23.92_{\pm 2.14}$	$52.17_{\pm 5.59}$	$62.78_{\pm 10.18}$	$53.93_{\pm 4.10}$
	EdgePreGprompt	$60.18_{\pm 7.66}$	$57.63_{\pm 7.77}$	$63.33_{\pm 2.98}$	$72.67_{\pm 3.45}$	$25.00_{\pm 1.56}$	$51.05_{\pm 7.26}$	$49.69_{\pm 13.79}$	$62.78_{\pm 0.65}$
	GraphCL	$61.80_{\pm 4.92}$	$57.04_{\pm 4.46}$	$69.69_{\pm 6.19}$	$69.33_{\pm 4.26}$	$26.71_{\pm 2.17}$	$56.41_{\pm 6.50}$	$51.67_{\pm 9.26}$	$60.62_{\pm 3.84}$
	SimGRACE	$63.05_{\pm 3.01}$	$54.01_{\pm 0.83}$	$68.72_{\pm 4.97}$	$74.27_{\pm 2.62}$	$26.67_{\pm 0.99}$	$62.95_{\pm 8.57}$	$53.64_{\pm 5.97}$	$59.32_{\pm 0.34}$
GPF	DGI	$59.90_{\pm 9.95}$	$59.65_{\pm 6.25}$	$60.99_{\pm 4.00}$	$72.93_{\pm 3.14}$	$24.75_{\pm 0.98}$	$52.98_{\pm 15.39}$	$61.05_{\pm 11.51}$	$58.70_{\pm 1.84}$
	GraphMAE	$66.02_{\pm 3.96}$	$29.68_{\pm 13.57}$	$63.37_{\pm 4.37}$	$70.13_{\pm 2.58}$	$23.92_{\pm 2.16}$	$55.87_{\pm 15.72}$	$54.63_{\pm 9.43}$	$60.57_{\pm 3.74}$
	EdgePreGPPT	$67.80_{\pm 5.58}$	$58.07_{\pm 6.03}$	$63.28_{\pm 4.33}$	$68.40_{\pm 3.44}$	$24.13_{\pm 1.56}$	$66.27_{\pm 14.57}$	$32.53_{\pm 10.35}$	$55.73_{\pm 3.64}$
	EdgePreGprompt	$62.62_{\pm 4.73}$	$56.62_{\pm 5.44}$	$62.34_{\pm 3.36}$	$74.00_{\pm 3.65}$	$24.29_{\pm 1.46}$	$64.77_{\pm 7.22}$	$50.99_{\pm 16.41}$	$61.06_{\pm 2.63}$
	GraphCL	$60.17_{\pm 9.46}$	$56.65_{\pm 2.44}$	$62.34_{\pm 4.37}$	$70.67_{\pm 3.89}$	$27.00_{\pm 0.78}$	$58.28_{\pm 9.07}$	$53.70_{\pm 10.84}$	$59.07_{\pm 0.65}$
	SimGRACE	$62.33_{\pm 4.53}$	$55.67_{\pm 7.42}$	$60.74_{\pm 3.46}$	$72.27_{\pm 4.51}$	$26.21_{\pm 0.96}$	$29.81_{\pm 15.53}$	$27.10_{\pm 4.50}$	$43.52_{\pm 5.74}$
GPF-plus	DGI	$63.50_{\pm 5.68}$	$54.91_{\pm 7.08}$	$63.51_{\pm 2.89}$	$71.87_{\pm 5.55}$	$25.67_{\pm 1.75}$	$60.16_{\pm 3.91}$	$62.84_{\pm 9.18}$	$62.76_{\pm 4.22}$
	GraphMAE	$63.72_{\pm 5.43}$	$37.23_{\pm 18.02}$	$63.15_{\pm 4.75}$	$72.67_{\pm 1.89}$	$24.79_{\pm 1.80}$	$67.35_{\pm 13.04}$	$49.51_{\pm 9.62}$	$59.21_{\pm 6.53}$
	EdgePreGPPT	$68.13_{\pm 3.31}$	$60.68_{\pm 4.67}$	$61.89_{\pm 4.59}$	$70.93_{\pm 3.00}$	$23.04_{\pm 1.13}$	$68.20_{\pm 12.65}$	$24.01_{\pm 5.97}$	$64.46_{\pm 3.57}$
	EdgePreGprompt	$63.85_{\pm 5.88}$	$57.76_{\pm 6.85}$	$58.83_{\pm 2.28}$	$72.67_{\pm 4.04}$	$26.87_{\pm 1.89}$	$72.87_{\pm 10.17}$	$22.35_{\pm 2.72}$	$64.80_{\pm 3.45}$
	GraphCL	$60.18_{\pm 10.28}$	$57.43_{\pm 3.27}$	$63.08_{\pm 4.23}$	$73.87_{\pm 3.51}$	$25.79_{\pm 1.76}$	$60.70_{\pm 11.73}$	$57.59_{\pm 5.89}$	$52.78_{\pm 10.34}$
	SimGRACE	$60.43_{\pm 9.87}$	$58.99_{\pm 4.32}$	$63.10_{\pm 4.06}$	$72.80_{\pm 3.30}$	$24.58_{\pm 1.68}$	$61.77_{\pm 16.67}$	$71.54_{\pm 14.81}$	$61.97_{\pm 3.07}$
MTG (Ours)	DGI	$62.88_{\pm 5.80}$	$59.50_{\pm 6.50}$	$63.98_{\pm 4.53}$	$75.27_{\pm 3.79}$	$28.92_{\pm 3.12}$	$51.69_{\pm 10.82}$	$76.37_{\pm 8.11}$	$59.12_{\pm 5.79}$
	GraphMAE	$63.72_{\pm 5.43}$	$32.22_{\pm 17.95}$	$63.48_{\pm 5.44}$	$72.80_{\pm 3.08}$	$28.08_{\pm 3.50}$	$54.48_{\pm 15.95}$	$55.99_{\pm 2.91}$	$57.82_{\pm 4.93}$
	EdgePreGPPT	$69.15_{\pm 4.09}$	$58.22_{\pm 5.93}$	$61.57_{\pm 3.42}$	$76.07_{\pm 2.17}$	$35.08_{\pm 3.28}$	$57.37_{\pm 13.16}$	$67.53_{\pm 7.79}$	$56.20_{\pm 3.52}$
	EdgePreGprompt	$63.88_{\pm 7.79}$	$55.35_{\pm 8.37}$	$67.22_{\pm 4.38}$	$81.60_{\pm 4.53}$	$25.21_{\pm 3.33}$	$57.45_{\pm 7.23}$	$55.12_{\pm 3.27}$	$66.07_{\pm 2.39}$
	GraphCL	$62.45_{\pm 6.81}$	$57.60_{\pm 3.54}$	$62.70_{\pm 3.70}$	$71.87_{\pm 5.60}$	$22.79_{\pm 2.59}$	$48.53_{\pm 6.77}$	$61.30_{\pm 6.75}$	$60.49_{\pm 3.62}$
	SimGRACE	$65.87_{\pm 4.73}$	$63.11_{\pm 1.88}$	$70.10_{\pm 1.12}$	$74.27_{\pm 3.69}$	$21.17_{\pm 1.10}$	$71.84_{\pm 2.75}$	$68.40_{\pm 12.20}$	$58.82_{\pm 4.24}$

Table 30: F1-score on 5-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	$61.53_{\pm 3.82}$	$46.72_{\pm 7.81}$	$60.12_{\pm 2.97}$	$68.77_{\pm 4.51}$	$21.36_{\pm 3.02}$	$50.40_{\pm 4.10}$	$49.85_{\pm 2.02}$	$61.16_{\pm 2.22}$
Fine-tuning	DGI	$37.27_{\pm 7.49}$	$60.77_{\pm 1.97}$	$59.48_{\pm 2.21}$	$69.60_{\pm 5.01}$	$23.93_{\pm 3.09}$	$46.29_{\pm 3.39}$	$21.92_{\pm 9.07}$	$53.50_{\pm 13.04}$
	GraphMAE	$48.52_{\pm 12.68}$	$47.93_{\pm 10.31}$	$58.76_{\pm 5.50}$	$61.19_{\pm 5.53}$	$21.10_{\pm 2.81}$	$46.30_{\pm 3.29}$	$34.64_{\pm 16.14}$	$60.62_{\pm 2.52}$
	EdgePreGPPT	$59.95_{\pm 10.96}$	$55.01_{\pm 7.38}$	$58.97_{\pm 1.92}$	$66.40_{\pm 3.88}$	$23.13_{\pm 4.27}$	$46.90_{\pm 4.51}$	$49.13_{\pm 5.88}$	$60.32_{\pm 3.32}$
	EdgePreGprompt	$61.87_{\pm 3.87}$	$58.08_{\pm 3.74}$	$57.72_{\pm 3.83}$	$67.43_{\pm 5.46}$	$24.23_{\pm 3.11}$	$45.22_{\pm 2.84}$	$29.92_{\pm 15.69}$	$55.91_{\pm 9.68}$
	GraphCL	$64.48_{\pm 3.90}$	$46.25_{\pm 8.10}$	$60.16_{\pm 1.94}$	$72.18_{\pm 2.14}$	$20.23_{\pm 2.20}$	$50.69_{\pm 2.30}$	$44.56_{\pm 0.94}$	$61.72_{\pm 1.92}$
	SimGRACE	$58.80_{\pm 4.71}$	$46.94_{\pm 6.45}$	$60.31_{\pm 2.61}$	$67.55_{\pm 3.52}$	$19.51_{\pm 3.69}$	$50.79_{\pm 10.90}$	$42.64_{\pm 2.95}$	$54.10_{\pm 1.53}$
GPPTPrompt	DGI	$33.29_{\pm 0.37}$	$44.26_{\pm 14.53}$	$51.17_{\pm 11.32}$	$62.76_{\pm 9.57}$	$15.91_{\pm 3.76}$	$31.80_{\pm 11.62}$	$40.57_{\pm 12.66}$	$49.10_{\pm 8.69}$
	GraphMAE	$33.15_{\pm 0.33}$	$16.01_{\pm 6.20}$	$56.63_{\pm 5.43}$	$61.64_{\pm 4.83}$	$18.70_{\pm 2.65}$	$36.00_{\pm 12.12}$	$45.03_{\pm 4.50}$	$48.97_{\pm 2.70}$
	EdgePreGPPT	$65.92_{\pm 3.82}$	$44.90_{\pm 15.76}$	$39.47_{\pm 4.13}$	$60.33_{\pm 5.13}$	$17.40_{\pm 1.52}$	$45.39_{\pm 5.02}$	$38.16_{\pm 10.87}$	$50.83_{\pm 4.76}$
	EdgePreGprompt	$33.15_{\pm 0.33}$	$19.75_{\pm 6.76}$	$44.10_{\pm 11.48}$	$33.67_{\pm 17.23}$	$19.46_{\pm 2.35}$	$42.06_{\pm 4.74}$	$41.95_{\pm 7.64}$	$55.63_{\pm 2.35}$
	GraphCL	$55.16_{\pm 14.72}$	$55.24_{\pm 1.26}$	$40.33_{\pm 5.84}$	$65.16_{\pm 6.09}$	$19.05_{\pm 1.73}$	$46.02_{\pm 10.48}$	$39.32_{\pm 8.94}$	$54.10_{\pm 2.94}$
	SimGRACE	$59.66_{\pm 9.14}$	$54.05_{\pm 3.81}$	$54.48_{\pm 10.12}$	$62.48_{\pm 5.78}$	$19.62_{\pm 1.57}$	$49.41_{\pm 6.89}$	$50.14_{\pm 3.71}$	$54.01_{\pm 3.13}$
Gprompt	DGI	$50.47_{\pm 12.00}$	$60.36_{\pm 1.92}$	$58.82_{\pm 4.98}$	$59.86_{\pm 14.59}$	$19.68_{\pm 2.10}$	$43.07_{\pm 4.95}$	$52.76_{\pm 11.53}$	$52.64_{\pm 4.28}$
	GraphMAE	$51.26_{\pm 10.62}$	$16.37_{\pm 0.04}$	$56.08_{\pm 6.65}$	$61.49_{\pm 7.64}$	$18.48_{\pm 2.23}$	$46.44_{\pm 3.42}$	$48.92_{\pm 4.67}$	$49.81_{\pm 4.74}$
	EdgePreGPPT	$66.49_{\pm 4.14}$	$57.50_{\pm 2.41}$	$61.89_{\pm 1.74}$	$61.80_{\pm 11.24}$	$16.31_{\pm 1.68}$	$48.08_{\pm 4.53}$	$46.79_{\pm 5.22}$	$51.83_{\pm 4.02}$
	EdgePreGprompt	$50.18_{\pm 11.72}$	$58.02_{\pm 1.57}$	$53.81_{\pm 6.27}$	$70.82_{\pm 1.39}$	$18.42_{\pm 3.25}$	$46.63_{\pm 4.27}$	$47.81_{\pm 6.75}$	$55.47_{\pm 3.78}$
	GraphCL	$60.64_{\pm 11.30}$	$59.83_{\pm 4.00}$	$57.86_{\pm 1.00}$	$60.49_{\pm 10.88}$	$19.29_{\pm 2.26}$	$47.44_{\pm 7.08}$	$50.30_{\pm 4.88}$	$57.01_{\pm 2.25}$
	SimGRACE	$53.16_{\pm 12.04}$	$60.09_{\pm 4.57}$	$59.49_{\pm 2.71}$	$59.62_{\pm 6.15}$	$20.75_{\pm 2.19}$	$49.12_{\pm 4.71}$	$47.73_{\pm 7.10}$	$52.97_{\pm 4.17}$
All-in-one	DGI	$59.53_{\pm 8.82}$	$57.32_{\pm 5.81}$	$70.60_{\pm 4.58}$	$76.43_{\pm 1.37}$	$17.66_{\pm 2.59}$	$44.71_{\pm 2.85}$	$52.91_{\pm 7.30}$	$52.63_{\pm 1.53}$
	GraphMAE	$58.72_{\pm 10.43}$	$17.43_{\pm 5.21}$	$66.84_{\pm 3.16}$	$76.59_{\pm 2.86}$	$15.28_{\pm 0.64}$	$45.02_{\pm 5.39}$	$50.25_{\pm 14.39}$	$57.48_{\pm 1.16}$
	EdgePreGPPT	$63.43_{\pm 2.34}$	$57.44_{\pm 6.03}$	$69.85_{\pm 2.07}$	$78.29_{\pm 1.00}$	$18.20_{\pm 5.07}$	$47.22_{\pm 3.60}$	$46.61_{\pm 3.47}$	$53.27_{\pm 3.89}$
	EdgePreGprompt	$58.91_{\pm 9.07}$	$56.52_{\pm 8.78}$	$61.72_{\pm 1.75}$	$69.01_{\pm 4.36}$	$19.95_{\pm 1.44}$	$48.65_{\pm 5.08}$	$44.03_{\pm 8.76}$	$59.90_{\pm 1.32}$
	GraphCL	$61.47_{\pm 5.33}$	$56.68_{\pm 4.02}$	$68.30_{\pm 5.95}$	$67.42_{\pm 3.89}$	$22.30_{\pm 0.99}$	$47.14_{\pm 2.57}$	$45.65_{\pm 4.16}$	$53.88_{\pm 4.86}$
	SimGRACE	$62.85_{\pm 2.99}$	$55.22_{\pm 0.57}$	$67.22_{\pm 4.47}$	$71.75_{\pm 2.16}$	$21.06_{\pm 0.77}$	$51.71_{\pm 4.63}$	$48.74_{\pm 3.72}$	$40.65_{\pm 6.82}$
GPF	DGI	$58.67_{\pm 10.64}$	$59.18_{\pm 6.42}$	$58.89_{\pm 2.28}$	$67.69_{\pm 4.30}$	$18.41_{\pm 2.04}$	$43.64_{\pm 5.83}$	$56.38_{\pm 8.64}$	$43.94_{\pm 8.16}$
	GraphMAE	$65.59_{\pm 4.08}$	$14.71_{\pm 5.24}$	$60.67_{\pm 2.38}$	$65.88_{\pm 1.64}$	$19.47_{\pm 2.56}$	$45.45_{\pm 8.77}$	$49.37_{\pm 6.61}$	$57.04_{\pm 2.26}$
	EdgePreGPPT	$67.73_{\pm 5.58}$	$57.75_{\pm 6.17}$	$61.14_{\pm 4.21}$	$63.56_{\pm 4.73}$	$19.89_{\pm 2.46}$	$43.97_{\pm 2.80}$	$30.06_{\pm 8.48}$	$55.24_{\pm 3.51}$
	EdgePreGprompt	$61.77_{\pm 4.51}$	$56.12_{\pm 5.23}$	$60.54_{\pm 2.00}$	$72.05_{\pm 3.02}$	$19.31_{\pm 2.59}$	$50.99_{\pm 3.91}$	$39.07_{\pm 4.52}$	$58.14_{\pm 3.07}$
	GraphCL	$58.58_{\pm 10.31}$	$54.42_{\pm 4.20}$	$58.69_{\pm 2.31}$	$67.61_{\pm 3.98}$	$21.36_{\pm 3.31}$	$50.58_{\pm 4.57}$	$47.55_{\pm 4.52}$	$40.86_{\pm 7.24}$
	SimGRACE	$61.27_{\pm 4.49}$	$54.22_{\pm 8.92}$	$59.44_{\pm 2.34}$	$67.56_{\pm 6.57}$	$23.55_{\pm 3.05}$	$25.14_{\pm 14.15}$	$26.14_{\pm 5.64}$	$33.65_{\pm 9.49}$
GPF-plus	DGI	$63.03_{\pm 5.96}$	$53.58_{\pm 8.50}$	$59.69_{\pm 2.41}$	$70.16_{\pm 4.83}$	$21.63_{\pm 0.48}$	$52.70_{\pm 2.44}$	$57.31_{\pm 6.19}$	$57.95_{\pm 3.57}$
	GraphMAE	$62.85_{\pm 5.85}$	$17.18_{\pm 6.87}$	$60.52_{\pm 2.74}$	$67.74_{\pm 2.44}$	$21.75_{\pm 1.83}$	$45.30_{\pm 2.36}$	$45.36_{\pm 7.65}$	$56.39_{\pm 7.29}$
	EdgePreGPPT	$67.66_{\pm 3.65}$	$60.45_{\pm 4.30}$	$58.78_{\pm 2.63}$	$68.26_{\pm 2.34}$	$20.46_{\pm 1.22}$	$45.80_{\pm 4.06}$	$21.07_{\pm 7.39}$	$60.10_{\pm 2.62}$
	EdgePreGprompt	$63.38_{\pm 6.30}$	$56.87_{\pm 7.08}$	$55.28_{\pm 2.90}$	$69.63_{\pm 3.06}$	$23.58_{\pm 3.05}$	$44.79_{\pm 1.97}$	$19.35_{\pm 3.96}$	$60.64_{\pm 2.13}$
	GraphCL	$58.81_{\pm 10.82}$	$57.11_{\pm 3.43}$	$60.02_{\pm 2.04}$	$69.79_{\pm 2.97}$	$22.03_{\pm 2.03}$	$47.74_{\pm 4.03}$	$53.04_{\pm 4.37}$	$47.52_{\pm 13.56}$
	SimGRACE	$59.10_{\pm 10.57}$	$58.56_{\pm 4.61}$	$60.30_{\pm 2.04}$	$69.16_{\pm 2.98}$	$22.60_{\pm 2.36}$	$51.12_{\pm 11.99}$	$43.15_{\pm 1.95}$	$53.98_{\pm 1.48}$
MTG (Ours)	DGI	$58.80_{\pm 10.80}$	$58.88_{\pm 6.78}$	$59.78_{\pm 3.39}$	$72.24_{\pm 3.73}$	$23.05_{\pm 3.33}$	$46.73_{\pm 4.46}$	$62.01_{\pm 4.48}$	$53.78_{\pm 4.13}$
	GraphMAE	$64.07_{\pm 5.62}$	$17.54_{\pm 6.80}$	$65.49_{\pm 4.00}$	$69.38_{\pm 3.75}$	$21.24_{\pm 5.03}$	$44.41_{\pm 7.24}$	$52.73_{\pm 2.09}$	$56.56_{\pm 3.40}$
	EdgePreGPPT	$69.06_{\pm 4.09}$	$59.90_{\pm 6.04}$	$59.41_{\pm 1.33}$	$74.08_{\pm 4.85}$	$33.66_{\pm 2.53}$	$48.74_{\pm 5.48}$	$59.70_{\pm 4.96}$	$56.92_{\pm 3.19}$
	EdgePreGprompt	$63.98_{\pm 6.56}$	$53.86_{\pm 9.80}$	$66.56_{\pm 3.42}$	$78.10_{\pm 4.12}$	$22.53_{\pm 3.93}$	$43.02_{\pm 4.67}$	$48.00_{\pm 2.37}$	$59.27_{\pm 3.15}$
	GraphCL	$61.24_{\pm 8.88}$	$56.93_{\pm 3.44}$	$64.43_{\pm 2.09}$	$69.40_{\pm 5.10}$	$17.40_{\pm 2.11}$	$43.12_{\pm 3.78}$	$55.88_{\pm 6.25}$	$56.50_{\pm 3.28}$
	SimGRACE	$65.40_{\pm 4.68}$	$62.26_{\pm 1.13}$	$69.70_{\pm 0.97}$	$70.64_{\pm 3.75}$	$17.82_{\pm 1.43}$	$50.79_{\pm 5.89}$	$60.50_{\pm 8.24}$	$55.08_{\pm 2.50}$

Table 31: AUROC on 5-shot graph classification.

Adaptation	Pre-training	IMDB-B	COLLAB	PROTEINS	MUTAG	ENZYMES	COX2	BZR	D&D
Supervised	-	$66.66_{\pm 4.29}$	$65.98_{\pm 11.45}$	$60.55_{\pm 1.27}$	$68.80_{\pm 5.31}$	$59.06_{\pm 0.90}$	$53.16_{\pm 1.88}$	$63.16_{\pm 1.91}$	$64.75_{\pm 3.62}$
Fine-tuning	DGI	$50.02_{\pm 0.04}$	$81.66_{\pm 0.61}$	$60.91_{\pm 1.77}$	$68.55_{\pm 5.83}$	$58.93_{\pm 1.27}$	$51.76_{\pm 2.64}$	$52.81_{\pm 2.18}$	$63.61_{\pm 4.41}$
	GraphMAE	$57.68_{\pm 8.77}$	$69.01_{\pm 10.12}$	$59.92_{\pm 2.11}$	$74.33_{\pm 4.22}$	$58.39_{\pm 0.82}$	$51.25_{\pm 2.21}$	$57.13_{\pm 7.05}$	$64.45_{\pm 3.57}$
	EdgePreGPPT	$61.63_{\pm 14.66}$	$76.71_{\pm 9.50}$	$60.88_{\pm 1.39}$	$72.84_{\pm 3.58}$	$59.64_{\pm 0.79}$	$53.48_{\pm 3.98}$	$60.58_{\pm 6.74}$	$63.82_{\pm 3.08}$
	EdgePreGprompt	$66.78_{\pm 4.49}$	$81.42_{\pm 0.65}$	$56.93_{\pm 4.04}$	$72.62_{\pm 1.98}$	$59.65_{\pm 1.08}$	$51.34_{\pm 2.69}$	$50.75_{\pm 1.31}$	$62.82_{\pm 3.74}$
	GraphCL	$69.68_{\pm 5.15}$	$76.71_{\pm 2.70}$	$61.80_{\pm 2.09}$	$70.63_{\pm 4.24}$	$57.49_{\pm 0.69}$	$55.91_{\pm 4.67}$	$50.86_{\pm 1.72}$	$64.37_{\pm 3.44}$
	SimGRACE	$64.69_{\pm 7.42}$	$72.81_{\pm 3.58}$	$61.49_{\pm 2.04}$	$71.48_{\pm 3.93}$	$55.98_{\pm 2.05}$	$61.95_{\pm 2.47}$	$48.92_{\pm 2.16}$	$57.39_{\pm 2.98}$
GPPTPrompt	DGI	$50.48_{\pm 0.70}$	$70.62_{\pm 7.22}$	$55.92_{\pm 14.32}$	$79.97_{\pm 4.66}$	$52.20_{\pm 2.26}$	$51.22_{\pm 2.96}$	$47.75_{\pm 9.82}$	$53.81_{\pm 9.82}$
	GraphMAE	$50.63_{\pm 0.56}$	$50.56_{\pm 3.78}$	$60.56_{\pm 6.96}$	$77.76_{\pm 5.00}$	$53.63_{\pm 1.06}$	$50.60_{\pm 6.83}$	$53.55_{\pm 6.08}$	$58.60_{\pm 5.56}$
	EdgePreGPPT	$69.02_{\pm 4.94}$	$68.78_{\pm 9.45}$	$68.50_{\pm 9.36}$	$72.44_{\pm 6.57}$	$53.43_{\pm 1.03}$	$53.02_{\pm 8.32}$	$47.93_{\pm 5.96}$	$51.76_{\pm 5.53}$
	EdgePreGprompt	$50.63_{\pm 0.56}$	$56.85_{\pm 1.07}$	$45.95_{\pm 18.22}$	$24.75_{\pm 25.84}$	$53.66_{\pm 1.73}$	$52.56_{\pm 3.29}$	$49.31_{\pm 5.04}$	$61.01_{\pm 3.67}$
	GraphCL	$59.75_{\pm 10.88}$	$76.73_{\pm 1.15}$	$69.17_{\pm 8.03}$	$77.80_{\pm 5.27}$	$53.37_{\pm 1.20}$	$56.22_{\pm 8.12}$	$50.45_{\pm 6.37}$	$57.89_{\pm 3.91}$
	SimGRACE	$61.35_{\pm 10.03}$	$70.93_{\pm 4.01}$	$59.76_{\pm 13.58}$	$77.27_{\pm 7.67}$	$53.63_{\pm 1.06}$	$57.78_{\pm 6.95}$	$59.49_{\pm 6.13}$	$57.49_{\pm 4.92}$
Gprompt	DGI	$53.91_{\pm 13.03}$	$70.52_{\pm 5.99}$	$57.55_{\pm 7.36}$	$64.03_{\pm 19.95}$	$53.87_{\pm 2.02}$	$50.06_{\pm 2.93}$	$56.36_{\pm 6.92}$	$54.21_{\pm 4.23}$
	GraphMAE	$55.46_{\pm 7.98}$	$44.83_{\pm 0.15}$	$54.75_{\pm 7.17}$	$63.37_{\pm 9.48}$	$52.82_{\pm 1.11}$	$49.40_{\pm 6.97}$	$54.99_{\pm 5.88}$	$51.72_{\pm 5.54}$
	EdgePreGPPT	$70.26_{\pm 4.44}$	$73.55_{\pm 4.01}$	$67.07_{\pm 0.80}$	$66.04_{\pm 10.33}$	$53.32_{\pm 2.77}$	$52.47_{\pm 6.25}$	$52.16_{\pm 2.83}$	$53.26_{\pm 4.39}$
	EdgePreGprompt	$56.66_{\pm 8.48}$	$75.59_{\pm 3.42}$	$56.24_{\pm 7.10}$	$74.43_{\pm 4.69}$	$54.57_{\pm 2.69}$	$50.60_{\pm 3.01}$	$53.90_{\pm 6.28}$	$56.36_{\pm 5.42}$
	GraphCL	$52.62_{\pm 12.32}$	$75.45_{\pm 3.52}$	$61.03_{\pm 5.77}$	$64.43_{\pm 10.15}$	$54.04_{\pm 2.81}$	$50.64_{\pm 7.63}$	$53.71_{\pm 2.16}$	$56.22_{\pm 3.97}$
	SimGRACE	$57.24_{\pm 12.21}$	$73.41_{\pm 3.61}$	$60.58_{\pm 5.34}$	$64.67_{\pm 8.28}$	$55.02_{\pm 1.57}$	$56.79_{\pm 5.66}$	$56.83_{\pm 6.49}$	$53.99_{\pm 6.27}$
All-in-one	DGI	$61.66_{\pm 12.01}$	$80.99_{\pm 1.96}$	$79.39_{\pm 1.76}$	$83.38_{\pm 1.95}$	$58.71_{\pm 0.88}$	$49.33_{\pm 1.56}$	$52.97_{\pm 11.15}$	$52.92_{\pm 1.58}$
	GraphMAE	$60.44_{\pm 13.36}$	$50.00_{\pm 0.00}$	$73.82_{\pm 1.89}$	$84.45_{\pm 1.78}$	$53.75_{\pm 1.13}$	$54.24_{\pm 10.01}$	$55.73_{\pm 11.76}$	$64.55_{\pm 1.63}$
	EdgePreGPPT	$69.53_{\pm 3.98}$	$81.50_{\pm 1.45}$	$77.60_{\pm 0.81}$	$85.46_{\pm 2.74}$	$54.18_{\pm 3.09}$	$52.22_{\pm 5.14}$	$39.69_{\pm 2.87}$	$58.41_{\pm 0.60}$
	EdgePreGprompt	$61.78_{\pm 11.31}$	$81.93_{\pm 1.28}$	$66.26_{\pm 1.84}$	$82.07_{\pm 1.61}$	$55.09_{\pm 2.15}$	$61.86_{\pm 2.18}$	$50.64_{\pm 4.57}$	$64.95_{\pm 0.92}$
	GraphCL	$65.55_{\pm 5.57}$	$77.86_{\pm 4.19}$	$73.65_{\pm 4.67}$	$77.59_{\pm 1.92}$	$59.50_{\pm 1.10}$	$49.77_{\pm 1.51}$	$52.63_{\pm 2.91}$	$59.95_{\pm 4.37}$
	SimGRACE	$66.52_{\pm 3.18}$	$75.84_{\pm 0.40}$	$77.13_{\pm 0.44}$	$79.68_{\pm 4.23}$	$55.76_{\pm 0.82}$	$62.42_{\pm 1.89}$	$54.60_{\pm 2.87}$	$55.43_{\pm 0.66}$
GPF	DGI	$61.28_{\pm 14.33}$	$80.91_{\pm 1.35}$	$59.36_{\pm 1.27}$	$68.29_{\pm 4.20}$	$57.38_{\pm 0.69}$	$55.65_{\pm 2.89}$	$69.56_{\pm 8.43}$	$52.10_{\pm 1.18}$
	GraphMAE	$71.36_{\pm 4.86}$	$50.00_{\pm 0.00}$	$60.79_{\pm 1.35}$	$73.75_{\pm 6.18}$	$57.51_{\pm 1.83}$	$55.80_{\pm 3.40}$	$57.87_{\pm 6.08}$	$59.78_{\pm 6.01}$
	EdgePreGPPT	$72.09_{\pm 6.35}$	$81.60_{\pm 1.83}$	$63.33_{\pm 6.33}$	$71.14_{\pm 4.08}$	$57.09_{\pm 1.47}$	$45.44_{\pm 3.51}$	$40.30_{\pm 2.41}$	$57.56_{\pm 1.55}$
	EdgePreGprompt	$66.56_{\pm 4.93}$	$82.11_{\pm 0.58}$	$59.32_{\pm 1.46}$	$78.72_{\pm 2.44}$	$56.88_{\pm 0.88}$	$57.11_{\pm 3.21}$	$38.86_{\pm 3.08}$	$59.75_{\pm 5.73}$
	GraphCL	$61.77_{\pm 14.84}$	$76.32_{\pm 2.61}$	$58.70_{\pm 2.93}$	$75.54_{\pm 4.82}$	$57.23_{\pm 1.32}$	$50.87_{\pm 1.24}$	$54.74_{\pm 5.41}$	$51.02_{\pm 2.03}$
	SimGRACE	$66.69_{\pm 4.51}$	$78.63_{\pm 3.50}$	$60.84_{\pm 0.83}$	$73.60_{\pm 4.98}$	$57.64_{\pm 0.64}$	$55.74_{\pm 2.01}$	$44.52_{\pm 2.78}$	$48.14_{\pm 1.17}$
GPF-plus	DGI	$67.76_{\pm 7.65}$	$80.06_{\pm 1.94}$	$60.56_{\pm 1.11}$	$75.93_{\pm 5.59}$	$57.99_{\pm 1.37}$	$56.06_{\pm 2.35}$	$67.92_{\pm 4.56}$	$62.75_{\pm 3.59}$
	GraphMAE	$68.56_{\pm 7.52}$	$50.00_{\pm 0.00}$	$60.88_{\pm 1.41}$	$74.54_{\pm 6.62}$	$56.78_{\pm 1.51}$	$48.93_{\pm 2.41}$	$54.11_{\pm 8.51}$	$62.56_{\pm 3.54}$
	EdgePreGPPT	$71.71_{\pm 3.89}$	$82.51_{\pm 0.50}$	$59.30_{\pm 1.96}$	$77.03_{\pm 2.00}$	$56.17_{\pm 0.68}$	$49.53_{\pm 4.57}$	$49.78_{\pm 0.43}$	$63.49_{\pm 4.00}$
	EdgePreGprompt	$67.36_{\pm 7.32}$	$81.68_{\pm 0.93}$	$54.63_{\pm 1.53}$	$75.31_{\pm 2.84}$	$60.04_{\pm 0.95}$	$50.95_{\pm 1.90}$	$48.60_{\pm 2.80}$	$64.20_{\pm 3.47}$
	GraphCL	$61.80_{\pm 14.82}$	$80.57_{\pm 0.98}$	$61.23_{\pm 1.64}$	$69.47_{\pm 3.56}$	$57.95_{\pm 1.27}$	$51.87_{\pm 1.80}$	$64.97_{\pm 3.46}$	$61.94_{\pm 4.01}$
	SimGRACE	$61.85_{\pm 14.76}$	$77.71_{\pm 3.28}$	$61.07_{\pm 1.57}$	$70.13_{\pm 2.87}$	$57.30_{\pm 0.70}$	$63.15_{\pm 2.98}$	$49.11_{\pm 1.78}$	$57.21_{\pm 2.94}$
MTG (Ours)	DGI	$64.24_{\pm 11.54}$	$81.70_{\pm 1.83}$	$62.28_{\pm 1.40}$	$79.88_{\pm 4.08}$	$60.12_{\pm 2.82}$	$46.65_{\pm 4.45}$	$70.58_{\pm 3.21}$	$53.66_{\pm 5.36}$
	GraphMAE	$65.61_{\pm 10.49}$	$50.00_{\pm 0.00}$	$61.23_{\pm 1.51}$	$76.70_{\pm 5.49}$	$58.65_{\pm 1.63}$	$52.08_{\pm 3.43}$	$59.66_{\pm 2.52}$	$57.52_{\pm 4.36}$
	EdgePreGPPT	$72.82_{\pm 5.50}$	$81.57_{\pm 1.80}$	$63.41_{\pm 2.01}$	$79.87_{\pm 5.60}$	$65.42_{\pm 2.00}$	$59.49_{\pm 5.74}$	$65.22_{\pm 4.05}$	$60.56_{\pm 2.93}$
	EdgePreGprompt	$63.17_{\pm 12.42}$	$81.92_{\pm 0.63}$	$67.57_{\pm 0.51}$	$85.99_{\pm 2.00}$	$59.08_{\pm 1.78}$	$47.32_{\pm 2.89}$	$50.15_{\pm 2.87}$	$62.66_{\pm 4.63}$
	GraphCL	$67.29_{\pm 4.11}$	$79.70_{\pm 2.70}$	$63.87_{\pm 1.69}$	$75.64_{\pm 5.17}$	$52.41_{\pm 1.80}$	$45.45_{\pm 5.87}$	$61.89_{\pm 6.88}$	$58.00_{\pm 3.33}$
	SimGRACE	$67.02_{\pm 4.32}$	$80.98_{\pm 0.46}$	$67.31_{\pm 1.03}$	$71.30_{\pm 1.90}$	$51.89_{\pm 1.51}$	$59.90_{\pm 5.00}$	$69.34_{\pm 3.12}$	$55.56_{\pm 1.76}$

Table 32: 1-shot node classification accuracy (%) on Wisconsin for various backbone models. Supervised learning baselines: GCN: 41.60$_{\pm 3.10}$, GAT: 34.51$_{\pm 18.02}$, GraphSAGE: 25.37$_{\pm 5.61}$, GIN: 28.91$_{\pm 11.51}$, GT: 20.91$_{\pm 7.07}$.

Fine-tuning

Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	37.49$_{\pm 5.13}$	36.80$_{\pm 7.17}$	35.31$_{\pm 9.31}$	**40.69**$_{\pm 4.13}$	33.94$_{\pm 7.74}$	37.37$_{\pm 3.68}$
GAT	16.00$_{\pm 6.24}$	**37.60**$_{\pm 10.69}$	20.00$_{\pm 3.82}$	33.37$_{\pm 4.76}$	18.86$_{\pm 1.88}$	28.00$_{\pm 9.40}$
GraphSAGE	40.69$_{\pm 9.46}$	**43.77**$_{\pm 12.43}$	26.06$_{\pm 5.38}$	29.94$_{\pm 3.75}$	36.57$_{\pm 4.88}$	9.37$_{\pm 2.72}$
GIN	**34.29**$_{\pm 10.40}$	26.29$_{\pm 7.81}$	25.14$_{\pm 7.70}$	33.49$_{\pm 7.69}$	22.63$_{\pm 8.62}$	16.46$_{\pm 4.53}$
GT	25.71$_{\pm 3.07}$	**39.77**$_{\pm 8.42}$	23.20$_{\pm 2.65}$	28.23$_{\pm 6.64}$	11.77$_{\pm 1.06}$	14.51$_{\pm 5.08}$

GPPT

Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	29.94$_{\pm 10.40}$	29.83$_{\pm 9.34}$	23.89$_{\pm 5.40}$	**30.40**$_{\pm 6.81}$	25.03$_{\pm 5.37}$	29.83$_{\pm 6.44}$
GAT	22.17$_{\pm 6.13}$	33.94$_{\pm 7.76}$	23.43$_{\pm 4.46}$	**37.94**$_{\pm 7.11}$	26.86$_{\pm 6.12}$	29.83$_{\pm 8.04}$
GraphSAGE	26.51$_{\pm 8.00}$	**30.51**$_{\pm 5.40}$	21.49$_{\pm 5.17}$	24.23$_{\pm 6.55}$	20.91$_{\pm 7.11}$	25.37$_{\pm 7.22}$
GIN	**27.20**$_{\pm 5.34}$	24.00$_{\pm 3.29}$	21.14$_{\pm 1.84}$	20.46$_{\pm 2.79}$	19.09$_{\pm 14.19}$	19.54$_{\pm 15.85}$
GT	27.20$_{\pm 10.48}$	**29.83**$_{\pm 5.80}$	28.00$_{\pm 6.01}$	23.31$_{\pm 3.01}$	27.66$_{\pm 0.69}$	25.03$_{\pm 5.43}$

Gprompt

Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	67.71$_{\pm 9.92}$	67.62$_{\pm 18.06}$	67.37$_{\pm 12.32}$	74.38$_{\pm 13.15}$	**77.07**$_{\pm 5.93}$	65.38$_{\pm 13.70}$
GAT	58.25$_{\pm 13.83}$	67.77$_{\pm 15.91}$	**94.17**$_{\pm 2.26}$	84.28$_{\pm 3.63}$	80.11$_{\pm 16.65}$	57.18$_{\pm 12.60}$
GraphSAGE	66.48$_{\pm 12.88}$	83.49$_{\pm 15.93}$	**87.52**$_{\pm 3.79}$	82.16$_{\pm 2.64}$	65.50$_{\pm 6.48}$	72.61$_{\pm 5.97}$
GIN	45.47$_{\pm 9.62}$	37.72$_{\pm 15.00}$	58.36$_{\pm 15.10}$	59.29$_{\pm 12.72}$	59.03$_{\pm 19.98}$	**71.80**$_{\pm 11.66}$
GT	56.03$_{\pm 7.33}$	73.50$_{\pm 9.72}$	76.97$_{\pm 13.39}$	**80.07**$_{\pm 2.84}$	59.31$_{\pm 10.17}$	69.30$_{\pm 10.57}$

All-in-one

Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	56.02$_{\pm 13.12}$	57.54$_{\pm 10.66}$	**66.29**$_{\pm 19.11}$	59.18$_{\pm 12.30}$	39.14$_{\pm 1.17}$	55.56$_{\pm 14.70}$
GAT	69.44$_{\pm 5.19}$	36.25$_{\pm 10.63}$	91.25$_{\pm 4.33}$	**92.65**$_{\pm 3.75}$	42.85$_{\pm 9.16}$	36.61$_{\pm 14.86}$
GraphSAGE	74.88$_{\pm 19.77}$	87.55$_{\pm 3.78}$	98.60$_{\pm 0.87}$	**99.12**$_{\pm 0.64}$	67.28$_{\pm 20.14}$	86.18$_{\pm 9.68}$
GIN	54.02$_{\pm 15.90}$	35.31$_{\pm 15.69}$	**58.77**$_{\pm 13.43}$	57.07$_{\pm 12.51}$	45.94$_{\pm 9.52}$	25.30$_{\pm 14.83}$
GT	60.22$_{\pm 11.02}$	97.42$_{\pm 2.13}$	94.61$_{\pm 1.73}$	**97.88**$_{\pm 2.24}$	51.33$_{\pm 15.56}$	83.26$_{\pm 16.29}$

GPF

Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	62.69$_{\pm 13.96}$	76.84$_{\pm 10.50}$	**78.35**$_{\pm 4.07}$	75.20$_{\pm 13.22}$	51.60$_{\pm 20.06}$	60.81$_{\pm 26.52}$
GAT	65.14$_{\pm 11.94}$	74.39$_{\pm 16.46}$	**94.96**$_{\pm 1.17}$	76.60$_{\pm 10.48}$	74.97$_{\pm 17.06}$	60.57$_{\pm 14.43}$
GraphSAGE	68.12$_{\pm 13.96}$	67.66$_{\pm 13.37}$	74.06$_{\pm 14.59}$	72.45$_{\pm 10.14}$	59.69$_{\pm 21.37}$	**78.37**$_{\pm 14.84}$
GIN	47.11$_{\pm 11.28}$	49.47$_{\pm 14.94}$	**66.99**$_{\pm 17.76}$	54.96$_{\pm 12.35}$	28.77$_{\pm 22.76}$	23.55$_{\pm 14.37}$
GT	39.85$_{\pm 4.83}$	71.26$_{\pm 14.43}$	72.67$_{\pm 13.36}$	**81.33**$_{\pm 3.41}$	78.19$_{\pm 2.19}$	67.90$_{\pm 10.53}$

GPF-plus

Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	74.68$_{\pm 11.81}$	**82.11**$_{\pm 13.95}$	72.66$_{\pm 12.05}$	78.76$_{\pm 13.63}$	52.35$_{\pm 19.69}$	73.49$_{\pm 14.17}$
GAT	93.34$_{\pm 6.13}$	83.28$_{\pm 12.20}$	**95.24**$_{\pm 1.58}$	92.03$_{\pm 4.64}$	87.49$_{\pm 7.17}$	63.06$_{\pm 18.45}$
GraphSAGE	71.83$_{\pm 17.50}$	85.47$_{\pm 1.45}$	**97.30**$_{\pm 1.68}$	80.35$_{\pm 14.37}$	50.35$_{\pm 8.91}$	71.95$_{\pm 9.43}$
GIN	57.55$_{\pm 16.90}$	66.88$_{\pm 14.55}$	**82.79**$_{\pm 10.37}$	74.40$_{\pm 13.11}$	29.94$_{\pm 22.25}$	24.30$_{\pm 17.29}$
GT	72.41$_{\pm 11.37}$	**95.19**$_{\pm 4.04}$	87.76$_{\pm 15.73}$	80.58$_{\pm 11.56}$	57.75$_{\pm 19.89}$	63.79$_{\pm 17.44}$

MTG (Ours)

Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	67.72$_{\pm 10.19}$	**83.32**$_{\pm 12.46}$	73.80$_{\pm 9.56}$	72.75$_{\pm 11.21}$	48.41$_{\pm 16.10}$	72.98$_{\pm 9.75}$
GAT	59.87$_{\pm 9.77}$	82.16$_{\pm 11.33}$	**95.84**$_{\pm 1.15}$	81.99$_{\pm 12.78}$	77.01$_{\pm 12.03}$	61.75$_{\pm 13.22}$
GraphSAGE	76.90$_{\pm 9.36}$	**99.29**$_{\pm 1.41}$	87.23$_{\pm 4.91}$	72.63$_{\pm 10.16}$	62.44$_{\pm 19.82}$	63.50$_{\pm 17.62}$
GIN	59.93$_{\pm 13.84}$	69.96$_{\pm 10.90}$	78.87$_{\pm 16.32}$	**83.57**$_{\pm 10.78}$	37.34$_{\pm 15.08}$	34.48$_{\pm 15.47}$
GT	58.35$_{\pm 10.12}$	**98.85**$_{\pm 1.02}$	84.45$_{\pm 8.03}$	72.99$_{\pm 10.45}$	93.45$_{\pm 4.15}$	89.24$_{\pm 2.66}$

Table 33: 1-shot graph classification accuracy (%) on PROTEINS for various backbone models. Supervised learning baselines: GCN: 56.36$_{\pm 7.97}$, GAT: 48.34$_{\pm 9.96}$, GraphSAGE: 60.54$_{\pm 2.95}$, GIN: 59.66$_{\pm 1.12}$, GT: 61.44$_{\pm 2.48}$.

Fine-tuning						
Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	60.00$_{\pm 4.48}$	62.40$_{\pm 1.494}$	58.27$_{\pm 10.66}$	61.84$_{\pm 2.59}$	**63.44**$_{\pm 3.64}$	60.07$_{\pm 3.21}$
GAT	58.34$_{\pm 6.52}$	61.06$_{\pm 4.13}$	**63.75**$_{\pm 3.71}$	54.09$_{\pm 4.03}$	60.04$_{\pm 3.06}$	58.65$_{\pm 6.71}$
GraphSAGE	60.70$_{\pm 4.08}$	60.56$_{\pm 5.12}$	61.60$_{\pm 1.78}$	**63.21**$_{\pm 1.80}$	61.80$_{\pm 3.77}$	58.56$_{\pm 1.84}$
GIN	59.71$_{\pm 1.16}$	59.75$_{\pm 1.22}$	64.83$_{\pm 3.56}$	**65.35**$_{\pm 2.36}$	58.52$_{\pm 0.77}$	58.49$_{\pm 0.80}$
GT	53.87$_{\pm 4.81}$	60.00$_{\pm 3.99}$	**64.92**$_{\pm 3.19}$	56.58$_{\pm 3.28}$	62.88$_{\pm 1.82}$	60.00$_{\pm 1.60}$
GPPT						
Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	60.81$_{\pm 1.55}$	60.72$_{\pm 1.70}$	**60.92**$_{\pm 2.47}$	57.03$_{\pm 4.55}$	59.24$_{\pm 1.01}$	55.42$_{\pm 8.81}$
GAT	57.71$_{\pm 8.98}$	57.80$_{\pm 10.55}$	**58.04**$_{\pm 9.92}$	54.97$_{\pm 7.45}$	52.29$_{\pm 7.83}$	55.15$_{\pm 9.84}$
GraphSAGE	56.56$_{\pm 6.73}$	57.73$_{\pm 7.95}$	**58.63**$_{\pm 11.78}$	56.94$_{\pm 5.67}$	58.00$_{\pm 7.80}$	54.74$_{\pm 6.59}$
GIN	**62.27**$_{\pm 2.54}$	52.13$_{\pm 11.00}$	52.52$_{\pm 6.97}$	55.53$_{\pm 8.92}$	55.78$_{\pm 7.22}$	55.78$_{\pm 7.22}$
GT	53.08$_{\pm 7.56}$	57.35$_{\pm 8.58}$	**60.27**$_{\pm 3.92}$	55.51$_{\pm 7.68}$	56.18$_{\pm 5.79}$	55.87$_{\pm 7.69}$
Gprompt						
Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	56.61$_{\pm 7.93}$	57.66$_{\pm 12.56}$	**59.17**$_{\pm 11.26}$	55.55$_{\pm 8.17}$	55.51$_{\pm 10.73}$	57.53$_{\pm 11.05}$
GAT	61.08$_{\pm 6.19}$	63.03$_{\pm 2.61}$	**64.47**$_{\pm 4.30}$	61.48$_{\pm 3.34}$	59.12$_{\pm 6.84}$	58.13$_{\pm 7.27}$
GraphSAGE	61.35$_{\pm 2.21}$	59.48$_{\pm 9.19}$	60.92$_{\pm 3.16}$	**63.30**$_{\pm 1.43}$	55.26$_{\pm 2.61}$	63.21$_{\pm 2.66}$
GIN	54.36$_{\pm 5.18}$	46.97$_{\pm 11.45}$	55.82$_{\pm 5.35}$	**57.84**$_{\pm 10.75}$	56.92$_{\pm 11.77}$	46.16$_{\pm 10.78}$
GT	56.65$_{\pm 5.81}$	60.99$_{\pm 1.62}$	**61.87**$_{\pm 5.60}$	55.33$_{\pm 3.69}$	54.81$_{\pm 7.62}$	58.97$_{\pm 1.16}$
All-in-one						
Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	62.58$_{\pm 7.07}$	**66.49**$_{\pm 6.26}$	65.71$_{\pm 5.49}$	61.82$_{\pm 7.53}$	64.36$_{\pm 7.30}$	61.17$_{\pm 1.73}$
GAT	60.04$_{\pm 3.84}$	60.00$_{\pm 6.04}$	62.11$_{\pm 2.85}$	**63.21**$_{\pm 2.22}$	58.36$_{\pm 4.93}$	59.37$_{\pm 5.59}$
GraphSAGE	59.53$_{\pm 4.94}$	60.70$_{\pm 4.89}$	63.12$_{\pm 1.59}$	59.98$_{\pm 8.46}$	**62.22**$_{\pm 3.81}$	62.04$_{\pm 2.07}$
GIN	**61.55**$_{\pm 3.02}$	60.72$_{\pm 4.32}$	59.78$_{\pm 3.28}$	58.29$_{\pm 12.13}$	40.81$_{\pm 1.04}$	59.19$_{\pm 1.04}$
GT	57.39$_{\pm 3.66}$	58.92$_{\pm 6.61}$	62.61$_{\pm 4.08}$	60.20$_{\pm 7.55}$	**62.81**$_{\pm 1.63}$	50.52$_{\pm 6.17}$
GPF						
Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	59.17$_{\pm 3.63}$	58.65$_{\pm 8.49}$	62.54$_{\pm 2.55}$	61.82$_{\pm 2.61}$	**63.91**$_{\pm 3.26}$	63.35$_{\pm 3.69}$
GAT	**63.01**$_{\pm 1.22}$	59.62$_{\pm 5.38}$	47.53$_{\pm 9.42}$	47.71$_{\pm 7.14}$	56.65$_{\pm 5.15}$	57.91$_{\pm 3.10}$
GraphSAGE	52.72$_{\pm 6.43}$	59.17$_{\pm 2.22}$	61.73$_{\pm 2.59}$	**64.54**$_{\pm 3.73}$	62.27$_{\pm 2.60}$	58.00$_{\pm 3.81}$
GIN	**61.19**$_{\pm 3.39}$	54.34$_{\pm 8.61}$	60.58$_{\pm 6.80}$	62.34$_{\pm 1.19}$	59.19$_{\pm 1.04}$	59.19$_{\pm 1.04}$
GT	**65.80**$_{\pm 7.42}$	60.16$_{\pm 5.81}$	64.54$_{\pm 7.18}$	61.21$_{\pm 2.91}$	58.74$_{\pm 5.51}$	59.57$_{\pm 2.93}$
GPF-plus						
Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	61.26$_{\pm 3.06}$	62.49$_{\pm 2.05}$	**63.06**$_{\pm 2.55}$	61.33$_{\pm 2.81}$	59.75$_{\pm 7.95}$	62.92$_{\pm 2.78}$
GAT	56.20$_{\pm 12.87}$	57.35$_{\pm 11.28}$	56.25$_{\pm 8.61}$	53.24$_{\pm 4.79}$	57.48$_{\pm 11.74}$	**57.48**$_{\pm 9.63}$
GraphSAGE	56.22$_{\pm 9.08}$	57.55$_{\pm 10.56}$	56.31$_{\pm 9.26}$	**57.71**$_{\pm 9.60}$	53.89$_{\pm 9.47}$	55.89$_{\pm 4.30}$
GIN	62.22$_{\pm 2.49}$	61.75$_{\pm 3.58}$	57.33$_{\pm 9.24}$	**64.99**$_{\pm 0.82}$	59.19$_{\pm 1.04}$	59.19$_{\pm 1.04}$
GT	53.39$_{\pm 5.23}$	57.37$_{\pm 10.95}$	57.39$_{\pm 11.88}$	52.61$_{\pm 5.30}$	**57.62**$_{\pm 12.27}$	56.16$_{\pm 5.07}$
MTG (Ours)						
Model	DGI	GraphMAE	EdgePreGPPT	EdgePreGprompt	GraphCL	SimGRACE
GCN	62.78$_{\pm 2.36}$	59.62$_{\pm 6.41}$	62.71$_{\pm 2.30}$	**65.66**$_{\pm 1.56}$	63.70$_{\pm 2.87}$	66.98$_{\pm 2.17}$
GAT	61.48$_{\pm 2.14}$	60.38$_{\pm 4.81}$	53.46$_{\pm 7.80}$	**63.53**$_{\pm 1.25}$	49.12$_{\pm 6.49}$	52.63$_{\pm 4.14}$
GraphSAGE	61.98$_{\pm 2.03}$	58.85$_{\pm 1.66}$	65.24$_{\pm 1.83}$	**65.88**$_{\pm 0.58}$	62.20$_{\pm 3.35}$	60.94$_{\pm 9.92}$
GIN	61.55$_{\pm 1.47}$	60.52$_{\pm 3.27}$	**65.64**$_{\pm 6.34}$	63.10$_{\pm 0.39}$	60.19$_{\pm 1.04}$	59.69$_{\pm 3.21}$
GT	61.83$_{\pm 6.86}$	57.19$_{\pm 8.75}$	63.64$_{\pm 5.55}$	59.91$_{\pm 5.85}$	**66.08**$_{\pm 2.70}$	61.72$_{\pm 0.83}$