RouterGFM: Inductive Routing and Local Expertise Transfer for Graph Learning

Anonymous authors
Paper under double-blind review

Abstract

Reusing pretrained *graph neural networks* (GNNs) requires deciding which experts suit a new application and how to combine their predictions across graph contexts. Application-level selection does not resolve instance-specific contributions, while learning a new local router can require more target supervision than is available. We propose RouterGFM, a graph mixture-of-experts framework that combines inductive expert selection with context-guided local integration. A shared application–expert scorer over a heterogeneous context graph selects experts from metadata and historical evaluations before executing candidate GNNs on target data. After fitting the selected task heads, expert-conditioned retrieval transfers historical local performance differences into instance-dependent prediction weights, while the pretrained GNNs and router parameters remain fixed. Our conditional analysis separates omitted local expertise, performance-estimation error, and mixture temperature, and characterizes the limits of metadata-only expert insertion. Experiments on nine datasets cover node classification, link prediction, graph classification, and graph regression. In the five-shot classification diagnosis with expert predictions fixed, local integration reduces average and worst-cell Brier loss by 11.7% and 17.3% relative to application-level weighting. Further analyses distinguish useful expert coverage from accurate integration and identify when unreliable historical evidence limits transfer.

1 Introduction

Graph foundation models aim to make pretrained knowledge reusable across graph domains and prediction tasks (Liu et al., 2024; Wang et al., 2024; 2025b). However, differences in node attributes, neighborhood structure, and supervision make the usefulness of a pretrained model application-dependent. A diverse pretrained expert pool preserves different model choices, but its use requires deciding both which experts to select and how to combine them.

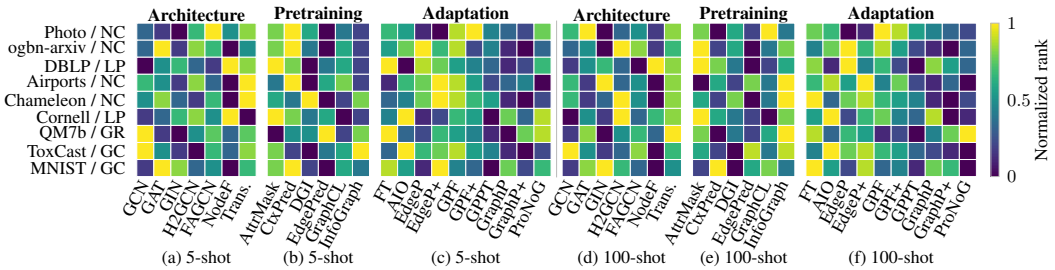

Figure 1: Configuration rankings across nine graph datasets. Each cell ranks the best reported configuration containing the indicated factor, with other factors varied and omitted results excluded. Ranks are normalized within each dataset and panel (higher: better; ties use average ranks), minimizing QM7b MAE and maximizing accuracy or ROC-AUC elsewhere. FT denotes full fine-tuning; LP panels follow the method-specific conventions in Appendix B.3.

Our study of 7 architectures, 6 self-supervised pretraining objectives, and 10 adaptation strategies across 9 datasets illustrates the difficulty of a fixed model choice (Figure 1). On Photo (Shchur et al., 2018), the architecture of the best observed configuration changes from FAGCN (Bo et al., 2021) at

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

five shots (81.27%) to GAT (Veličković et al., 2018) at 100 shots (89.35%). On Chameleon (Rozemberczki et al., 2021), it changes from Transformer (Vaswani et al., 2017) (35.39%) to H2GCN (Zhu et al., 2020) (42.30%). These configuration-level differences motivate application- and budget-dependent expert selection. A second decision remains after selection: an expert that performs well on average need not be the most useful expert in every local context. We ask: *which pretrained experts suit a new application, and how should their contributions depend on graph context?*

The first question concerns *inductive expert selection*. Graph mixtures exploit different processing rules and source models (Wang et al., 2023; Xia & Huang, 2026; Liu et al., 2025a), while model-selection methods learn application–model relationships from historical evaluations (Park et al., 2023a; Cai et al., 2026). Metadata and historical relations allow a shared scorer to represent new applications and checkpoints before executing candidates on target data.

The second question concerns *context-guided local integration*. Existing graph mixtures assign instance-dependent weights (Zeng et al., 2024; Ma et al., 2024; Meena et al., 2026), but a small target support set may not cover the contexts needed to learn a reliable gate. Historical evaluations record where support-adapted experts perform better or worse than their own application averages. The challenge is to transfer these local performance patterns to newly fitted task heads, rather than learn a gate from the limited supervision available for each target application.

We propose ROUTERGFM, a graph mixture-of-experts framework that separates application-level expert selection from local prediction integration. A heterogeneous context graph links applications, experts, and their construction factors, supporting a shared scorer that selects a small expert team before target-side candidate execution. Expert-conditioned retrieval combines the scorer's application-level estimates with centered historical performance differences to determine each selected expert's contribution to a query. Application-masked training hides the historical evaluations being predicted; at deployment, only the selected task heads are fitted on target support examples.

Our contribution is to reuse historical expert evaluations for both selection and integration across changing applications and expert teams, without fitting a target-specific router. The theoretical analysis relates prediction risk to omitted expertise and estimation reliability, and distinguishes representing an unseen expert from accurately predicting its behavior. Experiments span 9 datasets and 4 tasks; fixed-team controls isolate integration gains from changes in expert checkpoints, fitted heads, and predictions. Team-size, source-calibration, and incomplete-evidence analyses examine useful expert coverage and the reliability of transferred local information.

2 RELATED WORK

Graph foundation models. OFA, GFT, and GIT share knowledge across graph tasks through unified representations or transferable tree structures (Liu et al., 2024; Wang et al., 2024; 2025b). GraphAlign aligns features across graphs, MDGFM aligns topology, and BRIDGE combines feature alignment with mixture-based prompting (Hou et al., 2024; Wang et al., 2025a; Yuan et al., 2025). AutoGFM customizes architectures within a shared network (Chen et al., 2025). Graph prompting generally adapts a given pretrained GNN (Sun et al., 2022; Liu et al., 2023; Sun et al., 2023; Fang et al., 2023; Fu et al., 2025); ProNoG conditions prompts on non-homophilic patterns (Yu et al., 2025), while CTP supports cross-graph prompting without target-side parameter updates (Chen et al., 2027). ROUTERGFM instead reuses independently pretrained GNNs through expert selection and prediction integration after fitting task heads on target support examples.

Graph mixtures and model reuse. GMoE and Node-MoE select aggregation ranges or graph filters, while Mowst combines feature- and structure-based experts (Wang et al., 2023; Han et al., 2024; Zeng et al., 2024). GraphMoRE and GeoMoE use geometric experts; TopExpert specializes by molecular topology, and GraphMETRO aligns experts for distribution shifts (Guo et al., 2025; Cao et al., 2026; Kim et al., 2023; Wu et al., 2024). AnyGraph routes across graphs, OMOG integrates source-specific models, and GMoPE learns prompt-conditioned experts (Xia & Huang, 2026; Liu et al., 2025a; Wang et al., 2025c). KDEM/PPEM improve expert merging, whereas OGMM uses generated graphs to merge and adapt heterogeneous models (Liu et al., 2025b; Wang et al., 2026). Our framework retains the pretrained parameters and transfers historical evidence into prediction weights rather than merging the pretrained model parameters.

Inductive model selection. MetaGL learns evaluation-free graph-model selection, and GLEMOS benchmarks this problem (Park et al., 2023a;b). ModelLens recommends unseen models for unseen datasets and supplies candidate pools to downstream routers (Cai et al., 2026). GraphRouter uses a heterogeneous task–query–LLM graph, while RouteLLM learns from preference supervision (Feng et al., 2025; Ong et al., 2025). LogME and Model Spider assess pretrained-model transferability, and Quick-Tune also selects fine-tuning configurations (You et al., 2021; Zhang et al., 2023; Pineda Arango et al., 2024). We extend recommendation to transferring historical local evaluations across changing teams of support-adapted graph predictors on new target applications.

Local expert integration. Link-MoE learns node-pair-specific weights, and SAGMM uses topology-aware gating over diverse architectures, including a frozen-expert variant (Ma et al., 2024; Meena et al., 2026). META-DES learns local classifier competence for dynamic ensemble selection (Cruz et al., 2015). Our expert-conditioned retrieval transfers local performance across applications after task-head fitting, without learning a new target gate.

3 METHODOLOGY

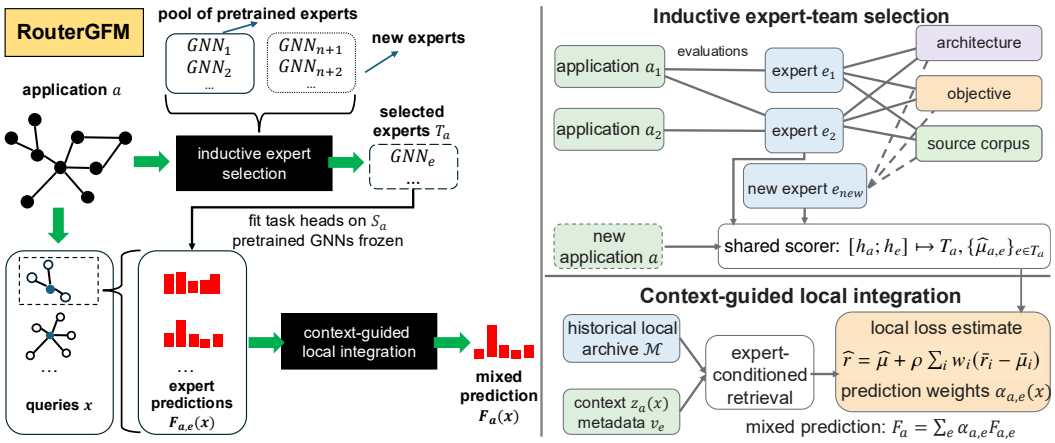

Figure 2: Overview of ROUTERGFM: embeddings h_a, h_e yield the team and global loss estimates; historical local corrections refine prediction weights after selected task heads are fitted.

ROUTERGFM first selects pretrained GNNs for an application and then combines their predictions according to each query's graph context (Figure 2). Application-level loss estimates determine the team and provide a baseline for local corrections retrieved from historical evaluations. Deployment fits only the selected task heads, keeping pretrained GNNs and router parameters fixed.

3.1 PROBLEM FORMULATION

An application a specifies a graph dataset, task, and labeled support set $S_a = \{(x_i, y_i)\}$. For an attributed graph $G = (V, E, \mathbf{X})$, V and E are its node and edge sets, and $\mathbf{X}$ is its node-feature matrix. An instance x is a graph with a marked node or unordered node pair for node or link prediction, or an entire graph for graph-level prediction. Its output y is a class or link label, multi-label vector, or numerical property; Q_a contains queries with unknown outputs.

Node and link contexts are subgraphs with the query node or endpoints marked (Zhang & Chen, 2018); a graph-level context is the input graph. The label-free vector $z_a(x) \in \mathbb{R}^{d_z}$ summarizes structure, node features, endpoint roles, and task information, where d_z is the common descriptor dimension. It is computed without candidate execution and is not an expert embedding.

Let $\mathcal{E}$ be the pretrained expert pool and $\mathcal{E}_a \subseteq \mathcal{E}$ its nonempty subset compatible with the target task, input processing, and task-head adaptation (Appendix B). Given an integer expert budget $K \geq 1$, we select $T_a \subseteq \mathcal{E}_a$ with $1 \leq |T_a| \leq K$. Expert e has a frozen encoder f_{Θ_e} and a head $g_{\omega_{a,e}}$ fitted on S_a, where Θ_e and $\omega_{a,e}$ denote their respective parameters. Writing $F_{a,e}(x) = g_{\omega_{a,e}}(f_{\Theta_e}(x))$, with

the task readout included in f_{Θ_e}, we combine predictions as

$$F_a(x) = \sum_{e \in T_a} \alpha_{a,e}(x) F_{a,e}(x), \qquad \alpha_{a,e}(x) \geq 0, \quad \sum_{e \in T_a} \alpha_{a,e}(x) = 1. \tag{1}$$

Each $\alpha_{a,e}(x)$ is a scalar query weight; predictions share class/assay coordinates or regression units.

3.2 Historical Evaluations and Context Graph

For each historical application a, task heads are fitted on S_a and evaluated on a separate labeled diagnostic set D_a. For the nonempty subset $D_{a,e} \subseteq D_a$ evaluated by e, $\ell_a(y, p)$ measures prediction p's task-specific error against output y. The recorded application-average loss is

$$\bar{\mu}_{a,e} = \frac{1}{|D_{a,e}|} \sum_{(x,y) \in D_{a,e}} \ell_a(y, F_{a,e}(x)). \tag{2}$$

A bar denotes a measured diagnostic average, a hat denotes a router estimate, and the corresponding unadorned symbol denotes a population risk. Average and context-specific losses supervise selection and integration, respectively, under compatible loss conventions.

The context graph $\mathcal{H}$ has application/dataset, expert, architecture, and pretraining-objective node types, distinct from input-graph nodes. Evaluation edges link applications to evaluated experts and carry $\bar{\mu}_{a,e}$ and the valid evaluation count $|D_{a,e}|$. Construction relations connect experts to their architecture, objective, and source corpus, with source corpora represented by application/dataset nodes. Metadata describes expert construction; evaluations describe adapted prediction behavior.

3.3 Inductive Expert Selection

A context-graph node v has metadata features x_v, combining a frozen text embedding with numerical descriptors. Application metadata describes the graph data, task, and support budget; expert metadata describes the architecture, objective, pretraining data, and adaptation rule. A type-specific projection initializes $h_v^{(0)} = P_{\kappa(v)}(x_v) \in \mathbb{R}^{d_h}$, where $\kappa(v)$ is the node type and d_h is the shared hidden dimension. A heterogeneous GNN then updates these vectors over $L_{\mathcal{H}}$ layers:

$$h_v^{(l+1)} = h_v^{(l)} + \sigma\left(\sum_{r \in \mathcal{R}} \sum_{u \in N_r(v)} \psi_r^{(l)}(h_u^{(l)}, h_v^{(l)}, b_{uv}) \right). \tag{3}$$

Here $l = 0, \ldots, L_{\mathcal{H}} - 1$, $\mathcal{R}$ contains the directed construction and evaluation relations and their reverses, and $N_r(v)$ is the set of incoming neighbors of v under relation r. The edge vector b_{uv} contains the available attributes of the relation, $\psi_r^{(l)}$ is its learned message function, and σ is a nonlinear activation (Schlichtkrull et al., 2018; Hu et al., 2020c). The final vectors $h_v = h_v^{(L_{\mathcal{H}})}$ give the application and expert embeddings h_a, h_e in Figure 2, not candidate-GNN query features.

The scorer $q_\theta : \mathbb{R}^{2d_h} \to \mathbb{R}$ predicts each compatible expert's average loss after head fitting:

$$\widehat{\mu}_{a,e} = \zeta(q_\theta([h_a; h_e])), \qquad T_a = \mathrm{TopK}_{e \in \mathcal{E}_a}(-\widehat{\mu}_{a,e}). \tag{4}$$

Brackets denote vector concatenation, $\zeta(t) = \log(1 + \exp t)$ ensures nonnegative estimates, and TopK returns up to K indices with the largest argument, hence the smallest predicted losses. We use θ collectively for the projections, context-graph encoder, and scorer. New applications enter through metadata without evaluation edges; new experts enter through metadata and construction relations, adding factor nodes as needed. Shared functions return T_a and $\{\widehat{\mu}_{a,e}\}_{e \in T_a}$ without router retraining or candidate GNN execution on the target application during selection.

3.4 Context-Guided Local Integration

For a random query and output (X, Y) from application a, define the fitted expert's conditional risk at descriptor z and its average over the query distribution:

$$r_{a,e}(z) = \mathbb{E}[\ell_a(Y, F_{a,e}(X)) \mid z_a(X) = z], \qquad \mu_{a,e} = \mathbb{E}[r_{a,e}(z_a(X))].$$

Expectations hold support, fitted predictors, router information, and team fixed; random instance X differs from feature matrix $\mathbf{X}$. We transfer adapted experts' relative local performance, not their application-specific predictions, labels, or fitted head parameters.

Let b index a historical application, and partition its diagnostic descriptors into B_b cells $C_{b,j}$ with representative vectors $c_{b,j}$, for $j = 1, \ldots, B_b$. Cells are shared across experts and group similar descriptors, not necessarily adjacent graph nodes. For $D_{b,j,e} = \{(x, y) \in D_{b,e} : z_b(x) \in C_{b,j}\}$ with count $n_{b,j,e} = |D_{b,j,e}| > 0$, the recorded local loss is

$$\bar{r}_{b,j,e} = \frac{1}{n_{b,j,e}} \sum_{(x,y)\in D_{b,j,e}} \ell_b(y, F_{b,e}(x)). \tag{5}$$

Archive $\mathcal{M}$ stores records indexed by i, comprising application b_i, expert e_i, representative c_i, local loss $\bar{r}_i$, application average $\bar{\mu}_i = \bar{\mu}_{b_i,e_i}$, and count n_i. Descriptor standardization uses historical training applications only; link contexts exclude candidate and held-out positive edges.

The learned key map $k_\phi : \mathbb{R}^{d_z} \times \mathbb{R}^{d_h} \to \mathbb{R}^{d_k}$ embeds context and expert metadata into d_k dimensions, with parameters ϕ. Its expert input is $v_e = h_e^{(0)}$, the projected metadata before message passing, rather than the evaluation-conditioned final embedding h_e. For query context $z = z_a(x)$, the Euclidean distance to record i is $d_i(z, e) = \|k_\phi(z, v_e) - k_\phi(c_i, v_{e_i})\|_2$. Let $\mathcal{N}_J(z, e)$ contain the indices of up to $J \geq 1$ nearest compatible records from other applications, matching task/loss family, support budget, and normalization convention, with a per-application cap. For a nonempty retrieved set, a scalar kernel bandwidth $\beta > 0$ defines normalized record weights:

$$w_i(z, e) = \frac{\exp(-d_i(z, e)^2/\beta^2)}{\sum_{i'\in\mathcal{N}_J(z,e)} \exp(-d_{i'}(z, e)^2/\beta^2)}, \qquad i \in \mathcal{N}_J(z, e). \tag{6}$$

Smaller β favors closer records; larger β spreads weight across the retrieved set, independently of record counts. Expert conditioning retrieves different evidence for different team members, including related experts' records when an expert lacks its own evaluations.

The local estimate adds the retrieved centered losses to the target expert's application-level estimate:

$$\widehat{r}_{a,e}(z) = \widehat{\mu}_{a,e} + \rho \sum_{i\in\mathcal{N}_J(z,e)} w_i(z, e)(\bar{r}_i - \bar{\mu}_i), \qquad \rho \in [0, 1]. \tag{7}$$

Correction strength $\rho = 1$ uses local evidence fully, whereas $\rho = 0$ recovers the global estimate. For an empty retrieved set, the correction is defined to be zero without evaluating Equation 6. Negative centered loss means better-than-average local behavior, not superiority to other experts, so comparisons still require $\widehat{\mu}_{a,e}$. Centering removes additive offsets but does not guarantee that local performance patterns survive task-head adaptation, even when task and loss conventions match.

A separate scalar mixture temperature $\tau > 0$ converts estimated losses into prediction weights:

$$\alpha_{a,e}(x) = \frac{\exp(-\widehat{r}_{a,e}(z_a(x))/\tau)}{\sum_{e'\in T_a} \exp(-\widehat{r}_{a,e'}(z_a(x))/\tau)}. \tag{8}$$

Record weights w_i normalize across history; prediction weights $\alpha_{a,e}$ normalize across the team. Smaller τ favors lower estimated losses; larger τ approaches uniform mixing, without changing retrieval. The probability simplex $\Delta(T_a) = \{\alpha : \alpha_e \geq 0,\ \sum_{e\in T_a} \alpha_e = 1\}$ contains distributions over the team. For fixed z, Equation 8 minimizes the following entropy-regularized objective:

$$\min_{\alpha\in\Delta(T_a)} \sum_{e\in T_a} \alpha_e \widehat{r}_{a,e}(z) + \tau \sum_{e\in T_a} \alpha_e \log \alpha_e.$$

The same τ weights negative entropy ($0 \log 0 = 0$), trading estimated individual-expert loss against concentrated weights rather than directly optimizing mixture loss. Equation 1 combines expert outputs, not retrieval keys, without query-specific parameter updates.

3.5 Optimization and Inference

Let $\mathcal{A}_{\text{tr}}$ be the historical training applications, and treat each $b \in \mathcal{A}_{\text{tr}}$ in turn as a pseudo-target. Before encoding and retrieval, we hide its evaluation edges and reverses, local records, and label-derived aggregates; all hatted training quantities below use this masked information. The hidden

evaluations supervise predictions from other applications, teaching cross-application correspondence after historical task-head fitting. For the nonempty evaluated set $\Omega_b = \{e \in \mathcal{E}_b : |D_{b,e}| > 0\}$, we supervise application averages and context-dependent losses with the following objectives:

$$\mathcal{L}_{\text{glob}}(b) = \frac{1}{|\Omega_b|} \sum_{e \in \Omega_b} \text{Huber}(\widehat{\mu}_{b,e}, \bar{\mu}_{b,e}) + \lambda_r \mathcal{L}_{\text{rank}}(b), \tag{9}$$

$$\mathcal{L}_{\text{loc}}(b) = \frac{1}{|\Omega_b|} \sum_{e \in \Omega_b} \frac{1}{|D_{b,e}|} \sum_{(x,y) \in D_{b,e}} \big(\widehat{r}_{b,e}(z_b(x)) - \ell_b(y, F_{b,e}(x))\big)^2. \tag{10}$$

Huber regression penalizes average-loss residuals quadratically near zero and linearly for large errors; $\lambda_r \geq 0$ weights the ranking loss relative to this regression term. Specifically, set $m_b = |\Omega_b|$ and let $\pi_b(1), \ldots, \pi_b(m_b)$ order evaluated experts by nondecreasing measured loss $\bar{\mu}_{b,e}$. The ListMLE objective (Xia et al., 2008) uses negative predicted losses as scores:

$$\mathcal{L}_{\text{rank}}(b) = -\sum_{j=1}^{m_b} \log \frac{\exp(-\widehat{\mu}_{b,\pi_b(j)})}{\sum_{t=j}^{m_b} \exp(-\widehat{\mu}_{b,\pi_b(t)})}. \tag{11}$$

Each factor favors the next ranked expert over those remaining; either order is permitted for tied measured losses. Regression supervises loss scale; ranking supervises expert order for selection. Local supervision averages instance-level loss-estimation errors within experts and then across experts. With $\lambda_\ell \geq 0$ controlling local supervision, the trainable router parameters minimize

$$\min_{\theta,\phi} \mathcal{L} = \frac{1}{|\mathcal{A}_{\text{tr}}|} \sum_{b \in \mathcal{A}_{\text{tr}}} \big[\mathcal{L}_{\text{glob}}(b) + \lambda_\ell \mathcal{L}_{\text{loc}}(b)\big]. \tag{12}$$

Encoders, historical heads, and evaluations remain fixed while the router learns global scores and retrieval keys. Local supervision also reaches θ through $\widehat{\mu}$ and projected metadata v_e. Retrieved record identities are fixed within each update, but their kernel weights are differentiable; search keys are refreshed between updates. Hyperparameters $\beta, \tau, \rho, \lambda_r, \lambda_\ell$ and router checkpoints use historical validation, grouping tasks, budgets, and splits from each base dataset. At deployment, we select T_a, fit its task heads on S_a with encoders and router fixed, and predict Q_a using Equations 6–8 (Algorithm 1). Source calibration adds source-application evaluations without updating router parameters or using target-query labels from the new target application.

3.6 Theoretical Analysis

Fix the application, fitted experts, archive, and team, and abbreviate $Z = z_a(X)$, $\mathcal{E} = \mathcal{E}_a$, $T = T_a$, and $F = F_a$, suppressing a in losses and risks. The mixture risk is $R(F) = \mathbb{E}[\ell(Y, F(X))]$, and $R_{\text{loc}} = \mathbb{E}[\min_{e \in \mathcal{E}} r_e(Z)]$ is the risk of choosing a best individual expert separately for each context.

Theorem 1 (Expert coverage and local integration)**.** *Assume the loss is convex in its prediction argument, the risks are integrable, and mixture weights depend only on the permitted context. For realized team size $K = |T|$, define the omission cost $\Gamma_K = \mathbb{E}[\min_{e \in T} r_e(Z) - \min_{e \in \mathcal{E}} r_e(Z)]$ and the largest within-team estimation error $\epsilon(Z) = \max_{e \in T} |\widehat{r}_e(Z) - r_e(Z)|$. The mixture in Equation 1, with weights from Equation 8, satisfies the following conditional risk bound:*

$$R(F) - R_{\text{loc}} \leq \Gamma_K + 2\mathbb{E}[\epsilon(Z)] + \tau \log K. \tag{13}$$

The bound characterizes a selected team; application-average ranking need not minimize omitted expertise Γ_K. Error ϵ includes global estimation, local transfer, and shrinkage; the entropy term uses the same τ as Equation 8. The comparator is a context-wise individual expert, not an optimal convex ensemble, and the guarantee concerns convex prediction loss rather than accuracy or ROC-AUC. Under a compatibility condition on support-adapted predictors, Theorem 3 relates local error to retrieval distance, within-cell variation, cross-application change, and evaluation noise. Small retrieval distance alone does not guarantee reliable transfer; assumptions and proofs appear in Appendix A.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Applications and expert pool. We evaluate nine datasets across four tasks: node classification (NC) on Photo (Shchur et al., 2018), ogbn-arxiv (Hu et al., 2020a), Airports (Ribeiro et al., 2017), and Chameleon (Rozemberczki et al., 2021); link prediction (LP) on DBLP (Bojchevski & Günnemann, 2018) and Cornell (Craven et al., 1998); graph classification (GC) on ToxCast (Wu et al., 2018) and MNIST (LeCun et al., 1998; Dwivedi et al., 2023); and graph regression (GR) on QM7b (Montavon et al., 2013). The expert pool contains 588 pretrained GNNs from seven architectures, seven objectives, and twelve source datasets before task-compatibility filtering. Single-label classification uses five or 100 support examples per class, regression uses five or 100 support graphs, and ToxCast uses multitask support partitions. LP uses edge-based supervision; blocks A and B identify its two displayed result sets, not five- and 100-link support budgets. Appendix B summarizes adaptation and evaluation, with detailed implementation and configurations in the accompanying code.

Comparisons and metrics. Table 1 compares graph mixtures and matched-pool methods on all nine datasets. SAGMM-PE uses the matched frozen-expert inventory (Meena et al., 2026), and MetaGL-U uniformly averages experts selected by MetaGL (Park et al., 2023a). ROUTERGFM-G shares ROUTERGFM's selected team and fitted heads but uses application-level prediction weights; both select $K = 5$ experts. Prediction tables report five-seed means and standard deviations, using accuracy for single-label classification, ROC-AUC for LP and ToxCast, and raw-unit MAE for QM7b; mechanism diagnostics report point estimates. Hit@K measures inclusion of a best eligible expert, and regret@K is the loss gap between the best shortlisted and best eligible individual experts. The configuration study and task-scoped comparisons appear in Appendices D and E.

Table 1: Graph mixtures and matched-pool methods on nine datasets, with means and standard deviations over five seeds. Bold and underline identify the best and second distinct means within each displayed block. Shot labels apply to non-LP tasks; LP blocks A and B retain the reported results under the same edge-split convention (Appendix B.3), with repeated baseline entries.

Methods	Datasets: Homophilic graphs			Heterophilic graphs			Graph-level tasks		
	E-commerce	Academic	Academic	Transport	Web	Web	Chemistry	Chemistry	Vision
	Photo	**ogbn-arxiv**	**DBLP**	**Airports**	**Chameleon**	**Cornell**	**QM7b**	**ToxCast**	**MNIST**
	NC / Acc. ↑	NC / Acc. ↑	LP / AUC ↑	NC / Acc. ↑	NC / Acc. ↑	LP / AUC ↑	GR / MAE ↓	GC / AUC ↑	GC / Acc. ↑
5-shot / LP block A: graph mixture-of-experts models									
AnyGraph	$40.29_{\pm 6.39}$	$24.91_{\pm 4.18}$	$\underline{83.81}_{\pm 1.78}$	$37.08_{\pm 1.85}$	$25.92_{\pm 1.52}$	$57.20_{\pm 8.08}$	$15.41_{\pm 1.90}$	$50.32_{\pm 1.32}$	$24.59_{\pm 4.02}$
GMoE	$71.03_{\pm 5.25}$	$5.86_{\pm 6.24}$	$54.83_{\pm 0.77}$	$26.29_{\pm 1.78}$	$22.27_{\pm 2.35}$	$58.28_{\pm 10.40}$	$115.69_{\pm 1.04}$	$49.02_{\pm 2.86}$	$11.48_{\pm 4.00}$
Mowst	$75.67_{\pm 4.87}$	$23.61_{\pm 2.09}$	$55.32_{\pm 0.46}$	$28.44_{\pm 3.52}$	$29.73_{\pm 4.72}$	$64.21_{\pm 9.67}$	$114.77_{\pm 1.20}$	$49.87_{\pm 2.13}$	$22.88_{\pm 3.38}$
GraphMoRE	$77.74_{\pm 1.35}$	$21.70_{\pm 2.36}$	$54.10_{\pm 0.40}$	$33.97_{\pm 5.06}$	$28.22_{\pm 5.89}$	$57.19_{\pm 9.03}$	$26.07_{\pm 46.29}$	$48.79_{\pm 1.35}$	$23.40_{\pm 2.67}$
GMoPE	$81.48_{\pm 1.26}$	$32.91_{\pm 1.20}$	$81.22_{\pm 1.62}$	$40.42_{\pm 3.11}$	$36.78_{\pm 3.01}$	$75.91_{\pm 6.10}$	$5.64_{\pm 0.28}$	$57.01_{\pm 1.24}$	$\underline{30.66}_{\pm 2.31}$
5-shot / LP block A: matched-pool routing and frozen-expert mixtures									
SAGMM-PE	$82.10_{\pm 0.92}$	$33.15_{\pm 0.88}$	$82.80_{\pm 1.23}$	$41.35_{\pm 2.51}$	$\underline{37.05}_{\pm 2.74}$	$\underline{76.80}_{\pm 5.20}$	$5.47_{\pm 0.21}$	$57.00_{\pm 1.90}$	$30.11_{\pm 1.91}$
MetaGL-U	$80.48_{\pm 0.92}$	$31.20_{\pm 0.88}$	$80.43_{\pm 1.40}$	$39.95_{\pm 2.52}$	$34.82_{\pm 2.42}$	$73.36_{\pm 4.79}$	$5.73_{\pm 0.22}$	$56.35_{\pm 1.19}$	$28.90_{\pm 1.73}$
ROUTERGFM-G	$\underline{82.48}_{\pm 1.06}$	$\underline{33.47}_{\pm 0.88}$	$83.31_{\pm 1.26}$	$\underline{41.55}_{\pm 2.46}$	$36.43_{\pm 2.74}$	$76.33_{\pm 4.65}$	$\underline{5.40}_{\pm 0.26}$	$\underline{58.29}_{\pm 1.69}$	$30.37_{\pm 1.96}$
ROUTERGFM	$\mathbf{83.74}_{\pm 0.89}$	$\mathbf{42.92}_{\pm 1.55}$	$\mathbf{91.08}_{\pm 0.45}$	$\mathbf{47.09}_{\pm 1.87}$	$\mathbf{39.53}_{\pm 2.68}$	$\mathbf{79.50}_{\pm 4.48}$	$\mathbf{5.19}_{\pm 0.26}$	$\mathbf{60.98}_{\pm 0.96}$	$\mathbf{32.28}_{\pm 1.84}$
100-shot / LP block B: graph mixture-of-experts models									
AnyGraph	$84.00_{\pm 1.90}$	$50.24_{\pm 1.76}$	$\underline{83.81}_{\pm 1.78}$	$40.40_{\pm 2.89}$	$36.63_{\pm 2.21}$	$57.20_{\pm 8.08}$	$11.92_{\pm 0.73}$	$53.15_{\pm 0.65}$	$33.23_{\pm 1.60}$
GMoE	$85.76_{\pm 2.14}$	$25.68_{\pm 0.57}$	$54.83_{\pm 0.77}$	$38.18_{\pm 3.77}$	$33.09_{\pm 1.69}$	$58.28_{\pm 10.40}$	$113.74_{\pm 2.41}$	$55.51_{\pm 2.71}$	$34.65_{\pm 3.08}$
Mowst	$87.19_{\pm 0.77}$	$38.97_{\pm 0.64}$	$55.32_{\pm 0.46}$	$42.68_{\pm 1.89}$	$35.24_{\pm 0.93}$	$64.21_{\pm 9.67}$	$103.16_{\pm 1.53}$	$52.23_{\pm 2.52}$	$28.32_{\pm 6.74}$
GraphMoRE	$85.56_{\pm 0.72}$	$33.85_{\pm 0.92}$	$54.10_{\pm 0.40}$	$38.94_{\pm 1.90}$	$35.06_{\pm 0.70}$	$57.19_{\pm 9.03}$	$22.93_{\pm 39.40}$	$52.78_{\pm 0.47}$	$29.49_{\pm 0.44}$
GMoPE	$89.66_{\pm 0.57}$	$50.26_{\pm 1.03}$	$81.22_{\pm 1.62}$	$48.11_{\pm 1.91}$	$\underline{44.52}_{\pm 1.73}$	$75.91_{\pm 6.10}$	$5.04_{\pm 0.16}$	$59.85_{\pm 0.56}$	$\underline{44.63}_{\pm 1.52}$
100-shot / LP block B: matched-pool routing and frozen-expert mixtures									
SAGMM-PE	$89.43_{\pm 0.47}$	$49.72_{\pm 0.80}$	$82.80_{\pm 1.23}$	$48.54_{\pm 1.44}$	$44.35_{\pm 1.17}$	$\underline{76.80}_{\pm 5.20}$	$4.99_{\pm 0.11}$	$60.10_{\pm 1.46}$	$43.62_{\pm 1.13}$
MetaGL-U	$89.03_{\pm 0.45}$	$48.06_{\pm 0.84}$	$80.43_{\pm 1.40}$	$47.48_{\pm 1.76}$	$42.33_{\pm 1.16}$	$73.36_{\pm 4.79}$	$5.14_{\pm 0.12}$	$59.84_{\pm 1.50}$	$42.51_{\pm 1.05}$
ROUTERGFM-G	$\underline{89.71}_{\pm 0.40}$	$\underline{50.79}_{\pm 0.81}$	$83.31_{\pm 1.26}$	$\underline{49.23}_{\pm 1.53}$	$44.05_{\pm 1.45}$	$76.33_{\pm 4.65}$	$\underline{4.91}_{\pm 0.12}$	$\underline{60.99}_{\pm 1.22}$	$44.16_{\pm 1.26}$
ROUTERGFM	$\mathbf{94.89}_{\pm 0.34}$	$\mathbf{57.60}_{\pm 0.60}$	$\mathbf{92.04}_{\pm 0.75}$	$\mathbf{62.18}_{\pm 1.20}$	$\mathbf{59.14}_{\pm 1.90}$	$\mathbf{79.81}_{\pm 4.55}$	$\mathbf{4.42}_{\pm 0.43}$	$\mathbf{62.64}_{\pm 1.13}$	$\mathbf{46.18}_{\pm 1.12}$

4.2 PREDICTIVE PERFORMANCE ACROSS APPLICATIONS

Local integration complements application-level selection. ROUTERGFM has the best mean among the methods in every displayed comparison in Table 1. At five shots, local integration improves all displayed task metrics over ROUTERGFM-G, including QM7b MAE (5.40 to 5.19). Tox-

Cast ROC-AUC increases from 58.29% to 60.98% at five shots and from 60.99% to 62.64% at 100 shots, giving gains of 2.69 and 1.65 points. The broader adaptation catalog remains task-dependent: at 100 shots, ProNoG obtains lower QM7b MAE than RouterGFM (4.06 versus 4.42; Table 3).

Local integration remains useful at a larger support budget. Across the five single-label classification datasets, the mean accuracy gain over RouterGFM-G increases from 4.25 to 8.41 points. The increase is concentrated on Airports and Chameleon, where gains rise from 5.54 to 12.95 and from 3.10 to 15.09 points, respectively, while Photo improves from 1.26 to 5.18 points. In contrast, gains narrow on ogbn-arxiv (9.45 to 6.81 accuracy points) and ToxCast (2.69 to 1.65 ROC-AUC points). Integration remains useful at higher budgets, although expert selection can also change.

4.3 Inductive Selection and Expert Insertion

Historical relations improve expert recommendation. Figure 3(a) evaluates selection separately from prediction integration. At five shots, RouterGFM increases hit@5 from MetaGL's 72% to 84% and reduces regret@5 from 0.032 to 0.019 (40.6%). Model Spider matches MetaGL's hit@5 but has lower regret (0.029), distinguishing inclusion from the cost of missed selections. At 100 shots, RouterGFM reaches 88% hit@5 and 0.013 regret; LogME and Model Spider reach 80% hit@5, and Model Spider has the lowest competing regret (0.022).

Representing experts differs from predicting their behavior. With router parameters fixed, RouterGFM improves on MetaGL+metadata in all four insertion conditions, but loss rises from 0.225 for new configurations to 0.259 for unseen architectures and 0.263 under joint novelty (Figure 3(b)). This separates the ability to represent a new expert from the availability of historical evidence about its behavior. Adding evaluations on four source applications reduces joint-novelty loss to 0.227 and residual-estimation error from 0.086 to 0.046; eight applications give loss 0.224 (Table 11). The first four recover 92.3% of this loss reduction without router retraining, separating improved evidence about new experts from changes to the routing parameters.

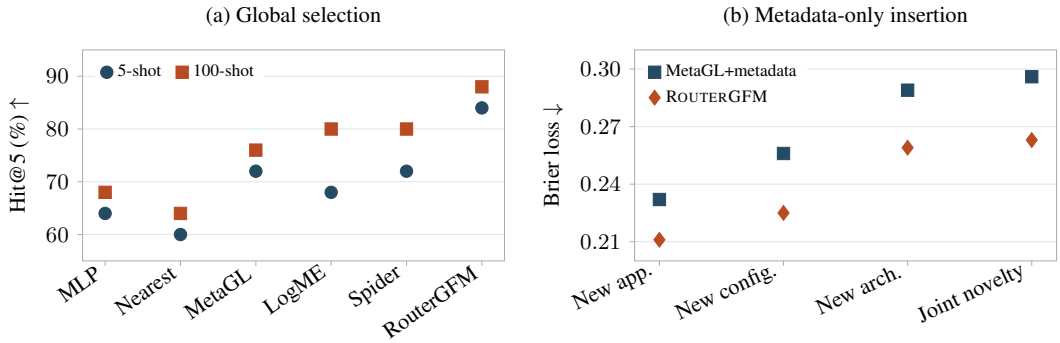

Figure 3: Application-level selection and metadata-only expert insertion. Hit@5 measures best-expert inclusion; insertion uses fixed router parameters without source calibration (Tables 9 and 10).

4.4 Local Integration with Expert Predictions Fixed

Transferred local information improves a fixed set of predictions. Figure 4 compares integration rules in the five-shot classification diagnosis with five expert checkpoints, fitted heads, and query predictions fixed. RouterGFM reduces average Brier loss from 0.239 to 0.211 (11.7%) and worst-cell loss from 0.388 to 0.321 (17.3%) relative to application-level weighting. Average loss also improves over simplex stacking (0.235) and Local-MLP (0.226), with corresponding reductions in worst-cell loss. The fixed predictors exclude expert or head improvements; differing historical supervision across gates supports procedure-level rather than retrieval-only attribution.

Informative historical correspondence makes local weighting useful. Without centering, local transfer reaches Brier loss 0.228, improving on global weighting at 0.239 but remaining above the centered method at 0.211. Shuffling context–evaluation correspondence instead raises loss to 0.244, above the global baseline: query-dependent weights alone are not sufficient. These controls show

that historical local information can transfer, that centering improves its use, and that incorrect correspondence can negate its benefit. Winner agreement rises from 43.6% to 67.3% (Table 4), measuring alignment with local expert preference rather than another classification accuracy.

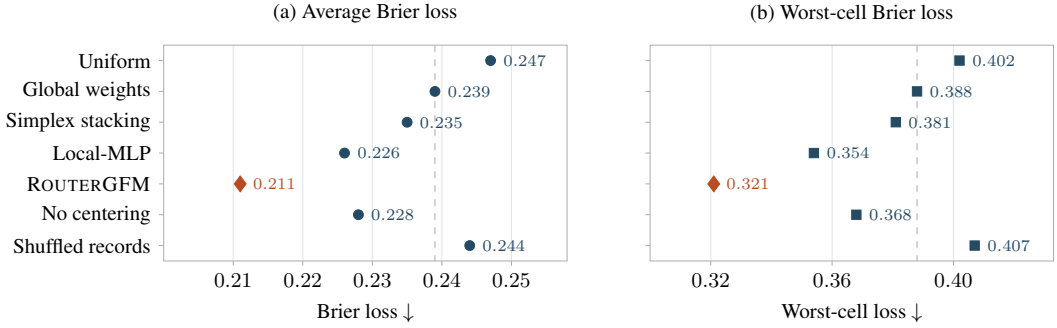

Figure 4: Five-shot classification diagnosis with five checkpoints, fitted heads, and predictions fixed (Table 4). Diamonds denote RouterGFM; dashed lines denote application-level weighting. Worst-cell loss is the largest cell-average mixture loss within an episode (Appendix B.4).

4.5 Expert Coverage and Integration Limits

Additional coverage has diminishing predictive value. Increasing K from one to five reduces Brier loss from 0.269 to 0.211 and raises local-winner coverage from 40.4% to 90.6% (Figure 7). Between $K = 8$ and $K = 16$, coverage increases from 96.1% to 98.8%, while loss changes from 0.209 to 0.212. Coverage can therefore increase without improving the mixture, although these point estimates do not identify a unique cause of the plateau. Winner coverage diagnoses inclusion, not the population loss-weighted omission term Γ_K defined in Theorem 1.

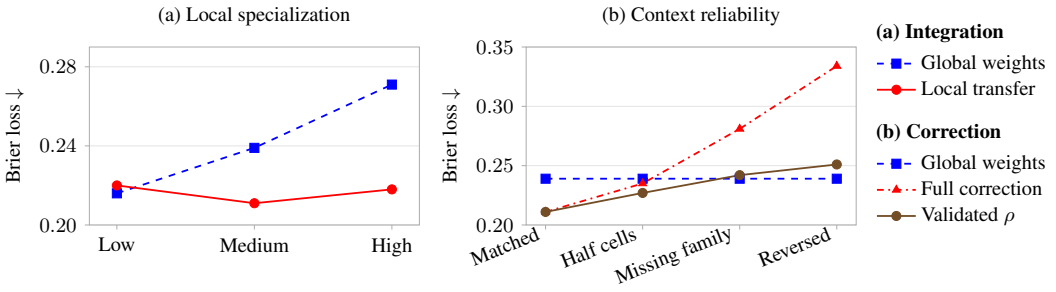

Figure 5: Local transfer under (a) specialization changes and (b) incomplete or incompatible evidence. Source-validated shrinkage mitigates the missing-family and reversal conditions relative to full correction ($\rho = 1$), but does not match global weighting (Tables 13 and 14).

Shrinkage moderates unreliable evidence but does not remove it. Local integration raises loss from 0.216 to 0.220 under low specialization, but lowers it by 0.028 and 0.053 in the medium and high conditions (Figure 5(a)). With half the historical cells removed, source-validated shrinkage gives loss 0.227, compared with 0.235 for full correction ($\rho = 1$) and 0.239 for global weighting. Missing a context family or reversing residual relationships instead gives losses 0.242 and 0.251, both above global weighting, although below the unshrunk losses 0.281 and 0.334 (Figure 5(b)). Incomplete evidence and incompatible behavior impose different limits: shrinkage reduces reliance on history, but cannot restore missing or reversed performance relationships (Theorem 3). Additional comparisons across the evaluated distribution shifts appear in Appendix D.6.

5 Concluding Remarks

RouterGFM reuses historical expert evaluations for inductive selection and context-guided local integration, adapting only the selected task heads. Fixed-team controls support the value of local information beyond application-level weighting, while the theory and diagnostics distinguish between containing useful expertise and reliably estimating its contributions. Neither larger teams nor metadata alone ensure both; incomplete or incompatible history can limit transfer despite shrinkage.

AI USE STATEMENT

Generative AI tools assisted with literature checking, manuscript polishing, coding, and experiment monitoring. The authors are responsible for verifying the final text, mathematical claims, experimental evidence, and bibliography.

ETHICS STATEMENT

Historical expert evaluations may underrepresent particular graph populations or structural contexts, and routing decisions can reproduce these coverage biases. We examine missing-context behavior and worst-cell Brier loss alongside average predictive performance. The results do not establish suitability for high-stakes deployment, and this work does not involve a new human-subject study.

REPRODUCIBILITY STATEMENT

Appendix A provides the theoretical assumptions and proofs, and Appendix B summarizes the algorithm, expert adaptation, and evaluation conventions. Appendices C–E describe the comparisons and report the detailed results. Implementation details and experiment configurations are provided with the accompanying code at `https://anonymous.4open.science/r/RouterGFM-B7B9`.

REFERENCES

Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the 2021 AAAI Conference on Artificial Intelligence (AAAI)*, pp. 3950–3957, 2021.

Aleksandar Bojchevski and Stephan Günnemann. Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking. In *Proceedings of the 2018 International Conference on Learning Representations (ICLR)*, 2018.

Rui Cai, Weijie Jacky Mo, Xiaofei Wen, Qiyao Ma, Wenhui Zhu, Xiwen Chen, Muhao Chen, and Zhe Zhao. ModelLens: Finding the best for your task from myriads of models. *arXiv preprint arXiv:2605.07075*, 2026.

Haifang Cao, Yu Wang, Timing Li, Xinjie Yao, and Pengfei Zhu. Geometric mixture-of-experts with curvature-guided adaptive routing for graph representation learning. *CoRR*, abs/2603.22317, 2026.

Haibo Chen, Xin Wang, Zeyang Zhang, Haoyang Li, Ling Feng, and Wenwu Zhu. AutoGFM: Automated graph foundation model with adaptive architecture customization. In *Proceedings of the 2025 International Conference on Machine Learning (ICML)*, volume 267, pp. 9295–9315. PMLR, 2025.

Yaqi Chen, Shixun Huang, Lei Wang, Ryan Twemlow, John Le, Sheng Wang, Willy Susilo, Jun Yan, and Jun Shen. A cross-graph tuning-free GNN prompting framework. *Neural Networks*, 205:109466, 2027.

Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom M. Mitchell, Kamal Nigam, and Seán Slattery. Learning to extract symbolic knowledge from the world wide web. In *Proceedings of the 1998 AAAI Conference on Artificial Intelligence (AAAI)*, pp. 509–516, 1998.

Rafael M. O. Cruz, Robert Sabourin, George D. C. Cavalcanti, and Tsang Ing Ren. META-DES: A dynamic ensemble selection framework using meta-learning. *Pattern Recognition*, 48(5):1925–1935, 2015.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 4171–4186. ACL, 2019.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research (JMLR)*, 24(43):1–48, 2023.

Taoran Fang, Yunchao Zhang, Yang Yang, Chunping Wang, and Lei Chen. Universal prompt tuning for graph neural networks. In *Proceedings of the 2023 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 52464–52489, 2023.

Tao Feng, Yanzhen Shen, and Jiaxuan You. GraphRouter: A graph-based router for LLM selections. In *Proceedings of the 2025 International Conference on Learning Representations (ICLR)*, 2025.

Matthias Fey and Jan E. Lenssen. Fast graph representation learning with pytorch geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

Xingbo Fu, Yinhan He, and Jundong Li. Edge prompt tuning for graph neural networks. In *Proceedings of the 2025 International Conference on Learning Representations (ICLR)*, 2025.

Zihao Guo, Qingyun Sun, Haonan Yuan, Xingcheng Fu, Min Zhou, Yisen Gao, and Jianxin Li. GraphMoRE: Mitigating topological heterogeneity via mixture of riemannian experts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 11754–11762, 2025.

Haoyu Han, Juanhui Li, Wei Huang, Xianfeng Tang, Hanqing Lu, Chen Luo, Hui Liu, and Jiliang Tang. Node-wise filtering in graph neural networks: A mixture of experts approach. *CoRR*, abs/2406.03464, 2024.

Zhenyu Hou, Haozhan Li, Yukuo Cen, Jie Tang, and Yuxiao Dong. GraphAlign: Pretraining one graph neural network on multiple graphs via feature alignment. *arXiv preprint arXiv:2406.02953*, 2024.

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. In *Proceedings of the 2020 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 22118–22133, 2020a.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay S. Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *Proceedings of the 2020 International Conference on Learning Representations (ICLR)*, 2020b.

Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *Proceedings of the 2020 International Conference on World Wide Web (WWW)*, pp. 2704–2710. ACM, 2020c.

Suyeon Kim, Dongha Lee, SeongKu Kang, Seonghyeon Lee, and Hwanjo Yu. Learning topology-specific experts for molecular property prediction. In *Proceedings of the 2023 AAAI Conference on Artificial Intelligence (AAAI)*, pp. 8291–8299, 2023.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 2015 International Conference on Learning Representations (ICLR)*, 2015.

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

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Hao Liu, Jiarui Feng, Lecheng Kong, Ningyue Liang, Dacheng Tao, Yixin Chen, and Muhan Zhang. One for all: Towards training one graph model for all classification tasks. In *Proceedings of the 2024 International Conference on Learning Representations (ICLR)*, 2024.

Jingzhe Liu, Haitao Mao, Zhikai Chen, Bingheng Li, Wenqi Fan, Mingxuan Ju, Tong Zhao, Neil Shah, and Jiliang Tang. One model for one graph: A new perspective for pretraining with cross-domain graphs. *arXiv preprint arXiv:2412.00315*, 2025a.

Lei Liu, Xingyu Xia, Qianqian Xie, Ben Liu, Wenjie Xu, and Min Peng. Enhanced expert merging for mixture-of-experts in graph foundation models. In *Proceedings of the 2025 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 22173–22207, 2025b.

Zemin Liu, Xingtong Yu, Yuan Fang, and Xinming Zhang. Graphprompt: Unifying pre-training and downstream tasks for graph neural networks. In *Proceedings of the 2023 International Conference on World Wide Web (WWW)*, pp. 417–428. ACM, 2023.

Li Ma, Haoyu Han, Juanhui Li, Harry Shomer, Hui Liu, Xiaofeng Gao, and Jiliang Tang. Mixture of link predictors on graphs. In *Advances in Neural Information Processing Systems*, volume 37, pp. 16043–16070, 2024.

Mohit Meena, Yash Punjabi, Abhishek A, Vishal Sharma, and Mahesh Chandran. Self-adaptive graph mixture of models. In *Proceedings of the 2026 AAAI Conference on Artificial Intelligence (AAAI)*, volume 40, pp. 24344–24352, 2026.

Grégoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New Journal of Physics*, 15(9): 095003, 2013.

Isaac Ong, Amjad Almahairi, Vincent Wu, Wei-Lin Chiang, Tianhao Wu, Joseph E. Gonzalez, M Waleed Kadous, and Ion Stoica. RouteLLM: Learning to route LLMs from preference data. In *Proceedings of the 2025 International Conference on Learning Representations (ICLR)*, 2025.

Namyong Park, Ryan A. Rossi, Nesreen Ahmed, and Christos Faloutsos. MetaGL: Evaluation-free selection of graph learning models via meta-learning. In *Proceedings of the 2023 International Conference on Learning Representations (ICLR)*, 2023a.

Namyong Park, Ryan A. Rossi, Xing Wang, Antoine Simoulin, Nesreen K. Ahmed, and Christos Faloutsos. GLEMOS: Benchmark for instantaneous graph learning model selection. In *Advances in Neural Information Processing Systems*, volume 36, pp. 69887–69899, 2023b.

Sebastian Pineda Arango, Fabio Ferreira, Arlind Kadra, Frank Hutter, and Josif Grabocka. Quicktune: Quickly learning which pretrained model to finetune and how. In *Proceedings of the 2024 International Conference on Learning Representations (ICLR)*, 2024.

Leonardo F. R. Ribeiro, Pedro H. P. Savarese, and Daniel R. Figueiredo. *struc2vec*: Learning node representations from structural identity. In *Proceedings of the 2017 ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 385–394, 2017.

Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2), 2021.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference (ESWC)*, pp. 593–607. Springer, 2018.

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *CoRR*, abs/1811.05868, 2018.

Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. In *Proceedings of the 2020 International Conference on Learning Representations (ICLR)*, 2020.

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 2022 ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 1717–1727. ACM, 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 2023 ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 2120–2131. ACM, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 2017 Annual Conference on Neural Information Processing Systems (NIPS)*, pp. 5998–6008, 2017.

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

Petar Velickovic, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. Deep graph infomax. In *Proceedings of the 2019 International Conference on Learning Representations (ICLR)*, 2019.

Haotao Wang, Ziyu Jiang, Yuning You, Yan Han, Gaowen Liu, Jayanth Srinivasa, Ramana Rao Kompella, and Zhangyang Wang. Graph mixture of experts: Learning on large-scale graphs with explicit diversity modeling. In *Proceedings of the 2023 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 50825–50837, 2023.

Shuo Wang, Bokui Wang, Zhixiang Shen, Boyan Deng, and Zhao Kang. Multi-domain graph foundation models: Robust knowledge transfer via topology alignment. In *Proceedings of the 2025 International Conference on Machine Learning (ICML)*, pp. 64806–64821. PMLR, 2025a.

Yidi Wang, Ziyue Qiao, Jiawei Gu, Xubin Zheng, Pengyang Wang, Xiaobing Pei, and Xiao Luo. Out-of-distribution graph models merging. In *Proceedings of the 2026 International Conference on Learning Representations (ICLR)*, 2026.

Zehong Wang, Zheyuan Zhang, Nitesh V Chawla, Chuxu Zhang, and Yanfang Ye. GFT: Graph foundation model with transferable tree vocabulary. In *Proceedings of the 2024 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 107403–107443, 2024.

Zehong Wang, Zheyuan Zhang, Tianyi Ma, Nitesh V. Chawla, Chuxu Zhang, and Yanfang Ye. Towards graph foundation models: Learning generalities across graphs via task-trees. In *Proceedings of the 2025 International Conference on Machine Learning (ICML)*, volume 267, pp. 65518–65555. PMLR, 2025b.

Zhibin Wang, Zhixing Zhang, Shuqi Wang, Xuanting Xie, and Zhao Kang. GMoPE: A prompt-expert mixture framework for graph foundation models. *CoRR*, abs/2511.03251, 2025c.

Qitian Wu, Wentao Zhao, Zenan Li, David P. Wipf, and Junchi Yan. Nodeformer: A scalable graph structure learning transformer for node classification. In *Proceedings of the 2022 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 27387–27401, 2022.

Shirley Wu, Kaidi Cao, Bruno Ribeiro, James Y. Zou, and Jure Leskovec. GraphMETRO: Mitigating complex graph distribution shifts via mixture of aligned experts. In *Proceedings of the 2024 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 9358–9387, 2024.

Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical Science*, 9(2):513–530, 2018.

Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang, and Hang Li. Listwise approach to learning to rank: theory and algorithm. In *Proceedings of the 2008 International Conference on Machine Learning (ICML)*, pp. 1192–1199. ACM, 2008.

Lianghao Xia and Chao Huang. AnyGraph: Graph foundation model in the wild. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 882–896. Association for Computational Linguistics, 2026.

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

Kaichao You, Yong Liu, Jianmin Wang, and Mingsheng Long. LogME: Practical assessment of pre-trained models for transfer learning. In *Proceedings of the 2021 International Conference on Machine Learning (ICML)*, volume 139, pp. 12133–12143. PMLR, 2021.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. In *Proceedings of the 2020 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 5812–5823, 2020.

Xingtong Yu, Zhenghao Liu, Yuan Fang, Zemin Liu, Sihong Chen, and Xinming Zhang. Generalized graph prompt: Toward a unification of pre-training and downstream tasks on graphs. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 36(11):6237–6250, 2024.

Xingtong Yu, Jie Zhang, Yuan Fang, and Renhe Jiang. Non-homophilic graph pre-training and prompt learning. In *Proceedings of the 2025 ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 1844–1854. ACM, 2025.

Haonan Yuan, Qingyun Sun, Junhua Shi, Xingcheng Fu, Bryan Hooi, Jianxin Li, and Philip S. Yu. How much can transfer? BRIDGE: bounded multi-domain graph foundation model with generalization guarantees. In *Proceedings of the 2025 International Conference on Machine Learning (ICML)*, pp. 73604–73644. PMLR, 2025.

Hanqing Zeng, Hanjia Lyu, Diyi Hu, Yinglong Xia, and Jiebo Luo. Mixture of weak and strong experts on graphs. In *Proceedings of the 2024 International Conference on Learning Representations (ICLR)*, 2024.

Muhan Zhang and Yixin Chen. Link prediction based on graph neural networks. In *Proceedings of the 2018 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 5171–5181, 2018.

Yi-Kai Zhang, Ting-Ji Huang, Yao-Xiang Ding, De-Chuan Zhan, and Han-Jia Ye. Model spider: Learning to rank pre-trained models efficiently. In *Proceedings of the 2023 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 13692–13719, 2023.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. In *Proceedings of the 2020 Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 7793–7804, 2020.

A Complete Theoretical Analysis

This appendix analyzes inductive expert selection and context-guided prediction integration, with the assumptions needed for each guarantee. The statements are conditional on the realized support set, the fitted experts, and the information permitted at deployment.

A.1 Notation and Measurability

Let (X, Y) follow the target application's query distribution, let $Z = z_a(X)$, and let $\mathcal{E}$ be a nonempty finite eligible pool. For each $e \in \mathcal{E}$, $F_e(X)$ is a prediction in a common convex set and $r_e(Z) = \mathbb{E}[\ell(Y, F_e(X)) \mid Z]$ is an integrable conditional risk. For any application a, the risks refer to its support-adapted predictors $F_{a,e} = g_{\omega_{a,e}} \circ f_{\Theta_e}$, and their mean risks and centered performance residuals are

$$\mu_{a,e} = \mathbb{E}[r_{a,e}(Z)], \qquad \delta_{a,e}(z) = \mu_{a,e} - r_{a,e}(z). \tag{14}$$

The sign convention makes $\delta_{a,e}(z)$ positive when expert e has lower loss in context z than its own application average. For the measured local and application-average losses of record i, define $\bar{\delta}_i = \bar{\mu}_i - \bar{r}_i$. With the retrieval weights of Equation 6, let

$$\widehat{\delta}_{a,e}(z) = \sum_{i \in \mathcal{N}_J(z,e)} w_i(z, e)\bar{\delta}_i, \qquad \widehat{r}_{a,e}(z) = \widehat{\mu}_{a,e} - \rho\widehat{\delta}_{a,e}(z). \tag{15}$$

Equation 15 is the residual form of Equation 7; both produce the same local loss estimates and mixture weights. The support set, archive, application metadata, and shortlist are fixed when these conditional expectations are formed. Equivalently, they can be included in the conditioning sigma-field. All comparisons use the same support-adapted experts; changing adaptation algorithms changes the comparison problem. A local gate is measurable with respect to Z and the permitted fixed information, so it cannot inspect the realized query label. This requirement is essential when moving its weights outside a conditional expectation. For squared local-risk supervision, we additionally assume finite second moments of the diagnostic loss.

A.2 The Opportunity for Local Expert Selection

Proposition 2 (Value of local expertise)**.** *For a finite eligible expert pool, define* $R_{\text{fixed}} = \min_e \mathbb{E}[r_e(Z)]$ *and* $R_{\text{loc}} = \mathbb{E}[\min_e r_e(Z)]$. *Then* $\Delta_{\text{loc}} = R_{\text{fixed}} - R_{\text{loc}} \geq 0$*, with equality if and only if a fixed expert minimizes* $r_e(Z)$ *almost surely.*

The difference measures the opportunity for local individual-expert selection, not improvement over the best constant convex mixture.

Proof. Choose $e_0 \in \arg\min_e \mathbb{E}r_e(Z)$, which exists because the pool is finite. For every z, $r_{e_0}(z) - \min_e r_e(z) \geq 0$. Taking expectations proves

$$\Delta_{\text{loc}} = \mathbb{E}\big[r_{e_0}(Z) - \min_e r_e(Z)\big] \geq 0. \tag{16}$$

If one expert e_* minimizes conditional risk almost surely, its mean risk equals $\mathbb{E}\min_e r_e(Z)$, so equality follows. Conversely, if $\Delta_{\text{loc}} = 0$, the nonnegative integrable random variable $r_{e_0}(Z) - \min_e r_e(Z)$ has zero expectation and therefore equals zero almost surely. Thus e_0 minimizes conditional risk almost surely, proving both directions. □

Example with equal global scores. Let two context regions have equal probability and let two experts have risks (a, b) and (b, a) in those regions, where $0 \leq a < b$. Both global risks equal $(a + b)/2$, but $R_{\text{loc}} = a$ and $\Delta_{\text{loc}} = (b - a)/2$. A global ranking cannot reveal which expert to prefer in either region. The construction isolates local specialization while keeping the two global risks equal.

An exact probabilistic example. Figure 6 evaluates an exact two-region probability model. In two equally likely regions, let Y have Bernoulli means 0.8 and 0.2 and let the experts return opposite fixed predictions 0.8 and 0.2. Their local squared probability losses are $(0.16, 0.52)$ and $(0.52, 0.16)$, so both global risks equal 0.34 and the local oracle risk equals 0.16. In this example only, $\sigma(t) =$

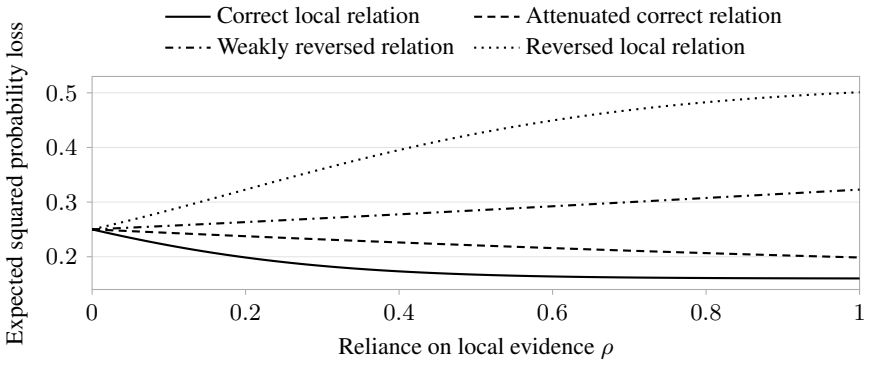

Figure 6: Exact analytical two-region example of residual orientation and shrinkage. With equally likely regions and label probabilities 0.8 and 0.2, two experts exchange predictions 0.8 and 0.2. Increasing ρ helps only when the transferred residual relation has the correct orientation. The global uniform mixture has loss 0.25 and the conditional individual-expert oracle has loss 0.16.

$1/(1+e^{-t})$ denotes the logistic sigmoid. Scale the true local residual by $1-2q$, where q is a declared mismatch parameter, and use temperature 0.1. In the first region the first expert's weight is $\sigma(3.6\rho(1-2q))$, so the exact mixture risk is

$$R(\rho, q) = 0.16 + \left[0.2 + 0.6\sigma(3.6\rho(1-2q)) - 0.8\right]^2. \tag{17}$$

Symmetry gives the same loss in the other region. Thus stronger local weighting reduces risk for $q < 1/2$, increases it for $q > 1/2$, and changes nothing for $q = 1/2$. The closed-form curves separate useful local expertise from the reliability of the transferred residuals.

Why global shortlisting can fail. Let a region have probability $p \in (0,1)$ and consider a generalist with constant risk c and a specialist with risk zero inside that region and one outside it. When $0 < c < 1-p$, the generalist has lower global risk and a size-one shortlist excludes the specialist. Its local omission cost is nevertheless $pc > 0$. Hence accurate global risk prediction alone does not make Γ_K vanish. Additional experts with the same generalist behavior extend the example to any fixed shortlist size. This example motivates measuring shortlist coverage in addition to dataset-level ranking accuracy.

A.3 Proof of Theorem 1

Write $T = T_a$, $\mathcal{E} = \mathcal{E}_a$, and $K = |T|$. The local omission cost and maximum within-team risk-estimation error are

$$\Gamma_K = \mathbb{E}[\min_{e\in T} r_e(Z) - \min_{e\in\mathcal{E}} r_e(Z)], \qquad \epsilon(Z) = \max_{e\in T} |\widehat{r}_e(Z) - r_e(Z)|. \tag{18}$$

For fixed z, write $q_e = \widehat{r}_e(z)$ and let $\alpha_e = \exp(-q_e/\tau)/\sum_{j\in T}\exp(-q_j/\tau)$. For any comparison distribution $p = (p_e)_{e\in T} \in \Delta(T)$, where $p_e \geq 0$ and $\sum_e p_e = 1$, nonnegativity of the Kullback–Leibler divergence $\mathrm{KL}(p\|\alpha) = \sum_e p_e \log(p_e/\alpha_e)$ gives

$$\langle \alpha, q\rangle - \tau H(\alpha) \leq \langle p, q\rangle - \tau H(p), \qquad H(p) = -\sum_{e\in T} p_e \log p_e. \tag{19}$$

This also follows by expanding the KL divergence and canceling the common log-normalizer. Here H is Shannon entropy, and p is a comparison distribution rather than the scalar retrieval bandwidth β. Choose p concentrated on a minimizer of q and use $H(\alpha) \leq \log K$ to obtain

$$\sum_{e\in T} \alpha_e \widehat{r}_e(z) \leq \min_{e\in T} \widehat{r}_e(z) + \tau \log K. \tag{20}$$

By the definition of $\epsilon(z)$, both replacing a weighted estimated risk by its true risk and replacing its minimum by the true minimum cost at most $\epsilon(z)$. Consequently,

$$\sum_{e\in T}\alpha_e r_e(z) \le \sum_{e\in T}\alpha_e \widehat{r}_e(z) + \epsilon(z) \tag{21}$$

$$\le \min_{e\in T}\widehat{r}_e(z) + \tau\log K + \epsilon(z) \tag{22}$$

$$\le \min_{e\in T} r_e(z) + \tau\log K + 2\epsilon(z). \tag{23}$$

Convexity of ℓ in the prediction gives the pointwise inequality

$$\ell\left(Y, \sum_{e\in T}\alpha_e(Z)F_e(X)\right) \le \sum_{e\in T}\alpha_e(Z)\ell(Y, F_e(X)). \tag{24}$$

Conditional expectation given $Z = z$ is legitimate because the weights are measurable with respect to Z. Combining the resulting inequality with Equation 23, integrating over Z, and subtracting R_{loc} proves

$$R(F) - R_{\mathrm{loc}} \le \Gamma_K + 2\mathbb{E}\epsilon(Z) + \tau\log K. \tag{25}$$

If $\mathbb{E}\epsilon$ is infinite, the same statement is valid but uninformative. In particular, $\Gamma_K + 2\mathbb{E}[\epsilon(Z)] + \tau\log K < \Delta_{\mathrm{loc}}$ suffices to outperform the best fixed individual expert. □

Interpretation of the comparator. The conditional single-expert oracle is not an upper limit on the performance of a convex mixture. A mixture can have lower loss than every individual expert through error cancellation. The theorem provides an upper bound on excess risk relative to that oracle, without requiring this excess to be nonnegative. For hard selection $\widehat{e}(z) \in \arg\min_{e\in T}\widehat{r}_e(z)$, the same proof removes the $\tau\log K$ term and does not require convexity of the loss. However, a dataset-level best-expert score remains a valid ceiling only for selecting one fixed expert from exactly that pool on exactly that evaluation split. Neither an oracle from a different checkpoint pool nor a test-selected team is a deployable baseline.

A.4 Reliability of Context-Guided Estimation

Theorem 3 (Context-conditioned estimation and its failure conditions)**.** *Assume that the centered performance residuals of compatible support-adapted predictors satisfy*

$$|\delta_{a,e}(z) - \delta_{b,f}(z')| \le L\|k_\phi(z, v_e) - k_\phi(z', v_f)\|_2 + \eta, \tag{26}$$

Here $L \ge 0$ bounds variation with retrieval-key distance, and $\eta \ge 0$ permits changes in local performance patterns across applications, including those induced by fitting different task heads. For record i, let s_i be its mean within-cell key radius and suppose its residual estimate differs from its population cell mean by at most ε_i. For normalized nonnegative weights, set $d_i = \|k_\phi(z, v_e) - k_\phi(c_i, v_{e_i})\|_2$. Then

$$|\widehat{\delta}_{a,e}(z) - \delta_{a,e}(z)| \le L\sum_i w_i(d_i + s_i) + \eta + \sum_i w_i\varepsilon_i. \tag{27}$$

If $|\widehat{\mu}_{a,e} - \mu_{a,e}| \le \varepsilon_g$ and $|\delta_{a,e}(z)| \le D$, the local-risk error is at most

$$\varepsilon_g + (1-\rho)D + \rho\left[L\sum_i w_i(d_i + s_i) + \eta + \sum_i w_i\varepsilon_i\right]. \tag{28}$$

Proof. Let record i originate from application b_i, expert f_i, and cell C_i. Define its population centered residual by $\delta_i^{\mathrm{cell}} = \mathbb{E}[\delta_{b_i,f_i}(Z') \mid Z' \in C_i]$ and assume $|\bar{\delta}_i - \delta_i^{\mathrm{cell}}| \le \varepsilon_i$. Set

$$s_i = \mathbb{E}[\|k_\phi(Z', v_{f_i}) - k_\phi(c_i, v_{f_i})\|_2 \mid Z' \in C_i]. \tag{29}$$

The representative need not be a Euclidean centroid after the nonlinear key map; the radius above accounts for that mismatch. Apply Equation 26 and the triangle inequality inside this conditional expectation to get

$$|\delta_i^{\mathrm{cell}} - \delta_{a,e}(z)| \le \mathbb{E}[|\delta_{b_i,f_i}(Z') - \delta_{a,e}(z)| \mid Z' \in C_i] \tag{30}$$

$$\le L(d_i + s_i) + \eta. \tag{31}$$

Since $w_i \geq 0$ and $\sum_i w_i = 1$, another triangle inequality yields

$$|\widehat{\delta}_{a,e}(z) - \delta_{a,e}(z)| \leq \sum_i w_i \big(|\bar{\delta}_i - \delta_i^{\text{cell}}| + |\delta_i^{\text{cell}} - \delta_{a,e}(z)|\big) \tag{32}$$

$$\leq \sum_i w_i \varepsilon_i + L \sum_i w_i (d_i + s_i) + \eta. \tag{33}$$

Finally, expand the difference between estimated and true local risks:

$$\widehat{r} - r = (\widehat{\mu} - \mu) + (1 - \rho)\delta + \rho(\delta - \widehat{\delta}). \tag{34}$$

Taking absolute values and applying the stated bounds proves Equation 28. No step treats neighboring graph queries as independent observations. □

Learning correspondence and interpreting the condition. Application-masked training predicts already adapted historical predictors from other applications, providing supervision for transferable context–expert relationships. It does not establish Equation 26 for every unseen application: a newly fitted head can change which contexts favor an expert, even after additive offsets are removed. The error decomposition includes both the application-level estimate and local transfer; shrinkage trades reduced reliance on the latter against bias toward the former. Likewise, $\tau \log K$ is a bound term, not evidence that lowering temperature or enlarging a team must improve observed prediction loss.

Centering and finite diagnostic evidence. Let $\bar{\mu} = \sum_j n_j \bar{r}_j / \sum_j n_j$, using exactly the same valid observations and weights as the cell risks. Then $\sum_j n_j(\bar{\mu} - \bar{r}_j) = 0$ by algebra. When labels are missing, the counts must match the loss's valid-query weighting rather than the raw number of nodes or targets. If the global diagnostic mean and cell risk have errors at most u and v_j, the residual error is at most $u + v_j$. Small cells can therefore have large ε_i even when their centroids are close to a target query. Empirical counts alone do not imply a $1/\sqrt{n_i}$ rate on a dependent graph. Such a rate requires an additional sampling or dependence model and is not asserted here.

A.5 Inductive Insertion: Equivariance and Ambiguity

Proposition 4 (Type-preserving equivariance)**.** *Suppose node feature maps are shared within each type, neighborhood aggregation is permutation invariant, and scoring and retrieval use shared functions with permutation-equivariant tie handling. Relabeling nodes within each type, while relabeling their features, relations, and evidence consistently, relabels all scores and mixture weights in the same way.*

Proof. The initial node states commute with such a relabeling because $P_{\kappa(v)}$ depends on type and features, not identity. Assume equivariance at layer l. Relabeling induces a bijection between the typed neighborhoods of corresponding nodes, preserving each message's inputs. Permutation-invariant aggregation and the shared residual update then give equivariance at layer $l + 1$. Induction proves the claim for the final embeddings. The shared pair head preserves corresponding scores, and metadata-based distance computations preserve corresponding retrieval weights. Top-K selection is equivariant away from boundary ties; ties can be handled by a consistently coupled random order or by equal-weight treatment rather than by arbitrary integer node IDs. The prediction mixture is a sum over corresponding experts, so it is unchanged as a numerical prediction. □

The proposition supports processing additional nodes with existing type-specific parameters. It does not imply that inserting a new node leaves every old embedding or ranking unchanged, because its messages can alter the graph neighborhood. It also does not imply prediction consistency for a new architecture outside the observed metadata distribution.

Proposition 5 (Metadata-only ambiguity)**.** *There exist two newly inserted experts and two target risk assignments that induce identical permitted router inputs but require opposite optimal choices. For any randomized size-one selection rule, the larger expected regret across these assignments is at least $\Delta/2$, where $\Delta > 0$ is their risk gap.*

Proof. Give the two experts identical observable metadata and no diagnostic observations that distinguish them. In the first assignment their risks are $(0, \Delta)$; in the second they are $(\Delta, 0)$. Both

assignments produce the same router input. If the first expert is chosen with probability p, the two expected regrets are $(1-p)\Delta$ and $p\Delta$. Their maximum is at least their average, namely $\Delta/2$. □

This is an information limit on metadata-only selection, not a claim that all real experts have ambiguous descriptions. Source calibration may supply distinguishing evidence, but its value depends on its relevance to the target application.

B Experimental Details

Algorithm 1 RouterGFM: historical router learning and target prediction

Require: Pretrained experts $\mathcal{E}$ and metadata; historical (S_b, D_b); target (S_a, Q_a); team budget K, retrieval budget J, bandwidth β, temperature τ, correction strength ρ, and loss weights λ_r, λ_ℓ.
Ensure: Predictions $F_a(x)$ for $x \in Q_a$.

Historical evaluation

for each historical application b **do**
  Fit evaluated expert heads on S_b with encoders frozen; record losses on disjoint D_b.
  Compute application averages by Equation 2; group label-free contexts and compute local averages by Equation 5.
end for
Build the context graph $\mathcal{H}$ and local evaluation archive $\mathcal{M}$.

Router learning

for each training episode with application b **do**
  Mask b's evaluation edges and their reverses, local records, and label-derived aggregates.
  Encode the masked graph into h_v and predict application losses $\widehat{\mu}_{b,e}$ for $e \in \Omega_b$.
  Retrieve evaluations from other applications and estimate local losses using Equation 7.
  Update (θ, ϕ) by Equation 12, keeping historical predictions and losses fixed.
end for
Select the router checkpoint and hyperparameters on separate historical validation applications.

Target prediction: router parameters fixed

Insert target and new-expert metadata with construction edges; filter $\mathcal{E}$ to $\mathcal{E}_a$.
Select T_a by Equation 4, without executing candidate encoders on the target.
Fit selected task heads once on S_a with pretrained GNNs frozen.
for each query $x \in Q_a$ **do**
  Compute context $z_a(x)$ and expert predictions $F_{a,e}(x)$ for $e \in T_a$.
  Retrieve $\mathcal{N}_J(z_a(x), e)$ and compute record weights w_i by Equation 6, or use zero correction when this set is empty.
  Compute local loss estimates and prediction weights using Equations 7 and 8.
  Return the weighted prediction $F_a(x)$ from Equation 1.
end for

B.1 Experts and Adaptation

The pool contains $7 \times 7 \times 12 = 588$ pretrained checkpoints before task-compatibility filtering. The architectures are GCN, GAT, GIN, H2GCN, FAGCN, NodeFormer, and Transformer (Kipf & Welling, 2017; Veličković et al., 2018; Xu et al., 2019; Zhu et al., 2020; Bo et al., 2021; Wu et al., 2022; Vaswani et al., 2017). Pretraining uses attribute masking, context prediction, DGI, edge prediction, GraphCL, InfoGraph, and supervised objectives (Hu et al., 2020b; Velickovic et al., 2019; You et al., 2020; Sun et al., 2020). The twelve source datasets are Cora, PubMed, Computers, Flickr, Actor, Squirrel, Email, BBBP, Tox21, QM9, PROTEINS, and CIFAR10. Each expert retains its checkpoint-specific preprocessing and task readout; matched-pool comparisons use the same eligible inventory and support observations. Node tasks read the marked node, link tasks use an enclosing-subgraph and symmetric endpoint readout, and graph tasks use pooled representations. Selected encoders remain frozen while task heads are fitted on target support examples, with output coordinates aligned before mixing. The implementation uses PyTorch Geometric (Fey & Lenssen, 2019), frozen BERT description embeddings (Devlin et al., 2019), and Adam optimization (Kingma

& Ba, 2015). Detailed model, preprocessing, and optimizer settings are provided with the accompanying code.

B.2 Task Losses and Local Evidence

For single-label classification, the routing loss is $\ell(y, p) = \frac{1}{2}\|p - \mathrm{onehot}(y)\|_2^2$; binary link prediction uses squared probability error. Multi-label predictions retain separate assay probabilities, and unobserved targets are masked in task losses. For regression, targets are normalized using support medians and median absolute deviations with a 10^{-6} scale floor; predictions are transformed back to raw units for MAE reporting. Archive records retain local and application-average losses, valid evaluation counts, context representatives, and expert identities. Retrieval matches task/loss families, support budgets, and normalization conventions, and caps records from any one source application. Counts determine diagnostic averages but do not multiply the distance-based weights in Equation 6. Source-trained descriptor standardization is reused at deployment, and an empty compatible retrieval set gives zero local correction.

For predicted loss u and measured loss v, $\mathrm{Huber}(u, v)$ equals $\frac{1}{2}(u - v)^2$ when $|u - v| \le c_{\mathrm{H}}$ and $c_{\mathrm{H}}(|u - v| - \frac{1}{2}c_{\mathrm{H}})$ otherwise, where $c_{\mathrm{H}} > 0$ is the transition threshold.

B.3 Training and Evaluation

Historical support sets fit expert heads, while disjoint diagnostic observations supervise the router. Application-masked training removes the pseudo-target's evaluation edges, reverse edges, local records, and label-derived aggregates before its losses are predicted. Router selection uses source validation, and tasks, budgets, and splits of one base dataset remain grouped. Target support labels fit the selected heads; query labels are used only for evaluation. Single-label classification uses examples per class, regression uses support graphs in total, and ToxCast uses the multitask partitions specified in the code. Prediction results use seeds 42, 0, 100, 123, and 2024; mechanism tables report point estimates rather than significance tests.

The routed LP comparisons use the 10%–5%–10% positive-edge convention, with remaining positives unused and held-out positives excluded from message passing. LP blocks A and B in Table 1 retain the two reported result sets under this convention; their labels identify displayed comparisons, not different link budgets, and repeated baselines are shared results. The configuration catalog uses 5%–10%–10% at five shots and 10%–5%–10% at 100 shots; ProNoG additionally uses remaining positives as message-only context. These LP groups are interpreted within their respective information conditions. Within each configuration, validation selects the checkpoint; the adaptation catalog then reports the best observed test mean across completed configurations with that configuration's standard deviation. Dataset loaders, support partitions, and complete configurations are documented in the code rather than repeated here.

B.4 Diagnostic Measures and Scope

The fixed-team classification diagnosis uses five-shot support and holds the selected checkpoints, fitted heads, query predictions, and evaluation cells fixed across integration rules. Within an evaluation episode, worst-cell Brier loss is the largest cell-average loss of the mixed predictions under that shared partition, not a worst-case guarantee over unobserved contexts. Local-winner coverage measures the mass of cells whose best eligible expert is in the shortlist, while winner agreement compares the highest-weight expert with a cellwise minimizing expert. These quantities answer different questions: whether useful experts are retained, whether they receive high weights, and whether their mixed predictions improve loss. Neither coverage nor agreement is a direct estimate of the population omission cost or risk-estimation error in Theorem 1.

Selection scores metadata pairs without executing every candidate GNN on target data; prediction still uses the selected experts. For N queries, K experts, R compatible archive records, and key dimension d_k, exhaustive retrieval requires $O(NKRd_k)$ distance work and $O(Rd_k)$ cached-key storage. These counts do not establish a measured latency or memory advantage. Local transfer depends on compatible historical evidence, and metadata alone need not distinguish experts with different behavior.

C BASELINE COMPARISONS

Selection and integration. The metadata MLP, nearest-application rule, MetaGL, LogME, and Model Spider compare application-level expert recommendation (Park et al., 2023a; You et al., 2021; Zhang et al., 2023). MetaGL+metadata denotes the metadata extension used for expert insertion. MetaGL-U averages its selected predictions, while ROUTERGFM-G uses ROUTERGFM's team and fitted heads with application-level weights. The fixed-team study compares uniform weighting, global weighting, simplex stacking, Local-MLP, and historical local transfer on the same expert predictions. The comparison isolates the complete integration procedure from changes in predictors; it does not separately match the historical supervision of every learned integration rule. The no-centering and shuffled-record rows assess sensitivity to changes in the archived evidence.

Graph mixtures and adaptation. SAGMM-PE uses the matched frozen-expert pool, whereas AnyGraph, GMoE, Mowst, GraphMoRE, and GMoPE retain their method-specific expert organizations (Meena et al., 2026; Xia & Huang, 2026; Wang et al., 2023; Zeng et al., 2024; Guo et al., 2025; Wang et al., 2025c). Node-MoE is evaluated on node tasks, Link-MoE on link tasks, and META-DES on classification; KDEM/PPEM use compatible architectures for expert merging (Han et al., 2024; Ma et al., 2024; Cruz et al., 2015; Liu et al., 2025b). Task-aware adaptations outside a method's original setting follow the implementations in the code. GraphMETRO, OGMM, and GeoMoE supply shift-oriented comparisons with their own expert inventories (Wu et al., 2024; Wang et al., 2026; Cao et al., 2026). The configuration catalog includes full fine-tuning and nine prompting variants, with GraphPrompt+ denoting the generalized extension of GraphPrompt (Yu et al., 2024). These catalog and method-specific results complement the matched-pool comparison rather than define equal-compute or identical-adaptation comparisons.

D ADDITIONAL RESULTS AND ANALYSES

D.1 APPLICATION-LEVEL EFFECTS OF LOCAL INTEGRATION

Table 2 gives the differences between ROUTERGFM and ROUTERGFM-G in Table 1, retaining each task's reporting metric. The mean accuracy improvement increases from 4.25 to 8.41 points across the five single-label classification datasets, while ToxCast's ROC-AUC improvement decreases from 2.69 to 1.65 points. For QM7b, the MAE reductions are 0.21 and 0.49, corresponding to 3.9% and 10.0% relative improvements. The gains therefore vary across applications and metrics rather than defining one universal supervision-response curve. The LP differences summarize the displayed comparisons without attributing their change to different link-label budgets.

Table 2: Application-specific improvement from local integration, computed from the displayed ROUTERGFM and ROUTERGFM-G means in Table 1. Accuracy and ROC-AUC gains are percentage points; regression gains are reductions in raw-unit MAE. For LP only, the last two columns correspond to blocks A and B rather than shot budgets; the reported results remain separate. The accuracy mean excludes ToxCast, LP, and regression.

Application	Improvement measure	5-shot / A	100-shot / B
Photo	Accuracy increase	1.26	5.18
ogbn-arxiv	Accuracy increase	9.45	6.81
Airports	Accuracy increase	5.54	12.95
Chameleon	Accuracy increase	3.10	15.09
MNIST	Accuracy increase	1.91	2.02
Accuracy mean	Five single-label datasets	4.25	8.41
ToxCast	ROC-AUC increase	2.69	1.65
QM7b	MAE reduction	0.21	0.49
DBLP	ROC-AUC increase by reported block	7.77	8.73
Cornell	ROC-AUC increase by reported block	3.17	3.48

Table 3: 100-shot performance summary on nine datasets. Each adaptation entry is the best observed test mean over the architecture–objective grid, with the standard deviation of the same configuration. Accuracy and ROC-AUC are percentages; QM7b uses MAE. Bold and underline identify the best and second distinct means; LP conventions are specified in Appendix B.3. LP rankings are computed separately for the configuration and routed comparison groups. Both ROUTERGFM and ROUTERGFM-G fine-tune only task heads.

Models	Homophilic graphs			Heterophilic graphs			Graph-level tasks		
	Photo	**ogbn-arxiv**	**DBLP**	**Airports**	**Chameleon**	**Cornell**	**QM7b**	**ToxCast**	**MNIST**
	NC / Acc. ↑	NC / Acc. ↑	LP / AUC ↑	NC / Acc. ↑	NC / Acc. ↑	LP / AUC ↑	GR / MAE ↓	GC / AUC ↑	GC / Acc. ↑
Supervised graph models									
GCN	$87.37_{\pm 0.78}$	$37.72_{\pm 0.68}$	$55.60_{\pm 0.65}$	$39.82_{\pm 3.13}$	$34.27_{\pm 1.32}$	$62.78_{\pm 5.43}$	$5.08_{\pm 0.08}$	$52.49_{\pm 1.21}$	$24.37_{\pm 2.61}$
GAT	$86.41_{\pm 0.57}$	$37.84_{\pm 0.53}$	$56.01_{\pm 0.59}$	$35.57_{\pm 1.87}$	$35.26_{\pm 1.35}$	$69.72_{\pm 9.48}$	$21.51_{\pm 0.59}$	$53.17_{\pm 1.03}$	$27.96_{\pm 0.88}$
GIN	$87.05_{\pm 0.63}$	$31.82_{\pm 0.91}$	$55.24_{\pm 1.26}$	$40.68_{\pm 1.53}$	$37.59_{\pm 0.38}$	$69.74_{\pm 8.13}$	$36.71_{\pm 2.31}$	$51.45_{\pm 1.86}$	$38.59_{\pm 1.96}$
H2GCN	$87.01_{\pm 0.79}$	$38.89_{\pm 1.00}$	$55.47_{\pm 0.73}$	$27.14_{\pm 0.84}$	$41.26_{\pm 1.68}$	$73.42_{\pm 6.76}$	$5.64_{\pm 0.16}$	$52.14_{\pm 1.25}$	$33.61_{\pm 1.15}$
FAGCN	$87.21_{\pm 0.68}$	$42.00_{\pm 0.64}$	$54.62_{\pm 1.33}$	$36.51_{\pm 2.99}$	$37.33_{\pm 1.12}$	$62.81_{\pm 13.20}$	$5.47_{\pm 0.10}$	$52.05_{\pm 0.82}$	$23.41_{\pm 2.00}$
Transformer	$88.44_{\pm 0.36}$	$39.24_{\pm 0.68}$	$57.03_{\pm 0.42}$	$41.09_{\pm 2.15}$	$39.64_{\pm 0.96}$	$\mathbf{76.79}_{\pm 7.10}$	$10.71_{\pm 1.28}$	$52.74_{\pm 0.54}$	$37.32_{\pm 1.53}$
NodeFormer	$83.94_{\pm 0.84}$	$40.47_{\pm 2.33}$	$53.29_{\pm 1.73}$	$37.42_{\pm 3.58}$	$39.40_{\pm 1.28}$	$62.68_{\pm 3.99}$	$13.71_{\pm 0.06}$	$51.34_{\pm 0.74}$	$14.54_{\pm 4.72}$
Best observed configurations by adaptation rule									
Full fine-tuning	$88.99_{\pm 0.57}$	$42.45_{\pm 1.23}$	$\underline{58.65}_{\pm 2.34}$	$42.66_{\pm 2.37}$	$42.30_{\pm 1.70}$	$69.37_{\pm 3.03}$	$5.26_{\pm 0.07}$	$53.56_{\pm 1.56}$	$42.46_{\pm 0.31}$
All-in-One	$88.80_{\pm 0.41}$	$45.06_{\pm 0.14}$	$54.83_{\pm 0.32}$	$43.34_{\pm 0.86}$	$39.95_{\pm 0.43}$	$\underline{74.13}_{\pm 5.46}$	$11.78_{\pm 0.57}$	$60.96_{\pm 1.64}$	$37.50_{\pm 0.55}$
EdgePrompt	$83.10_{\pm 1.76}$	$46.89_{\pm 0.59}$	$\mathbf{58.84}_{\pm 3.95}$	$42.56_{\pm 1.66}$	$42.10_{\pm 0.96}$	$66.56_{\pm 12.92}$	$9.22_{\pm 0.33}$	$53.98_{\pm 0.63}$	$33.69_{\pm 3.79}$
EdgePrompt+	$85.07_{\pm 1.50}$	$44.96_{\pm 1.34}$	$57.27_{\pm 2.83}$	$47.67_{\pm 1.34}$	$42.15_{\pm 1.08}$	$72.01_{\pm 4.86}$	$9.26_{\pm 0.32}$	$54.54_{\pm 3.25}$	$39.31_{\pm 2.50}$
GPF	$89.35_{\pm 0.46}$	$46.68_{\pm 0.80}$	$58.22_{\pm 1.30}$	$41.95_{\pm 1.55}$	$41.36_{\pm 1.30}$	$72.60_{\pm 3.42}$	$33.24_{\pm 1.56}$	$53.79_{\pm 2.33}$	$34.22_{\pm 4.13}$
GPF+	$89.23_{\pm 0.78}$	$43.02_{\pm 0.27}$	$55.63_{\pm 0.73}$	$42.86_{\pm 1.64}$	$41.01_{\pm 1.30}$	$69.06_{\pm 10.62}$	$32.62_{\pm 0.72}$	$53.72_{\pm 0.91}$	$35.53_{\pm 1.48}$
GPPT	$88.21_{\pm 0.67}$	$36.06_{\pm 0.60}$	$53.84_{\pm 3.53}$	$40.53_{\pm 1.70}$	$40.25_{\pm 0.64}$	$66.83_{\pm 9.07}$	$43.58_{\pm 9.14}$	$50.82_{\pm 0.62}$	$31.34_{\pm 1.48}$
GraphPrompt	$87.81_{\pm 0.51}$	$24.46_{\pm 0.65}$	$\underline{58.65}_{\pm 3.19}$	$40.51_{\pm 1.98}$	$32.76_{\pm 4.18}$	$72.98_{\pm 6.05}$	$9.73_{\pm 0.32}$	$47.17_{\pm 1.26}$	$38.30_{\pm 0.52}$
GraphPrompt+	$87.78_{\pm 0.95}$	$22.58_{\pm 0.48}$	$57.49_{\pm 2.09}$	$41.47_{\pm 1.55}$	$32.13_{\pm 4.58}$	$57.29_{\pm 14.44}$	$8.10_{\pm 4.09}$	$46.52_{\pm 1.64}$	$38.11_{\pm 0.40}$
ProNoG	$87.40_{\pm 0.47}$	$41.35_{\pm 0.68}$	$56.44_{\pm 0.53}$	$37.97_{\pm 1.24}$	$37.56_{\pm 1.46}$	$65.97_{\pm 10.09}$	$\mathbf{4.06}_{\pm 0.07}$	$45.69_{\pm 1.34}$	$31.31_{\pm 0.75}$
Graph mixture-of-experts models									
AnyGraph	$84.00_{\pm 1.90}$	$50.24_{\pm 1.76}$	$\underline{83.81}_{\pm 1.78}$	$40.40_{\pm 2.89}$	$36.63_{\pm 2.21}$	$57.20_{\pm 8.08}$	$11.92_{\pm 0.73}$	$53.15_{\pm 0.65}$	$33.23_{\pm 1.60}$
GMoE	$85.76_{\pm 2.14}$	$25.68_{\pm 0.57}$	$54.83_{\pm 0.77}$	$38.18_{\pm 3.77}$	$33.09_{\pm 1.69}$	$58.28_{\pm 10.40}$	$113.74_{\pm 2.41}$	$55.51_{\pm 2.71}$	$34.65_{\pm 3.08}$
Mowst	$87.19_{\pm 0.77}$	$38.97_{\pm 0.64}$	$55.32_{\pm 0.46}$	$42.68_{\pm 1.89}$	$35.24_{\pm 0.93}$	$64.21_{\pm 9.67}$	$103.16_{\pm 1.53}$	$52.23_{\pm 2.52}$	$28.32_{\pm 6.74}$
GraphMoRE	$85.56_{\pm 0.72}$	$33.85_{\pm 0.92}$	$54.10_{\pm 0.40}$	$38.94_{\pm 1.90}$	$35.06_{\pm 0.70}$	$57.19_{\pm 9.03}$	$22.93_{\pm 39.40}$	$52.78_{\pm 0.47}$	$29.49_{\pm 0.44}$
GMoPE	$89.66_{\pm 0.57}$	$50.26_{\pm 1.03}$	$81.22_{\pm 1.62}$	$48.11_{\pm 1.91}$	$\underline{44.52}_{\pm 1.73}$	$75.91_{\pm 6.10}$	$5.04_{\pm 0.16}$	$59.85_{\pm 0.56}$	$\underline{44.63}_{\pm 1.52}$
Matched-pool routing and frozen-expert mixtures									
SAGMM-PE	$89.43_{\pm 0.47}$	$49.72_{\pm 0.80}$	$82.80_{\pm 1.23}$	$48.54_{\pm 1.44}$	$44.35_{\pm 1.17}$	$\underline{76.80}_{\pm 5.20}$	$4.99_{\pm 0.11}$	$60.10_{\pm 1.46}$	$43.62_{\pm 1.13}$
MetaGL-U	$89.03_{\pm 0.45}$	$48.06_{\pm 0.84}$	$80.43_{\pm 1.40}$	$47.48_{\pm 1.76}$	$42.33_{\pm 1.16}$	$73.36_{\pm 4.79}$	$5.14_{\pm 0.12}$	$59.84_{\pm 1.50}$	$42.51_{\pm 1.05}$
ROUTERGFM-G	$\underline{89.71}_{\pm 0.40}$	$\underline{50.79}_{\pm 0.81}$	$83.31_{\pm 1.26}$	$\underline{49.23}_{\pm 1.53}$	$44.05_{\pm 1.45}$	$76.33_{\pm 4.65}$	$4.91_{\pm 0.12}$	$\underline{60.99}_{\pm 1.22}$	$44.16_{\pm 1.26}$
ROUTERGFM	$\mathbf{94.89}_{\pm 0.34}$	$\mathbf{57.60}_{\pm 0.60}$	$\mathbf{92.04}_{\pm 0.75}$	$\mathbf{62.18}_{\pm 1.20}$	$\mathbf{59.14}_{\pm 1.90}$	$\mathbf{79.81}_{\pm 4.55}$	$\underline{4.42}_{\pm 0.43}$	$\mathbf{62.64}_{\pm 1.13}$	$\mathbf{46.18}_{\pm 1.12}$

D.2 Fixed-Team Prediction Integration

Table 4 gives the measurements in Figure 4 and adds local-winner agreement. With predictors fixed, uniform weighting, global weighting, and simplex stacking give average losses 0.247, 0.239, and 0.235, whereas local transfer reaches 0.211. Its advantage over Local-MLP is 0.015 in average loss and 0.033 in worst-cell loss. Removing centering or shuffling correspondence increases loss to 0.228 or 0.244, respectively, and lowers winner agreement from 67.3% to 55.1% or 42.7%. These observations support transferring structured local evidence rather than treating all historical evaluations as equally useful. The controls alter different aspects of the procedure, so their gaps are not additive component contributions.

Table 4: Five-shot classification diagnosis with identical checkpoints, adapted heads, and expert predictions. Brier losses are point estimates; worst-cell loss is defined within an evaluation episode, and winner agreement is not classification accuracy. The comparison does not alone establish matched historical supervision for learned integration rules.

Integration on the same team	Brier loss ↓	Worst-cell loss ↓	Winner agreement (%) ↑
Uniform	0.247	0.402	39.2
Global risk weights	0.239	0.388	43.6
Simplex stacking	0.235	0.381	45.6
Local-MLP	0.226	0.354	58.8
Local residual transfer	0.211	0.321	67.3
No residual centering	0.228	0.368	55.1
Shuffled context–residual pairs	0.244	0.407	42.7

D.3 Task-Scoped Mixtures and Expert Merging

Tables 5–8 compare methods within their evaluated task scopes. On Chameleon, RouterGFM exceeds Node-MoE by 1.14 and 13.17 accuracy points at five and 100 shots, respectively. On DBLP, its AUC exceeds Link-MoE by 6.32 and 7.28 points in the displayed blocks; Cornell's corresponding margins are 1.08 and 1.39 points. On MNIST, RouterGFM exceeds GMoPE by 1.62 and 1.55 accuracy points. These comparisons retain method-specific expert inventories; the fixed-team study provides the more direct test of prediction integration.

Table 5: Node-level 5-shot accuracy comparisons (percent, higher is better). KDEM/PPEM use compatible expert architectures; META-DES uses the matched graph-context representation. Best/second-best means in each column are bold/underlined.

Method	Photo	Arxiv	Airports	Chameleon
Node-MoE	$80.32_{\pm 0.93}$	$29.33_{\pm 0.93}$	$40.22_{\pm 2.40}$	$\underline{38.39}_{\pm 2.26}$
META-DES	$81.46_{\pm 1.00}$	$32.20_{\pm 0.93}$	$41.10_{\pm 2.88}$	$37.12_{\pm 2.56}$
SAGMM-PE	$\underline{82.10}_{\pm 0.92}$	$33.15_{\pm 0.88}$	$41.35_{\pm 2.51}$	$37.05_{\pm 2.74}$
KDEM	$81.82_{\pm 1.04}$	$\underline{33.32}_{\pm 1.01}$	$\underline{41.37}_{\pm 2.86}$	$36.65_{\pm 2.66}$
PPEM	$81.37_{\pm 1.00}$	$32.84_{\pm 0.87}$	$40.84_{\pm 2.87}$	$36.18_{\pm 2.58}$
RouterGFM	$\mathbf{83.74}_{\pm 0.89}$	$\mathbf{42.92}_{\pm 1.55}$	$\mathbf{47.09}_{\pm 1.87}$	$\mathbf{39.53}_{\pm 2.68}$

D.4 Selection, Insertion, and Source Calibration

Table 9 reports both best-expert inclusion and shortlist regret. At five shots, Model Spider and MetaGL have the same 72% hit@5 but different regrets (0.029 and 0.032); RouterGFM improves both criteria to 84% and 0.019. At 100 shots, RouterGFM retains the strongest values, 88% and 0.013. These diagnostics concern the globally best individual expert, not coverage of every local specialist.

Table 10 separates new applications, new configurations, unseen architectures, and joint novelty while keeping router parameters fixed. Relative to MetaGL+metadata, RouterGFM reduces risk by 0.021, 0.031, 0.030, and 0.033, respectively. Nevertheless, its risk is 0.034 higher for unseen

Table 6: Node-level 100-shot accuracy comparisons (percent, higher is better). KDEM/PPEM use compatible expert architectures; META-DES uses the matched graph-context representation. Best/second-best means in each column are bold/underlined.

Method	Photo	Arxiv	Airports	Chameleon
Node-MoE	$88.82_{\pm 0.43}$	$44.94_{\pm 0.94}$	$47.80_{\pm 1.42}$	$\underline{45.97}_{\pm 1.44}$
META-DES	$\underline{89.50}_{\pm 0.40}$	$49.70_{\pm 0.87}$	$\underline{49.18}_{\pm 1.60}$	$44.35_{\pm 1.24}$
SAGMM-PE	$89.43_{\pm 0.47}$	$49.72_{\pm 0.80}$	$48.54_{\pm 1.44}$	$44.35_{\pm 1.17}$
KDEM	$89.03_{\pm 0.41}$	$\underline{50.71}_{\pm 0.94}$	$47.75_{\pm 1.67}$	$43.88_{\pm 1.44}$
PPEM	$88.74_{\pm 0.49}$	$50.13_{\pm 0.78}$	$47.24_{\pm 1.73}$	$43.40_{\pm 1.21}$
ROUTERGFM	$\mathbf{94.89}_{\pm 0.34}$	$\mathbf{57.60}_{\pm 0.60}$	$\mathbf{62.18}_{\pm 1.20}$	$\mathbf{59.14}_{\pm 1.90}$

Table 7: Link-prediction ROC-AUC for blocks A and B in Table 1 (percent, higher is better); block labels do not denote link-supervision budgets. The two blocks share the edge-split convention in Appendix B.3; shared baseline results are repeated for comparison and do not represent additional runs. Task-specific predictors retain their declared eligible inventories; bold and underline identify the best and second-best means within each block.

Method	LP block A		LP block B	
	DBLP	Cornell	DBLP	Cornell
Link-MoE	$\underline{84.76}_{\pm 1.35}$	$\underline{78.42}_{\pm 5.75}$	$\underline{84.76}_{\pm 1.35}$	$\underline{78.42}_{\pm 5.75}$
SAGMM-PE	$82.80_{\pm 1.23}$	$76.80_{\pm 5.20}$	$82.80_{\pm 1.23}$	$76.80_{\pm 5.20}$
KDEM	$84.15_{\pm 1.47}$	$75.82_{\pm 5.62}$	$84.15_{\pm 1.47}$	$75.82_{\pm 5.62}$
PPEM	$83.96_{\pm 1.34}$	$75.28_{\pm 5.17}$	$83.96_{\pm 1.34}$	$75.28_{\pm 5.17}$
ROUTERGFM	$\mathbf{91.08}_{\pm 0.45}$	$\mathbf{79.50}_{\pm 4.48}$	$\mathbf{92.04}_{\pm 0.75}$	$\mathbf{79.81}_{\pm 4.55}$

Table 8: Graph-classification accuracy on MNIST at both support budgets (percent, higher is better). Best/second-best means in each column are bold/underlined.

Method	5-shot	100-shot
META-DES	$30.28_{\pm 1.80}$	$43.77_{\pm 1.13}$
GMoPE	$\underline{30.66}_{\pm 2.31}$	$\underline{44.63}_{\pm 1.52}$
SAGMM-PE	$30.11_{\pm 1.91}$	$43.62_{\pm 1.13}$
ROUTERGFM	$\mathbf{32.28}_{\pm 1.84}$	$\mathbf{46.18}_{\pm 1.12}$

Table 9: Global expert-selection diagnostics with $K = 5$ on node and single-label graph classification. Hit@5 is a percentage (higher is better), and regret@5 uses Brier loss (lower is better). All methods use the same eligible inventory, while target-execution costs of transferability estimators are not included in these selection-quality scores.

Method	5-shot		100-shot	
	Hit@5	Regret@5	Hit@5	Regret@5
Metadata MLP	64.0	0.045	68.0	0.036
Nearest application	60.0	0.053	64.0	0.044
MetaGL	72.0	0.032	76.0	0.025
LogME	68.0	0.036	80.0	0.023
Model Spider	72.0	0.029	80.0	0.022
ROUTERGFM	84.0	0.019	88.0	0.013

architectures than for new configurations built from observed factors. Source calibration reduces joint-novelty loss by 0.036 with four applications and by a further 0.003 with eight (Table 11). Residual-estimation error falls from 0.086 without calibration to 0.046 with four applications and 0.041 with eight. The paired changes in residual error and prediction loss support the usefulness of source evidence for the adapted expert, while the router parameters remain unchanged.

Table 10: Inductive insertion diagnosis using Brier loss (lower is better). MetaGL+metadata is an explicitly extended baseline. Architecture-family holdouts group checkpoints differing only by seed.

Frozen-router setting	MetaGL+metadata	RouterGFM
New application; known experts	0.232	0.211
New expert; seen metadata factors	0.256	0.225
New architecture family	0.289	0.259
New application and expert	0.296	0.263

Table 11: Calibration of a newly inserted expert on separate source applications. Router parameters remain fixed; no target query labels are used. Calibration runs count toward the experiment budget.

Source calibration applications	Risk ↓	Mean residual error ↓
0	0.263	0.086
1	0.247	0.070
2	0.235	0.057
4	0.227	0.046
8	0.224	0.041

D.5 Coverage, Specialization, and Evidence Reliability

Table 12 follows nested teams under one global ranking. Moving from one to five experts lowers loss by 0.058 and raises winner coverage by 50.2 points. Beyond eight experts, coverage continues to increase while the observed prediction loss remains near its minimum. The loss-reduction column is a difference of mixture losses, not an estimate of the omission term Γ_K. Five experts recover 96.7% of the observed one-to-eight loss reduction with 37.5% fewer expert contributions. This compares expert contribution counts, not measured wall-clock latency.

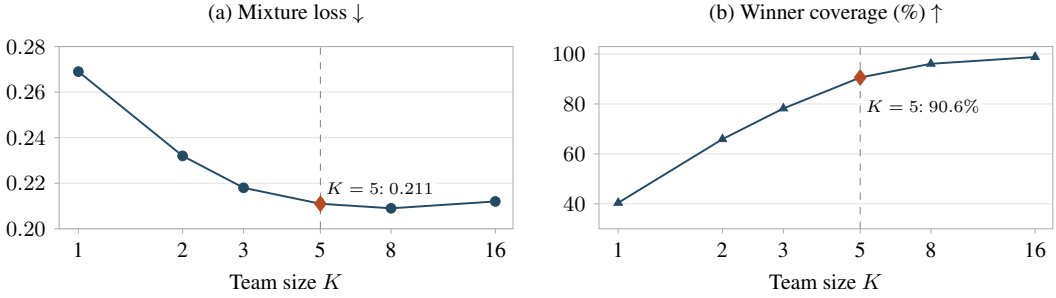

Figure 7: Expert coverage and prediction quality are distinct. Coverage increases with team size while prediction loss reaches a plateau; the marked $K = 5$ team captures most of the observed loss reduction (Table 12).

With half the archived cells removed, source-validated shrinkage retains an advantage over global weighting, but removing a context family or reversing residual relationships gives higher loss (Table 13). Shrinkage reduces the corresponding excess losses relative to full correction without eliminating them. This agrees with the role of residual shift in Theorem 3: a nearby context need not carry transferable expert behavior.

Table 14 and Figure 5 compare global and local weighting under the reported specialization conditions. Local transfer increases loss by 0.004 in the low condition but decreases it by 0.028 and 0.053 in the medium and high conditions. The conditions can differ in both expert behavior and estimation

Table 12: Shortlist size, prediction risk, and local-winner coverage. Loss reduction is the absolute difference from the singleton mixture, calculated within this table. Expert contributions count predictors in the mixture, not wall-clock executions.

Team size K	Brier loss ↓	Loss reduction	Winner coverage (%) ↑	Expert contributions
1	0.269	0.000	40.4	1
2	0.232	0.037	65.9	2
3	0.218	0.051	78.2	3
5	0.211	0.058	90.6	5
8	0.209	0.060	96.1	8
16	0.212	0.057	98.8	16

Table 13: Coverage and failure analysis using Brier loss. Source-selected shrinkage may still fail under an unanticipated target shift; it is not an oracle that always recovers the global baseline.

Available context evidence	Global only	Local, $\rho = 1$	Source-validated ρ
Matched coverage	0.239	0.211	0.211
Half of context cells missing	0.239	0.235	0.227
Entire context family missing	0.239	0.281	0.242
Residual relation reversed	0.239	0.334	0.251

difficulty, so these contrasts are descriptive rather than causal effects of one graph property. The analytical example in Appendix A.2 separately isolates the direction of transferred residuals.

Table 14: Prediction risks within the reported specialization conditions at a fixed support budget. The final column is local minus global risk, not an oracle gap or a paired uncertainty estimate.

Specialization condition	Global only	Local transfer	Risk difference
Low	0.216	0.220	$+0.004$
Medium	0.239	0.211	-0.028
High	0.271	0.218	-0.053

D.6 Distribution-Shift Comparisons

Table 15 shows complementary strengths across shift conditions. RouterGFM gives the lowest feature-shift risk (0.232), whereas GeoMoE is better under structural shift (0.231 versus 0.247) and mixed shift (0.263 versus 0.269). Thus historical local integration is not uniformly preferable to a shift-oriented model. The comparison uses each method's own inventory and adaptation procedure and does not establish that inserting a standalone shift specialist into the pool would reproduce its performance.

Table 15: Distribution-shift comparisons with task-appropriate graph experts, reporting Brier loss. Methods retain task-appropriate expert inventories and training budgets; lower is better.

Shift condition	GraphMETRO	OGMM	GeoMoE	RouterGFM
Feature shift	0.245	<u>0.237</u>	0.238	**0.232**
Structural shift	<u>0.239</u>	0.251	**0.231**	0.247
Mixed unseen shift	0.277	<u>0.266</u>	**0.263**	0.269

E CONFIGURATION-LEVEL RESULTS

Tables 16 and 3 summarize seven architectures and ten adaptation strategies on nine datasets; the following tables retain their configuration-level results. Each adaptation summary takes the best observed test mean over the completed architecture–objective grid and retains that configuration's standard deviation. Figure 1 ranks the best configurations containing each factor, so its cells summarize attainable choices rather than isolated factor effects. Photo's leading architecture changes from FAGCN (81.27%) to GAT (89.35%) across budgets, while Chameleon's changes from Transformer (35.39%) to H2GCN (42.30%). The task dependence also appears in graph prediction: ProNoG obtains QM7b MAE 4.06 ± 0.07 with Transformer/InfoGraph at 100 shots, whereas All-in-One gives the best ToxCast adaptation mean, 60.96 ± 1.64 ROC-AUC, with tied GCN/EdgePred and GCN/InfoGraph configurations. These results motivate selecting among different pretrained choices, while the matched-pool and fixed-team experiments assess routing and integration more directly.

Table 16: Five-shot single-model configuration summary on nine datasets. Each adaptation entry is the best observed test mean over the architecture–objective grid, with the standard deviation of the same configuration. Accuracy and ROC-AUC are percentages; QM7b uses MAE. Bold and underline identify the best and second distinct means; LP conventions are specified in Appendix B.3.

Models	Homophily			Heterophily			Graph		
	Photo	**ogbn-arxiv**	**DBLP**	**Airports**	**Chameleon**	**Cornell**	**QM7b**	**ToxCast**	**MNIST**
	NC / Acc. ↑	NC / Acc. ↑	LP / AUC ↑	NC / Acc. ↑	NC / Acc. ↑	LP / AUC ↑	GR / MAE ↓	GC / AUC ↑	GC / Acc. ↑
Supervised graph models									
GCN	$77.08_{\pm 1.52}$	$22.86_{\pm 1.72}$	$54.94_{\pm 0.45}$	$32.51_{\pm 3.89}$	$28.79_{\pm 4.71}$	$58.29_{\pm 8.87}$	**5.59**$_{\pm 0.28}$	$48.76_{\pm 0.98}$	$21.08_{\pm 0.72}$
GAT	$72.23_{\pm 2.53}$	$19.79_{\pm 1.41}$	$55.01_{\pm 1.13}$	$28.80_{\pm 3.49}$	$27.74_{\pm 3.74}$	$63.72_{\pm 4.61}$	$21.56_{\pm 1.26}$	$48.58_{\pm 1.53}$	$22.77_{\pm 2.40}$
GIN	$71.84_{\pm 1.52}$	$17.86_{\pm 1.31}$	$55.91_{\pm 0.82}$	$35.18_{\pm 3.08}$	$29.01_{\pm 3.20}$	$61.91_{\pm 4.48}$	$41.65_{\pm 5.39}$	$45.97_{\pm 2.36}$	$26.79_{\pm 2.20}$
H2GCN	$76.77_{\pm 1.94}$	$20.25_{\pm 1.34}$	$55.33_{\pm 0.78}$	$25.50_{\pm 3.47}$	$29.80_{\pm 3.80}$	$70.13_{\pm 5.92}$	$6.14_{\pm 0.17}$	$48.83_{\pm 1.31}$	$23.83_{\pm 2.38}$
FAGCN	$75.89_{\pm 1.79}$	$23.60_{\pm 0.84}$	$53.81_{\pm 0.86}$	$28.02_{\pm 2.26}$	$31.36_{\pm 5.31}$	$62.37_{\pm 7.41}$	$6.16_{\pm 0.43}$	$48.22_{\pm 0.80}$	$21.89_{\pm 1.35}$
Transformer	$76.19_{\pm 3.34}$	$22.89_{\pm 1.68}$	$56.95_{\pm 0.52}$	$33.98_{\pm 3.06}$	$28.41_{\pm 3.29}$	$\underline{72.42}_{\pm 8.03}$	$24.64_{\pm 37.88}$	$48.73_{\pm 0.70}$	$22.79_{\pm 2.93}$
NodeFormer	$79.04_{\pm 2.30}$	$4.43_{\pm 1.98}$	$54.20_{\pm 4.08}$	$33.47_{\pm 3.84}$	$28.06_{\pm 6.79}$	$58.75_{\pm 9.62}$	$52.70_{\pm 47.67}$	$50.51_{\pm 0.75}$	$14.64_{\pm 3.46}$
Best observed configurations by adaptation rule									
Full fine-tuning	$80.32_{\pm 1.36}$	$29.02_{\pm 1.19}$	**59.91**$_{\pm 3.28}$	$34.63_{\pm 2.56}$	$32.67_{\pm 3.82}$	$63.58_{\pm 9.74}$	$\underline{5.60}_{\pm 0.69}$	$50.88_{\pm 1.32}$	$\underline{28.22}_{\pm 1.25}$
All-in-One	$79.45_{\pm 1.29}$	$30.31_{\pm 0.26}$	$54.40_{\pm 0.54}$	$37.64_{\pm 4.45}$	$33.61_{\pm 1.78}$	**74.63**$_{\pm 14.46}$	$35.77_{\pm 1.65}$	**57.51**$_{\pm 3.09}$	$26.89_{\pm 1.30}$
EdgePrompt	$71.53_{\pm 5.19}$	**31.38**$_{\pm 0.88}$	$\underline{57.38}_{\pm 1.50}$	$38.17_{\pm 6.37}$	$34.71_{\pm 3.01}$	$61.08_{\pm 11.77}$	$36.00_{\pm 1.98}$	$52.39_{\pm 1.93}$	$24.48_{\pm 2.13}$
EdgePrompt+	$70.42_{\pm 3.71}$	$29.51_{\pm 1.33}$	$57.16_{\pm 3.77}$	**38.80**$_{\pm 5.58}$	**35.39**$_{\pm 2.10}$	$61.81_{\pm 16.74}$	$32.62_{\pm 1.56}$	$52.46_{\pm 2.22}$	**28.54**$_{\pm 2.58}$
GPF	$\underline{80.88}_{\pm 4.46}$	$\underline{31.20}_{\pm 0.91}$	$56.91_{\pm 0.93}$	$\underline{38.67}_{\pm 4.85}$	$\underline{34.95}_{\pm 2.84}$	$65.54_{\pm 8.29}$	$33.95_{\pm 2.13}$	$\underline{52.60}_{\pm 1.47}$	$26.67_{\pm 2.78}$
GPF+	**81.27**$_{\pm 3.42}$	$26.40_{\pm 1.26}$	$56.28_{\pm 0.49}$	$37.62_{\pm 4.03}$	$34.40_{\pm 2.46}$	$65.32_{\pm 10.05}$	$34.99_{\pm 1.92}$	$52.35_{\pm 1.69}$	$25.53_{\pm 5.97}$
GPPT	$78.46_{\pm 1.94}$	$28.39_{\pm 4.30}$	$55.35_{\pm 2.88}$	$35.03_{\pm 2.05}$	$33.72_{\pm 2.27}$	$60.29_{\pm 6.19}$	$41.86_{\pm 1.66}$	$51.09_{\pm 1.67}$	$22.35_{\pm 2.61}$
GraphPrompt	$77.56_{\pm 1.63}$	$14.67_{\pm 2.23}$	$54.83_{\pm 0.55}$	$35.40_{\pm 4.13}$	$27.83_{\pm 6.74}$	$66.76_{\pm 7.43}$	$42.99_{\pm 2.25}$	$47.53_{\pm 1.92}$	$28.16_{\pm 2.70}$
GraphPrompt+	$77.22_{\pm 2.58}$	$13.59_{\pm 1.88}$	$55.20_{\pm 0.69}$	$35.61_{\pm 4.09}$	$26.15_{\pm 6.21}$	$61.90_{\pm 8.10}$	$25.02_{\pm 4.18}$	$44.65_{\pm 2.66}$	$24.61_{\pm 1.79}$
ProNoG	$80.56_{\pm 3.53}$	$27.10_{\pm 3.22}$	$57.03_{\pm 0.70}$	$32.03_{\pm 3.68}$	$30.57_{\pm 3.07}$	$70.00_{\pm 5.18}$	$8.40_{\pm 3.46}$	$48.06_{\pm 0.75}$	$27.28_{\pm 0.44}$

Table 17: 5-shot configuration results for GCN across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	76.29$_{\pm 1.77}$	18.33$_{\pm 8.05}$	54.08$_{\pm 0.32}$	32.70$_{\pm 3.78}$	28.80$_{\pm 4.97}$	48.80$_{\pm 7.63}$	69.86$_{\pm 58.95}$	48.84$_{\pm 0.86}$	22.72$_{\pm 1.93}$
All-in-One	75.48$_{\pm 2.56}$	10.60$_{\pm 2.23}$	52.22$_{\pm 0.75}$	**37.56**$_{\pm 2.58}$	28.04$_{\pm 5.05}$	53.25$_{\pm 4.92}$	106.36$_{\pm 0.49}$	**55.14**$_{\pm 2.79}$	12.29$_{\pm 3.90}$
EdgePrompt	64.04$_{\pm 4.16}$	24.04$_{\pm 3.29}$	53.99$_{\pm 0.53}$	34.12$_{\pm 4.93}$	**32.38**$_{\pm 3.84}$	**56.37**$_{\pm 8.07}$	109.57$_{\pm 0.17}$	50.89$_{\pm 1.49}$	19.54$_{\pm 6.46}$
EdgePrompt+	66.05$_{\pm 5.86}$	20.00$_{\pm 2.91}$	53.75$_{\pm 0.56}$	34.70$_{\pm 6.09}$	31.88$_{\pm 3.00}$	53.56$_{\pm 9.97}$	109.25$_{\pm 0.20}$	51.68$_{\pm 0.76}$	24.42$_{\pm 1.50}$
GPF	76.70$_{\pm 2.18}$	23.36$_{\pm 2.41}$	54.11$_{\pm 0.53}$	31.08$_{\pm 5.22}$	31.53$_{\pm 1.05}$	49.52$_{\pm 10.32}$	110.98$_{\pm 0.26}$	51.94$_{\pm 0.80}$	20.42$_{\pm 4.38}$
GPF+	**77.15**$_{\pm 2.55}$	20.41$_{\pm 1.81}$	53.68$_{\pm 0.49}$	32.50$_{\pm 4.48}$	31.30$_{\pm 2.12}$	49.32$_{\pm 8.56}$	110.93$_{\pm 0.28}$	50.31$_{\pm 2.41}$	23.03$_{\pm 1.96}$
GPPT	72.48$_{\pm 3.39}$	17.51$_{\pm 5.27}$	50.86$_{\pm 1.52}$	31.64$_{\pm 4.36}$	27.50$_{\pm 4.80}$	52.22$_{\pm 8.24}$	108.52$_{\pm 0.77}$	49.90$_{\pm 0.87}$	18.43$_{\pm 1.52}$
GraphPrompt	68.39$_{\pm 3.65}$	11.08$_{\pm 2.18}$	53.42$_{\pm 0.42}$	31.81$_{\pm 1.43}$	25.00$_{\pm 5.49}$	52.51$_{\pm 4.12}$	99.99$_{\pm 0.20}$	43.07$_{\pm 2.99}$	**24.64**$_{\pm 0.74}$
GraphPrompt+	45.13$_{\pm 8.54}$	12.63$_{\pm 2.55}$	**54.35**$_{\pm 0.43}$	35.21$_{\pm 3.95}$	22.14$_{\pm 2.37}$	54.76$_{\pm 9.26}$	83.94$_{\pm 8.25}$	42.83$_{\pm 2.77}$	20.04$_{\pm 1.49}$
ProNoG	72.48$_{\pm 4.64}$	**24.66**$_{\pm 3.36}$	54.12$_{\pm 1.01}$	28.36$_{\pm 2.87}$	26.79$_{\pm 2.43}$	52.50$_{\pm 5.90}$	**10.13**$_{\pm 4.09}$	43.23$_{\pm 2.32}$	23.38$_{\pm 1.05}$
Pretraining: ContextPred									
Full fine-tuning	76.95$_{\pm 4.53}$	19.71$_{\pm 10.06}$	**53.87**$_{\pm 0.30}$	30.91$_{\pm 5.89}$	29.33$_{\pm 4.53}$	44.88$_{\pm 6.25}$	**5.78**$_{\pm 0.61}$	49.65$_{\pm 1.42}$	22.53$_{\pm 2.85}$
All-in-One	72.52$_{\pm 1.87}$	6.62$_{\pm 1.03}$	52.49$_{\pm 0.27}$	32.55$_{\pm 2.42}$	29.86$_{\pm 4.80}$	53.77$_{\pm 10.17}$	109.54$_{\pm 0.70}$	50.08$_{\pm 0.56}$	18.29$_{\pm 4.34}$
EdgePrompt	68.26$_{\pm 7.93}$	29.24$_{\pm 2.51}$	53.27$_{\pm 0.43}$	34.43$_{\pm 4.80}$	31.16$_{\pm 2.51}$	49.87$_{\pm 10.24}$	110.51$_{\pm 0.10}$	50.51$_{\pm 0.68}$	21.66$_{\pm 0.83}$
EdgePrompt+	67.70$_{\pm 5.29}$	25.95$_{\pm 1.70}$	52.90$_{\pm 0.45}$	**35.95**$_{\pm 6.14}$	**31.92**$_{\pm 2.94}$	52.98$_{\pm 11.53}$	110.27$_{\pm 0.17}$	**50.67**$_{\pm 1.65}$	25.37$_{\pm 1.21}$
GPF	**78.37**$_{\pm 3.80}$	**29.75**$_{\pm 2.29}$	52.80$_{\pm 0.66}$	34.12$_{\pm 4.71}$	31.58$_{\pm 2.61}$	44.44$_{\pm 8.82}$	111.08$_{\pm 0.15}$	50.11$_{\pm 0.98}$	20.94$_{\pm 3.57}$
GPF+	78.31$_{\pm 2.82}$	26.40$_{\pm 1.26}$	52.71$_{\pm 0.64}$	34.87$_{\pm 3.87}$	31.49$_{\pm 2.13}$	45.33$_{\pm 9.70}$	111.02$_{\pm 0.12}$	50.48$_{\pm 1.38}$	24.54$_{\pm 1.77}$
GPPT	76.88$_{\pm 2.48}$	16.81$_{\pm 7.21}$	50.67$_{\pm 1.29}$	30.29$_{\pm 4.69}$	30.13$_{\pm 1.57}$	52.71$_{\pm 9.85}$	109.06$_{\pm 0.79}$	50.36$_{\pm 0.87}$	20.12$_{\pm 0.63}$
GraphPrompt	74.80$_{\pm 4.84}$	10.51$_{\pm 1.28}$	52.72$_{\pm 0.59}$	29.42$_{\pm 4.90}$	23.09$_{\pm 4.43}$	56.58$_{\pm 9.25}$	101.18$_{\pm 0.50}$	47.18$_{\pm 3.69}$	25.64$_{\pm 1.73}$
GraphPrompt+	64.50$_{\pm 8.04}$	8.33$_{\pm 3.20}$	53.00$_{\pm 0.71}$	33.28$_{\pm 2.52}$	22.41$_{\pm 3.57}$	**59.26**$_{\pm 10.70}$	87.03$_{\pm 8.01}$	42.97$_{\pm 3.00}$	19.62$_{\pm 3.50}$
ProNoG	76.78$_{\pm 1.88}$	22.72$_{\pm 3.23}$	53.29$_{\pm 0.91}$	29.98$_{\pm 4.45}$	27.53$_{\pm 2.75}$	55.84$_{\pm 8.02}$	12.16$_{\pm 1.99}$	45.57$_{\pm 1.87}$	**26.01**$_{\pm 1.25}$
Pretraining: DGI									
Full fine-tuning	77.37$_{\pm 2.31}$	21.77$_{\pm 1.58}$	**54.03**$_{\pm 0.16}$	31.79$_{\pm 4.06}$	29.53$_{\pm 5.36}$	49.63$_{\pm 11.54}$	**5.78**$_{\pm 0.50}$	46.52$_{\pm 2.48}$	19.93$_{\pm 5.94}$
All-in-One	74.79$_{\pm 1.49}$	17.58$_{\pm 2.20}$	52.47$_{\pm 0.60}$	**33.68**$_{\pm 5.48}$	30.21$_{\pm 4.22}$	53.42$_{\pm 4.30}$	104.46$_{\pm 1.05}$	46.69$_{\pm 1.00}$	16.50$_{\pm 4.01}$
EdgePrompt	68.90$_{\pm 2.33}$	22.87$_{\pm 3.45}$	53.67$_{\pm 0.79}$	32.29$_{\pm 5.15}$	**34.14**$_{\pm 3.98}$	53.17$_{\pm 9.72}$	108.74$_{\pm 0.38}$	49.07$_{\pm 0.36}$	20.05$_{\pm 5.04}$
EdgePrompt+	66.76$_{\pm 5.24}$	22.18$_{\pm 1.65}$	53.34$_{\pm 0.36}$	33.42$_{\pm 4.82}$	33.37$_{\pm 3.34}$	51.93$_{\pm 6.26}$	108.44$_{\pm 0.13}$	49.33$_{\pm 0.71}$	**24.37**$_{\pm 2.94}$
GPF	**77.84**$_{\pm 1.94}$	22.71$_{\pm 1.00}$	53.06$_{\pm 0.36}$	26.75$_{\pm 2.42}$	32.00$_{\pm 3.58}$	50.83$_{\pm 12.43}$	110.30$_{\pm 1.18}$	48.60$_{\pm 0.73}$	20.76$_{\pm 4.66}$
GPF+	77.04$_{\pm 2.08}$	21.67$_{\pm 1.75}$	53.07$_{\pm 0.35}$	28.79$_{\pm 4.09}$	33.29$_{\pm 3.64}$	53.17$_{\pm 8.78}$	109.29$_{\pm 0.10}$	48.71$_{\pm 0.55}$	23.98$_{\pm 1.89}$
GPPT	44.63$_{\pm 4.05}$	17.01$_{\pm 4.40}$	49.60$_{\pm 0.59}$	26.34$_{\pm 6.34}$	21.37$_{\pm 4.67}$	53.23$_{\pm 12.62}$	104.90$_{\pm 2.04}$	**50.32**$_{\pm 1.57}$	11.27$_{\pm 1.68}$
GraphPrompt	74.47$_{\pm 3.97}$	9.96$_{\pm 2.27}$	52.85$_{\pm 1.18}$	31.38$_{\pm 1.49}$	23.99$_{\pm 4.54}$	55.92$_{\pm 10.73}$	87.45$_{\pm 0.76}$	42.85$_{\pm 3.06}$	24.23$_{\pm 1.98}$
GraphPrompt+	70.06$_{\pm 9.35}$	10.39$_{\pm 2.27}$	53.15$_{\pm 0.69}$	30.82$_{\pm 2.30}$	25.30$_{\pm 6.51}$	**61.90**$_{\pm 8.10}$	63.87$_{\pm 9.30}$	43.07$_{\pm 2.83}$	21.13$_{\pm 0.50}$
ProNoG	74.13$_{\pm 5.42}$	**24.55**$_{\pm 3.32}$	52.33$_{\pm 0.61}$	27.06$_{\pm 3.29}$	28.69$_{\pm 2.84}$	53.34$_{\pm 3.67}$	17.14$_{\pm 6.97}$	45.35$_{\pm 1.40}$	23.17$_{\pm 0.85}$
Pretraining: EdgePred									
Full fine-tuning	**76.49**$_{\pm 1.88}$	15.55$_{\pm 9.12}$	**53.79**$_{\pm 0.28}$	30.19$_{\pm 4.24}$	29.41$_{\pm 5.13}$	43.88$_{\pm 5.64}$	**5.60**$_{\pm 0.69}$	46.66$_{\pm 2.30}$	19.93$_{\pm 5.94}$
All-in-One	24.39$_{\pm 2.10}$	3.06$_{\pm 1.32}$	50.90$_{\pm 1.31}$	34.50$_{\pm 3.67}$	29.80$_{\pm 4.51}$	**55.46**$_{\pm 7.28}$	114.39$_{\pm 0.75}$	46.67$_{\pm 2.19}$	13.97$_{\pm 2.99}$
EdgePrompt	45.40$_{\pm 10.29}$	10.80$_{\pm 2.40}$	53.29$_{\pm 0.18}$	**35.15**$_{\pm 3.98}$	24.31$_{\pm 4.56}$	49.66$_{\pm 7.58}$	110.51$_{\pm 0.10}$	**50.93**$_{\pm 1.20}$	19.66$_{\pm 4.50}$
EdgePrompt+	63.60$_{\pm 5.16}$	**23.01**$_{\pm 1.67}$	53.46$_{\pm 0.52}$	35.06$_{\pm 7.36}$	30.33$_{\pm 6.92}$	50.64$_{\pm 10.10}$	110.27$_{\pm 0.17}$	50.48$_{\pm 1.19}$	24.27$_{\pm 1.00}$
GPF	38.96$_{\pm 5.94}$	9.03$_{\pm 2.04}$	51.33$_{\pm 0.92}$	30.87$_{\pm 6.17}$	30.01$_{\pm 1.66}$	41.76$_{\pm 9.06}$	116.65$_{\pm 0.45}$	50.74$_{\pm 1.03}$	13.18$_{\pm 7.45}$
GPF+	63.61$_{\pm 5.21}$	12.52$_{\pm 2.47}$	53.34$_{\pm 0.25}$	29.74$_{\pm 4.87}$	**31.17**$_{\pm 1.89}$	41.00$_{\pm 8.14}$	111.02$_{\pm 0.12}$	49.20$_{\pm 1.65}$	22.35$_{\pm 1.67}$
GPPT	16.72$_{\pm 3.99}$	3.28$_{\pm 1.24}$	51.71$_{\pm 1.97}$	33.93$_{\pm 3.29}$	26.71$_{\pm 2.54}$	54.44$_{\pm 7.88}$	109.06$_{\pm 0.79}$	50.08$_{\pm 0.22}$	20.80$_{\pm 0.83}$
GraphPrompt	23.01$_{\pm 1.45}$	4.02$_{\pm 0.66}$	51.17$_{\pm 0.56}$	30.17$_{\pm 2.65}$	24.19$_{\pm 3.24}$	53.35$_{\pm 8.36}$	113.85$_{\pm 1.87}$	45.72$_{\pm 1.79}$	23.57$_{\pm 1.01}$
GraphPrompt+	45.89$_{\pm 9.69}$	2.74$_{\pm 0.92}$	53.22$_{\pm 0.62}$	32.21$_{\pm 2.46}$	25.13$_{\pm 5.83}$	55.23$_{\pm 8.87}$	87.03$_{\pm 8.01}$	42.84$_{\pm 2.71}$	19.16$_{\pm 1.09}$
ProNoG	20.78$_{\pm 3.10}$	5.09$_{\pm 0.55}$	52.76$_{\pm 0.30}$	32.03$_{\pm 3.68}$	27.14$_{\pm 2.07}$	51.33$_{\pm 8.06}$	12.16$_{\pm 1.99}$	42.49$_{\pm 2.71}$	**25.82**$_{\pm 0.90}$
Pretraining: GraphCL									
Full fine-tuning	77.31$_{\pm 3.64}$	12.17$_{\pm 8.16}$	**54.35**$_{\pm 0.07}$	32.32$_{\pm 3.37}$	28.84$_{\pm 4.94}$	43.85$_{\pm 5.70}$	34.35$_{\pm 13.23}$	46.43$_{\pm 1.29}$	21.22$_{\pm 5.40}$
All-in-One	71.01$_{\pm 3.11}$	6.54$_{\pm 2.18}$	52.37$_{\pm 0.25}$	34.94$_{\pm 3.67}$	28.13$_{\pm 4.35}$	51.36$_{\pm 5.57}$	114.31$_{\pm 0.68}$	**54.89**$_{\pm 2.61}$	20.03$_{\pm 2.76}$
EdgePrompt	68.43$_{\pm 6.51}$	22.10$_{\pm 1.67}$	53.31$_{\pm 0.63}$	34.79$_{\pm 4.61}$	30.81$_{\pm 2.31}$	46.99$_{\pm 5.81}$	110.34$_{\pm 0.27}$	47.87$_{\pm 1.82}$	21.55$_{\pm 1.56}$
EdgePrompt+	69.36$_{\pm 5.13}$	23.01$_{\pm 1.67}$	53.11$_{\pm 1.00}$	**35.20**$_{\pm 4.70}$	30.58$_{\pm 2.56}$	47.69$_{\pm 14.84}$	110.00$_{\pm 0.14}$	47.83$_{\pm 2.01}$	**25.16**$_{\pm 1.88}$
GPF	77.81$_{\pm 1.38}$	**23.22**$_{\pm 1.42}$	53.22$_{\pm 0.50}$	26.68$_{\pm 3.30}$	31.51$_{\pm 2.08}$	43.89$_{\pm 9.18}$	116.14$_{\pm 0.13}$	48.26$_{\pm 2.52}$	22.68$_{\pm 1.16}$
GPF+	**77.92**$_{\pm 1.33}$	22.32$_{\pm 1.53}$	53.50$_{\pm 0.37}$	27.57$_{\pm 4.50}$	**31.77**$_{\pm 2.64}$	44.41$_{\pm 7.81}$	116.07$_{\pm 0.13}$	48.11$_{\pm 2.30}$	24.28$_{\pm 1.36}$
GPPT	66.71$_{\pm 1.56}$	12.27$_{\pm 4.16}$	50.22$_{\pm 0.27}$	27.83$_{\pm 3.46}$	26.23$_{\pm 2.35}$	52.83$_{\pm 5.75}$	115.34$_{\pm 0.13}$	49.16$_{\pm 0.95}$	19.10$_{\pm 1.83}$
GraphPrompt	69.52$_{\pm 3.67}$	10.76$_{\pm 1.27}$	53.32$_{\pm 0.48}$	29.85$_{\pm 2.76}$	22.32$_{\pm 3.47}$	49.83$_{\pm 5.28}$	113.67$_{\pm 0.12}$	44.76$_{\pm 2.37}$	24.24$_{\pm 1.96}$
GraphPrompt+	74.01$_{\pm 4.90}$	11.00$_{\pm 1.44}$	53.76$_{\pm 0.61}$	32.24$_{\pm 1.34}$	24.59$_{\pm 6.20}$	**59.00**$_{\pm 10.48}$	97.09$_{\pm 12.95}$	42.63$_{\pm 3.21}$	19.38$_{\pm 2.37}$
ProNoG	67.47$_{\pm 1.59}$	18.54$_{\pm 1.36}$	53.74$_{\pm 0.82}$	28.85$_{\pm 2.59}$	26.85$_{\pm 3.00}$	53.95$_{\pm 8.12}$	**9.68**$_{\pm 2.42}$	45.93$_{\pm 1.66}$	24.26$_{\pm 1.22}$
Pretraining: InfoGraph									
Full fine-tuning	**77.58**$_{\pm 1.74}$	**21.94**$_{\pm 2.16}$	**53.81**$_{\pm 0.40}$	32.26$_{\pm 3.46}$	28.74$_{\pm 5.10}$	46.37$_{\pm 9.52}$	**5.67**$_{\pm 0.67}$	48.87$_{\pm 0.85}$	19.93$_{\pm 5.94}$
All-in-One	46.06$_{\pm 2.64}$	3.06$_{\pm 1.32}$	51.81$_{\pm 0.99}$	34.50$_{\pm 3.67}$	28.46$_{\pm 4.82}$	56.03$_{\pm 7.48}$	114.39$_{\pm 0.75}$	**57.51**$_{\pm 3.09}$	13.97$_{\pm 2.99}$
EdgePrompt	61.78$_{\pm 5.70}$	11.40$_{\pm 2.20}$	52.34$_{\pm 0.99}$	**35.15**$_{\pm 3.98}$	**31.40**$_{\pm 4.85}$	51.07$_{\pm 9.04}$	112.15$_{\pm 0.06}$	50.93$_{\pm 1.20}$	19.66$_{\pm 4.50}$
EdgePrompt+	64.39$_{\pm 7.89}$	14.63$_{\pm 0.98}$	53.04$_{\pm 0.37}$	34.15$_{\pm 5.76}$	30.87$_{\pm 3.67}$	50.02$_{\pm 11.48}$	111.88$_{\pm 0.09}$	50.48$_{\pm 1.19}$	**24.27**$_{\pm 1.00}$
GPF	70.38$_{\pm 4.68}$	10.97$_{\pm 1.53}$	52.52$_{\pm 0.39}$	27.98$_{\pm 4.51}$	28.32$_{\pm 1.57}$	46.75$_{\pm 9.86}$	116.30$_{\pm 0.05}$	50.78$_{\pm 1.08}$	20.23$_{\pm 2.85}$
GPF+	77.51$_{\pm 3.02}$	12.17$_{\pm 1.22}$	53.25$_{\pm 0.30}$	27.56$_{\pm 3.92}$	30.99$_{\pm 1.47}$	46.28$_{\pm 9.74}$	116.29$_{\pm 0.06}$	50.51$_{\pm 0.95}$	23.88$_{\pm 0.92}$
GPPT	48.07$_{\pm 3.59}$	4.35$_{\pm 0.51}$	50.57$_{\pm 1.29}$	29.11$_{\pm 2.75}$	29.79$_{\pm 3.05}$	52.29$_{\pm 7.87}$	114.89$_{\pm 0.39}$	50.08$_{\pm 0.22}$	20.80$_{\pm 0.83}$
GraphPrompt	44.42$_{\pm 2.94}$	5.32$_{\pm 0.64}$	52.40$_{\pm 0.72}$	30.09$_{\pm 2.17}$	25.09$_{\pm 5.51}$	53.28$_{\pm 7.46}$	112.52$_{\pm 0.06}$	45.68$_{\pm 1.85}$	23.25$_{\pm 0.91}$
GraphPrompt+	63.79$_{\pm 6.28}$	3.44$_{\pm 1.34}$	52.51$_{\pm 0.62}$	32.00$_{\pm 2.15}$	23.62$_{\pm 3.58}$	**58.26**$_{\pm 10.18}$	96.34$_{\pm 12.56}$	42.57$_{\pm 2.68}$	20.94$_{\pm 3.04}$
ProNoG	52.61$_{\pm 2.75}$	8.87$_{\pm 0.69}$	53.09$_{\pm 0.86}$	28.51$_{\pm 3.59}$	28.77$_{\pm 2.68}$	55.38$_{\pm 6.67}$	12.53$_{\pm 3.09}$	43.89$_{\pm 2.14}$	22.98$_{\pm 1.15}$

Table 18: 5-shot configuration results for GAT across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	76.45±4.21	26.60±1.53	54.11±0.47	28.82±2.61	28.80±4.19	47.98±5.27	33.13±23.94	50.63±0.71	22.25±2.22
All-in-One	76.71±4.34	15.03±1.60	51.95±0.92	31.76±1.57	24.43±2.91	56.06±5.81	108.70±0.06	**55.98**±2.50	11.20±1.64
EdgePrompt	59.57±18.75	**28.41**±1.25	53.31±0.75	22.32±7.46	26.92±2.53	53.96±11.29	111.97±0.21	50.19±0.36	13.45±5.17
EdgePrompt+	69.07±4.88	23.53±2.21	53.56±0.36	31.13±4.31	29.67±4.11	**56.49**±13.38	111.97±0.23	50.19±0.98	20.39±4.07
GPF	77.65±2.16	28.27±1.51	53.43±0.65	26.27±1.44	27.77±3.12	50.08±8.40	113.13±0.23	50.48±0.18	12.96±4.15
GPF+	**78.20**±1.15	23.43±1.69	53.36±0.56	23.79±4.49	**31.17**±3.29	48.51±7.95	113.12±0.18	50.83±0.40	13.56±5.60
GPPT	75.91±6.79	17.36±1.76	51.94±1.60	28.80±3.98	27.81±4.60	51.43±12.84	110.13±0.91	50.27±0.78	19.38±2.29
GraphPrompt	71.79±3.34	11.72±1.01	53.48±0.55	34.03±3.30	24.96±4.88	54.70±2.31	111.51±5.82	42.70±3.00	**27.46**±1.47
GraphPrompt+	61.42±5.15	12.22±1.30	53.63±1.01	**35.18**±3.94	22.32±2.43	47.97±3.91	99.66±6.87	42.94±2.51	24.61±1.79
ProNoG	76.55±5.00	22.80±2.43	**54.63**±0.47	26.63±1.60	27.59±3.21	50.94±4.46	**13.74**±3.36	42.28±2.62	25.92±1.27
Pretraining: ContextPred									
Full fine-tuning	**75.50**±3.20	29.02±1.19	54.21±0.44	27.32±2.20	30.29±4.70	55.84±10.35	20.35±0.75	49.58±0.93	25.26±2.54
All-in-One	69.36±3.59	20.00±0.92	51.56±0.51	28.07±3.82	25.12±3.13	52.38±5.15	115.82±0.08	46.74±1.31	22.51±2.40
EdgePrompt	68.58±5.23	**31.38**±0.88	52.99±0.98	36.99±2.41	29.56±1.60	54.03±13.24	115.57±0.10	49.34±0.76	22.20±5.26
EdgePrompt+	68.42±6.28	26.30±0.63	52.85±0.46	**37.45**±3.03	**31.07**±3.28	49.57±15.02	115.38±0.07	49.97±1.33	**28.54**±2.58
GPF	73.93±3.95	31.20±0.91	52.91±1.00	26.87±2.87	28.84±2.22	50.49±12.92	116.56±0.02	49.58±1.46	26.67±2.78
GPF+	74.48±3.02	20.10±11.29	53.02±0.47	27.15±2.61	28.68±3.74	49.68±8.84	116.62±0.13	**50.05**±0.84	25.53±5.97
GPPT	67.34±2.66	26.80±5.03	50.70±1.28	31.59±2.72	28.22±1.48	44.38±12.54	116.27±0.17	50.04±1.16	22.35±2.61
GraphPrompt	70.24±1.84	12.25±1.09	53.10±0.66	28.17±3.75	23.91±4.30	**60.99**±10.52	114.58±0.15	42.89±2.64	27.43±2.21
GraphPrompt+	69.02±2.53	11.13±1.57	53.01±0.71	34.50±3.81	22.45±2.24	47.05±4.02	113.57±1.09	42.73±2.77	21.15±2.82
ProNoG	71.75±2.34	25.39±2.42	**55.41**±0.50	26.60±1.27	28.07±2.66	51.89±11.92	**13.32**±2.16	45.16±1.85	25.59±1.69
Pretraining: DGI									
Full fine-tuning	**75.55**±2.05	22.73±1.25	54.05±0.26	28.26±1.93	30.47±5.14	**55.72**±9.44	31.11±7.76	49.48±1.01	26.08±3.15
All-in-One	59.23±5.81	20.51±0.88	52.26±0.75	28.21±3.58	33.61±1.78	54.33±11.78	111.70±0.12	50.06±0.90	19.87±0.92
EdgePrompt	48.10±2.80	22.15±1.22	53.06±0.63	28.14±4.12	**34.58**±3.45	42.71±10.24	110.52±0.09	**51.36**±1.59	18.63±4.05
EdgePrompt+	61.19±5.38	23.09±0.74	53.10±1.04	32.82±3.27	31.20±3.88	51.40±7.99	110.40±0.08	50.78±1.56	21.00±4.07
GPF	53.91±4.12	23.02±2.28	52.88±0.56	25.61±1.07	33.61±2.05	45.73±12.23	115.50±0.05	51.01±0.93	19.65±4.71
GPF+	57.67±3.47	19.22±9.35	52.71±0.64	26.55±2.57	32.09±4.75	44.58±11.80	115.67±0.37	50.05±0.74	20.66±3.07
GPPT	43.14±7.78	13.53±6.11	51.33±0.69	30.96±5.73	29.13±5.80	46.48±9.55	113.86±0.42	50.02±0.16	16.56±2.90
GraphPrompt	49.69±8.13	8.60±1.75	53.11±0.52	29.16±1.95	23.26±3.83	47.25±10.71	111.87±0.12	43.14±3.19	**26.91**±2.06
GraphPrompt+	53.07±7.97	10.22±2.53	53.14±0.52	**32.85**±6.66	23.77±5.27	50.42±6.69	110.87±0.63	42.95±2.57	21.61±2.51
ProNoG	44.42±8.12	**24.46**±2.79	**54.48**±0.45	27.37±2.32	30.47±3.51	44.08±10.84	**14.04**±2.33	42.77±2.60	20.81±1.66
Pretraining: EdgePred									
Full fine-tuning	**67.75**±6.64	**22.89**±1.70	**54.04**±0.61	26.82±2.17	27.53±4.45	53.19±9.87	20.65±2.88	46.57±1.73	**26.68**±1.81
All-in-One	26.01±1.79	5.04±0.61	50.84±0.58	30.60±4.98	25.99±1.53	54.60±9.31	115.82±0.08	45.16±2.05	19.44±5.10
EdgePrompt	28.00±2.49	10.40±1.99	51.54±0.53	24.17±2.28	**31.74**±2.67	58.67±7.16	115.57±0.10	51.35±2.27	16.52±5.74
EdgePrompt+	56.86±7.73	14.64±1.21	53.16±0.35	29.76±3.65	30.56±5.36	54.60±6.77	115.38±0.07	51.43±1.68	25.23±1.19
GPF	20.24±6.16	9.75±1.08	51.92±1.06	25.56±1.10	28.29±2.93	54.72±7.45	116.56±0.02	**52.60**±1.47	21.70±1.36
GPF+	32.61±5.46	11.37±1.52	52.81±0.58	25.11±0.04	29.25±2.39	54.67±6.61	116.56±0.02	52.35±1.69	22.34±1.97
GPPT	12.16±5.96	2.40±0.68	49.79±0.43	29.81±3.20	25.19±2.98	49.22±9.36	116.27±0.17	49.81±0.71	15.33±2.38
GraphPrompt	16.96±2.58	3.23±0.36	50.70±0.57	31.21±2.04	24.05±3.56	54.05±8.07	114.58±0.15	42.60±2.84	25.24±1.54
GraphPrompt+	36.22±6.18	3.78±0.88	52.65±0.54	**33.81**±3.34	25.25±2.49	54.28±8.29	113.57±1.09	42.67±2.91	21.79±2.74
ProNoG	14.28±3.62	3.97±0.68	53.70±0.60	25.03±7.24	27.31±1.86	**60.26**±5.04	**13.32**±2.16	42.74±2.75	24.50±1.51
Pretraining: GraphCL									
Full fine-tuning	73.68±3.84	**23.28**±1.58	**53.88**±0.42	27.37±2.31	30.08±3.03	**57.21**±5.98	45.35±7.04	48.46±1.10	17.91±1.95
All-in-One	65.18±2.67	18.55±1.29	51.59±0.92	28.60±2.60	26.60±3.80	51.95±11.05	112.93±0.05	48.35±1.59	14.08±2.59
EdgePrompt	69.60±6.19	22.63±2.97	53.45±0.44	35.04±2.69	30.32±2.41	56.72±5.14	111.21±0.25	49.04±0.67	13.52±2.39
EdgePrompt+	69.97±5.37	22.11±0.96	52.90±0.52	**36.65**±3.56	**31.18**±2.73	49.33±5.14	111.25±0.22	48.95±0.98	19.32±3.93
GPF	75.10±2.02	22.41±4.01	53.19±0.72	30.36±4.26	30.07±2.56	54.72±5.39	116.65±0.08	49.00±1.12	12.25±2.02
GPF+	**76.61**±2.27	20.52±3.33	53.52±0.45	27.98±2.13	30.46±2.48	55.43±4.06	116.61±0.15	48.93±1.27	11.40±1.25
GPPT	67.83±2.14	15.37±5.46	50.56±1.08	28.29±6.71	25.80±2.21	55.58±9.04	114.86±0.30	**49.41**±0.95	12.20±0.76
GraphPrompt	66.17±2.60	11.25±1.53	53.55±0.65	28.68±4.84	20.92±3.46	52.83±4.78	112.67±0.09	45.60±2.64	**19.94**±2.15
GraphPrompt+	71.98±2.32	10.50±1.42	53.67±0.41	31.23±3.99	26.15±6.21	51.52±6.37	110.84±0.41	42.86±3.15	17.73±3.20
ProNoG	64.90±1.41	21.93±3.18	53.46±1.21	28.56±3.75	26.60±2.46	56.51±10.77	**10.44**±1.66	46.82±2.01	14.82±1.22
Pretraining: InfoGraph									
Full fine-tuning	**73.73**±1.98	**23.14**±1.91	**54.61**±0.28	29.11±1.87	28.78±3.53	53.85±11.02	43.11±7.43	49.19±1.05	21.96±4.74
All-in-One	29.75±2.30	8.88±1.52	50.83±0.79	28.19±2.67	29.90±3.86	50.48±8.18	114.58±0.09	**52.84**±1.32	11.97±1.41
EdgePrompt	51.93±7.22	7.77±2.56	52.87±0.53	27.01±4.91	**32.35**±2.44	56.05±7.15	114.32±0.30	49.69±1.31	14.23±2.64
EdgePrompt+	66.75±5.36	13.69±1.65	53.47±0.44	**33.76**±2.98	31.82±3.56	52.87±10.30	114.35±0.20	49.59±1.82	17.85±4.51
GPF	43.87±3.98	9.23±1.84	52.34±0.77	25.97±2.41	27.35±1.48	52.15±14.02	116.64±0.03	50.79±1.15	12.61±1.74
GPF+	63.68±4.30	10.46±5.05	53.41±0.41	28.39±2.77	27.82±1.53	55.91±9.61	116.65±0.06	51.36±1.14	11.08±1.75
GPPT	34.86±1.98	3.66±0.27	49.93±0.40	27.59±3.88	29.14±3.00	52.98±4.26	116.41±0.30	49.97±0.17	18.01±2.15
GraphPrompt	29.55±2.17	5.06±0.67	52.37±0.78	29.52±0.99	25.54±6.20	61.57±13.42	116.01±1.18	47.53±1.92	**23.33**±1.33
GraphPrompt+	68.15±4.43	5.91±0.77	53.33±0.87	33.01±1.68	22.66±3.23	53.47±5.11	111.65±0.97	42.68±3.02	22.33±1.28
ProNoG	35.53±0.97	8.60±0.64	53.96±0.60	27.66±2.66	28.35±3.06	**64.57**±5.62	**9.10**±1.99	43.74±2.14	20.04±1.40

Table 19: 5-shot configuration results for GIN across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	**Photo** NC / Acc. ↑	**ogbn-arxiv** NC / Acc. ↑	**DBLP** LP / AUC ↑	**Airports** NC / Acc. ↑	**Chameleon** NC / Acc. ↑	**Cornell** LP / AUC ↑	**QM7b** GR / MAE ↓	**ToxCast** GC / AUC ↑	**MNIST** GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	**75.84**$_{\pm 1.45}$	**27.54**$_{\pm 0.71}$	53.49$_{\pm 0.84}$	33.64$_{\pm 2.66}$	**29.67**$_{\pm 2.38}$	54.68$_{\pm 4.60}$	**41.38**$_{\pm 2.62}$	47.99$_{\pm 2.53}$	**28.22**$_{\pm 1.25}$
All-in-One	72.98$_{\pm 3.70}$	10.30$_{\pm 2.10}$	52.15$_{\pm 0.11}$	32.14$_{\pm 2.67}$	27.98$_{\pm 4.92}$	54.79$_{\pm 4.78}$	59.26$_{\pm 1.43}$	**53.97**$_{\pm 4.02}$	19.82$_{\pm 5.25}$
EdgePrompt	60.45$_{\pm 4.78}$	14.13$_{\pm 1.32}$	54.31$_{\pm 0.77}$	33.21$_{\pm 3.23}$	24.47$_{\pm 4.28}$	46.11$_{\pm 6.56}$	95.57$_{\pm 1.49}$	48.47$_{\pm 1.23}$	22.92$_{\pm 0.97}$
EdgePrompt+	54.49$_{\pm 18.55}$	13.60$_{\pm 1.56}$	54.95$_{\pm 1.10}$	31.90$_{\pm 3.94}$	24.99$_{\pm 4.52}$	51.25$_{\pm 10.73}$	93.85$_{\pm 1.33}$	48.18$_{\pm 0.52}$	25.28$_{\pm 0.98}$
GPF	68.74$_{\pm 4.70}$	16.97$_{\pm 3.68}$	52.33$_{\pm 0.33}$	**35.85**$_{\pm 4.58}$	23.99$_{\pm 4.10}$	46.11$_{\pm 9.75}$	97.29$_{\pm 1.49}$	49.54$_{\pm 1.94}$	24.95$_{\pm 2.23}$
GPF+	72.10$_{\pm 3.12}$	18.47$_{\pm 2.32}$	52.57$_{\pm 0.68}$	35.78$_{\pm 4.30}$	26.54$_{\pm 4.25}$	46.61$_{\pm 8.09}$	97.07$_{\pm 0.69}$	49.39$_{\pm 1.64}$	24.14$_{\pm 2.88}$
GPPT	51.52$_{\pm 10.79}$	18.60$_{\pm 6.61}$	51.76$_{\pm 1.42}$	32.44$_{\pm 4.46}$	23.68$_{\pm 2.82}$	51.00$_{\pm 7.33}$	75.19$_{\pm 5.35}$	50.05$_{\pm 0.65}$	17.49$_{\pm 2.39}$
GraphPrompt	66.87$_{\pm 4.31}$	13.58$_{\pm 1.97}$	53.29$_{\pm 0.81}$	34.99$_{\pm 3.65}$	24.58$_{\pm 4.56}$	51.11$_{\pm 4.41}$	67.37$_{\pm 33.79}$	44.11$_{\pm 3.34}$	23.70$_{\pm 1.15}$
GraphPrompt+	44.22$_{\pm 7.38}$	8.90$_{\pm 1.83}$	53.78$_{\pm 0.46}$	35.33$_{\pm 4.17}$	22.23$_{\pm 2.29}$	49.63$_{\pm 7.23}$	41.73$_{\pm 1.57}$	42.73$_{\pm 2.79}$	21.36$_{\pm 1.07}$
ProNoG	74.40$_{\pm 5.06}$	25.51$_{\pm 4.74}$	**55.17**$_{\pm 0.79}$	30.77$_{\pm 4.48}$	25.79$_{\pm 1.81}$	**55.38**$_{\pm 8.61}$	55.81$_{\pm 41.58}$	42.81$_{\pm 2.67}$	24.03$_{\pm 1.18}$
Pretraining: ContextPred									
Full fine-tuning	71.93$_{\pm 2.94}$	**22.44**$_{\pm 1.70}$	56.71$_{\pm 0.52}$	33.91$_{\pm 3.31}$	29.80$_{\pm 3.09}$	58.44$_{\pm 9.09}$	42.98$_{\pm 3.78}$	47.05$_{\pm 1.07}$	**27.14**$_{\pm 1.92}$
All-in-One	67.78$_{\pm 2.00}$	12.93$_{\pm 2.18}$	54.40$_{\pm 0.54}$	32.60$_{\pm 4.09}$	27.80$_{\pm 4.15}$	59.01$_{\pm 5.36}$	35.77$_{\pm 1.65}$	**54.01**$_{\pm 3.01}$	21.17$_{\pm 7.56}$
EdgePrompt	60.50$_{\pm 1.28}$	4.97$_{\pm 0.66}$	**56.95**$_{\pm 0.26}$	30.34$_{\pm 3.27}$	26.71$_{\pm 4.45}$	57.12$_{\pm 5.31}$	42.33$_{\pm 2.04}$	52.39$_{\pm 1.93}$	24.27$_{\pm 3.04}$
EdgePrompt+	56.71$_{\pm 3.51}$	6.96$_{\pm 2.10}$	56.70$_{\pm 0.31}$	33.21$_{\pm 1.82}$	26.95$_{\pm 5.39}$	55.48$_{\pm 6.81}$	32.62$_{\pm 1.56}$	52.46$_{\pm 2.22}$	25.09$_{\pm 2.26}$
GPF	69.72$_{\pm 2.51}$	10.82$_{\pm 1.93}$	56.34$_{\pm 0.75}$	36.07$_{\pm 5.10}$	25.65$_{\pm 4.60}$	55.89$_{\pm 7.05}$	33.95$_{\pm 2.13}$	51.05$_{\pm 2.16}$	21.21$_{\pm 6.25}$
GPF+	70.03$_{\pm 4.72}$	10.21$_{\pm 4.18}$	56.28$_{\pm 0.49}$	**36.21**$_{\pm 4.91}$	**30.37**$_{\pm 5.85}$	54.08$_{\pm 8.60}$	34.99$_{\pm 1.92}$	49.50$_{\pm 2.52}$	21.76$_{\pm 6.75}$
GPPT	70.08$_{\pm 3.75}$	3.65$_{\pm 1.65}$	50.49$_{\pm 0.48}$	32.55$_{\pm 1.63}$	24.77$_{\pm 4.70}$	60.29$_{\pm 6.19}$	41.86$_{\pm 1.66}$	50.78$_{\pm 1.16}$	22.13$_{\pm 2.10}$
GraphPrompt	71.31$_{\pm 4.37}$	7.11$_{\pm 0.84}$	54.83$_{\pm 0.55}$	31.95$_{\pm 6.42}$	24.05$_{\pm 4.04}$	**66.76**$_{\pm 7.43}$	42.99$_{\pm 2.25}$	44.49$_{\pm 2.07}$	26.52$_{\pm 2.68}$
GraphPrompt+	51.63$_{\pm 13.25}$	3.80$_{\pm 1.93}$	55.20$_{\pm 0.69}$	35.11$_{\pm 4.03}$	22.28$_{\pm 2.17}$	59.57$_{\pm 6.54}$	43.85$_{\pm 1.33}$	43.21$_{\pm 2.30}$	21.79$_{\pm 1.00}$
ProNoG	**73.91**$_{\pm 3.66}$	13.65$_{\pm 3.07}$	55.78$_{\pm 0.50}$	27.56$_{\pm 6.16}$	23.25$_{\pm 2.73}$	54.57$_{\pm 8.97}$	**25.85**$_{\pm 11.80}$	46.28$_{\pm 1.06}$	24.79$_{\pm 1.25}$
Pretraining: DGI									
Full fine-tuning	**75.87**$_{\pm 5.11}$	17.91$_{\pm 2.13}$	53.79$_{\pm 0.73}$	34.34$_{\pm 3.18}$	31.27$_{\pm 4.41}$	53.34$_{\pm 7.67}$	41.31$_{\pm 4.51}$	46.90$_{\pm 2.46}$	22.33$_{\pm 2.04}$
All-in-One	70.25$_{\pm 3.60}$	5.25$_{\pm 3.19}$	51.43$_{\pm 1.38}$	32.36$_{\pm 3.03}$	29.57$_{\pm 4.17}$	47.63$_{\pm 8.56}$	63.35$_{\pm 1.50}$	**53.68**$_{\pm 1.61}$	16.72$_{\pm 8.07}$
EdgePrompt	58.65$_{\pm 6.71}$	20.95$_{\pm 3.05}$	53.68$_{\pm 0.71}$	26.75$_{\pm 8.63}$	**33.85**$_{\pm 4.08}$	**55.72**$_{\pm 12.21}$	98.26$_{\pm 0.65}$	48.71$_{\pm 2.53}$	18.33$_{\pm 4.27}$
EdgePrompt+	56.97$_{\pm 6.86}$	21.73$_{\pm 1.20}$	53.85$_{\pm 0.71}$	32.91$_{\pm 6.22}$	33.06$_{\pm 5.12}$	53.95$_{\pm 9.18}$	96.91$_{\pm 1.40}$	48.96$_{\pm 2.26}$	**22.46**$_{\pm 1.77}$
GPF	72.78$_{\pm 6.35}$	13.56$_{\pm 2.81}$	52.78$_{\pm 0.73}$	32.38$_{\pm 5.59}$	31.16$_{\pm 3.44}$	52.82$_{\pm 4.51}$	99.56$_{\pm 1.22}$	49.83$_{\pm 1.96}$	16.19$_{\pm 6.13}$
GPF+	72.34$_{\pm 5.20}$	19.39$_{\pm 5.46}$	52.72$_{\pm 0.57}$	31.50$_{\pm 4.01}$	32.59$_{\pm 3.91}$	51.56$_{\pm 4.47}$	98.82$_{\pm 0.97}$	47.86$_{\pm 0.99}$	19.94$_{\pm 2.30}$
GPPT	23.66$_{\pm 11.54}$	5.03$_{\pm 2.04}$	50.50$_{\pm 1.48}$	29.28$_{\pm 3.57}$	20.02$_{\pm 7.35}$	43.26$_{\pm 13.81}$	79.70$_{\pm 5.45}$	49.54$_{\pm 0.85}$	13.11$_{\pm 0.96}$
GraphPrompt	52.81$_{\pm 5.43}$	11.82$_{\pm 2.50}$	53.76$_{\pm 0.64}$	30.92$_{\pm 3.38}$	21.04$_{\pm 3.31}$	53.91$_{\pm 9.79}$	46.57$_{\pm 2.37}$	44.05$_{\pm 2.01}$	19.64$_{\pm 0.87}$
GraphPrompt+	44.14$_{\pm 16.99}$	5.03$_{\pm 1.13}$	53.01$_{\pm 0.53}$	**34.50**$_{\pm 5.75}$	23.05$_{\pm 1.92}$	54.22$_{\pm 8.02}$	44.83$_{\pm 1.98}$	43.22$_{\pm 2.79}$	21.84$_{\pm 2.06}$
ProNoG	67.63$_{\pm 6.50}$	**21.94**$_{\pm 4.95}$	**56.57**$_{\pm 0.52}$	27.79$_{\pm 2.31}$	28.23$_{\pm 4.19}$	50.84$_{\pm 14.59}$	**16.58**$_{\pm 9.57}$	43.71$_{\pm 2.65}$	15.43$_{\pm 0.51}$
Pretraining: EdgePred									
Full fine-tuning	**65.44**$_{\pm 4.00}$	**7.37**$_{\pm 1.30}$	52.35$_{\pm 0.75}$	33.69$_{\pm 2.13}$	29.02$_{\pm 3.19}$	48.56$_{\pm 10.42}$	44.29$_{\pm 6.18}$	45.32$_{\pm 2.05}$	**27.24**$_{\pm 1.80}$
All-in-One	22.49$_{\pm 1.91}$	5.09$_{\pm 1.17}$	50.86$_{\pm 0.68}$	28.68$_{\pm 2.54}$	25.38$_{\pm 3.24}$	51.25$_{\pm 12.17}$	36.08$_{\pm 2.00}$	**53.91**$_{\pm 5.96}$	24.76$_{\pm 3.33}$
EdgePrompt	50.07$_{\pm 10.53}$	3.95$_{\pm 1.43}$	52.55$_{\pm 0.93}$	33.73$_{\pm 5.49}$	27.05$_{\pm 5.05}$	**60.69**$_{\pm 10.24}$	36.00$_{\pm 1.98}$	45.07$_{\pm 1.91}$	23.28$_{\pm 2.63}$
EdgePrompt+	48.11$_{\pm 14.93}$	4.01$_{\pm 0.56}$	53.40$_{\pm 1.14}$	31.86$_{\pm 3.03}$	26.86$_{\pm 3.26}$	48.10$_{\pm 9.29}$	35.07$_{\pm 1.59}$	45.22$_{\pm 1.82}$	25.63$_{\pm 0.81}$
GPF	37.56$_{\pm 3.75}$	5.28$_{\pm 1.98}$	50.52$_{\pm 1.21}$	**35.44**$_{\pm 4.50}$	26.97$_{\pm 5.07}$	43.16$_{\pm 14.77}$	36.77$_{\pm 3.24}$	44.92$_{\pm 2.22}$	19.51$_{\pm 5.84}$
GPF+	48.47$_{\pm 3.53}$	3.61$_{\pm 0.73}$	52.28$_{\pm 0.52}$	32.38$_{\pm 1.99}$	**30.37**$_{\pm 2.21}$	45.14$_{\pm 16.25}$	36.74$_{\pm 1.65}$	45.45$_{\pm 2.17}$	23.45$_{\pm 2.98}$
GPPT	13.46$_{\pm 9.04}$	3.21$_{\pm 4.28}$	50.05$_{\pm 0.84}$	33.27$_{\pm 2.95}$	26.81$_{\pm 3.06}$	44.99$_{\pm 11.44}$	51.24$_{\pm 26.55}$	51.09$_{\pm 1.67}$	18.33$_{\pm 0.66}$
GraphPrompt	16.21$_{\pm 2.27}$	2.42$_{\pm 0.70}$	50.86$_{\pm 0.48}$	33.09$_{\pm 2.40}$	22.78$_{\pm 3.42}$	60.10$_{\pm 13.21}$	44.88$_{\pm 2.01}$	42.66$_{\pm 3.06}$	22.41$_{\pm 2.16}$
GraphPrompt+	25.92$_{\pm 7.09}$	2.60$_{\pm 0.73}$	51.54$_{\pm 0.49}$	31.76$_{\pm 3.72}$	23.99$_{\pm 5.01}$	50.40$_{\pm 6.48}$	44.96$_{\pm 1.99}$	42.23$_{\pm 2.11}$	23.55$_{\pm 2.48}$
ProNoG	15.35$_{\pm 2.77}$	2.62$_{\pm 0.46}$	**54.81**$_{\pm 0.54}$	32.02$_{\pm 3.34}$	24.81$_{\pm 1.91}$	59.69$_{\pm 7.11}$	**25.85**$_{\pm 11.80}$	42.60$_{\pm 2.71}$	20.59$_{\pm 1.70}$
Pretraining: GraphCL									
Full fine-tuning	**61.30**$_{\pm 6.00}$	17.52$_{\pm 1.40}$	52.90$_{\pm 0.61}$	**34.63**$_{\pm 2.56}$	28.93$_{\pm 3.53}$	51.61$_{\pm 6.09}$	39.88$_{\pm 4.84}$	46.74$_{\pm 1.51}$	26.51$_{\pm 2.06}$
All-in-One	50.02$_{\pm 2.39}$	**19.00**$_{\pm 0.64}$	51.97$_{\pm 0.93}$	29.74$_{\pm 2.07}$	27.20$_{\pm 2.66}$	46.75$_{\pm 10.58}$	**36.18**$_{\pm 2.03}$	**52.42**$_{\pm 3.35}$	22.39$_{\pm 7.37}$
EdgePrompt	59.67$_{\pm 3.13}$	17.05$_{\pm 1.24}$	52.97$_{\pm 1.10}$	32.92$_{\pm 5.26}$	**28.97**$_{\pm 6.29}$	55.83$_{\pm 10.57}$	41.99$_{\pm 2.30}$	47.65$_{\pm 1.36}$	23.24$_{\pm 2.97}$
EdgePrompt+	57.82$_{\pm 2.83}$	18.19$_{\pm 0.61}$	53.83$_{\pm 1.18}$	33.66$_{\pm 4.51}$	27.24$_{\pm 6.12}$	50.89$_{\pm 11.12}$	40.58$_{\pm 2.80}$	47.82$_{\pm 1.28}$	26.48$_{\pm 1.39}$
GPF	57.34$_{\pm 1.43}$	15.84$_{\pm 1.86}$	52.50$_{\pm 0.72}$	31.38$_{\pm 6.56}$	28.57$_{\pm 5.38}$	42.85$_{\pm 9.13}$	42.43$_{\pm 4.18}$	49.03$_{\pm 1.77}$	20.33$_{\pm 8.16}$
GPF+	56.19$_{\pm 3.49}$	18.27$_{\pm 1.39}$	52.91$_{\pm 0.63}$	29.30$_{\pm 3.77}$	28.93$_{\pm 4.05}$	45.02$_{\pm 6.30}$	41.83$_{\pm 2.97}$	48.99$_{\pm 1.35}$	24.06$_{\pm 2.79}$
GPPT	14.03$_{\pm 5.34}$	11.17$_{\pm 5.25}$	50.53$_{\pm 0.65}$	32.29$_{\pm 2.86}$	20.01$_{\pm 1.90}$	50.99$_{\pm 10.94}$	156.03$_{\pm 224.12}$	49.57$_{\pm 0.96}$	22.09$_{\pm 2.18}$
GraphPrompt	51.11$_{\pm 3.64}$	9.03$_{\pm 1.51}$	53.21$_{\pm 0.72}$	30.70$_{\pm 4.33}$	22.29$_{\pm 3.02}$	47.37$_{\pm 15.04}$	44.86$_{\pm 2.04}$	44.34$_{\pm 2.54}$	**27.42**$_{\pm 2.72}$
GraphPrompt+	61.25$_{\pm 7.89}$	7.76$_{\pm 1.63}$	53.31$_{\pm 0.63}$	34.51$_{\pm 4.83}$	22.18$_{\pm 3.58}$	**56.29**$_{\pm 10.48}$	44.97$_{\pm 1.96}$	42.84$_{\pm 2.96}$	21.71$_{\pm 0.78}$
ProNoG	56.27$_{\pm 2.78}$	18.07$_{\pm 2.34}$	**57.03**$_{\pm 0.70}$	30.91$_{\pm 3.78}$	25.99$_{\pm 3.98}$	49.34$_{\pm 10.99}$	51.79$_{\pm 37.24}$	46.77$_{\pm 1.18}$	26.85$_{\pm 2.38}$
Pretraining: InfoGraph									
Full fine-tuning	71.65$_{\pm 1.78}$	**14.77**$_{\pm 2.36}$	54.11$_{\pm 0.26}$	**33.95**$_{\pm 2.52}$	**30.04**$_{\pm 3.00}$	53.06$_{\pm 7.30}$	36.17$_{\pm 3.05}$	46.36$_{\pm 1.76}$	**25.50**$_{\pm 1.84}$
All-in-One	59.24$_{\pm 0.96}$	2.69$_{\pm 1.56}$	51.58$_{\pm 1.00}$	29.91$_{\pm 3.49}$	27.38$_{\pm 2.70}$	53.70$_{\pm 6.86}$	110.13$_{\pm 1.01}$	**53.90**$_{\pm 3.17}$	17.47$_{\pm 5.38}$
EdgePrompt	60.48$_{\pm 4.41}$	2.80$_{\pm 0.50}$	53.98$_{\pm 1.04}$	32.36$_{\pm 4.58}$	24.31$_{\pm 4.56}$	54.15$_{\pm 8.82}$	111.52$_{\pm 0.24}$	46.28$_{\pm 1.43}$	22.23$_{\pm 4.34}$
EdgePrompt+	62.64$_{\pm 4.87}$	5.02$_{\pm 1.11}$	53.78$_{\pm 0.91}$	32.48$_{\pm 3.89}$	23.93$_{\pm 3.76}$	52.93$_{\pm 9.53}$	111.29$_{\pm 0.04}$	47.28$_{\pm 1.38}$	23.77$_{\pm 2.18}$
GPF	71.86$_{\pm 3.63}$	5.92$_{\pm 1.62}$	54.08$_{\pm 0.78}$	32.38$_{\pm 2.66}$	25.39$_{\pm 4.79}$	47.90$_{\pm 9.46}$	113.70$_{\pm 0.21}$	48.86$_{\pm 1.01}$	18.30$_{\pm 6.27}$
GPF+	**73.38**$_{\pm 2.13}$	5.08$_{\pm 1.92}$	54.33$_{\pm 0.77}$	33.93$_{\pm 4.63}$	28.55$_{\pm 5.71}$	49.54$_{\pm 9.16}$	113.66$_{\pm 0.14}$	48.06$_{\pm 1.06}$	22.54$_{\pm 3.40}$
GPPT	61.38$_{\pm 2.15}$	3.43$_{\pm 4.11}$	49.68$_{\pm 0.08}$	32.56$_{\pm 2.69}$	23.80$_{\pm 2.17}$	51.25$_{\pm 14.05}$	112.68$_{\pm 0.58}$	50.23$_{\pm 0.43}$	18.26$_{\pm 1.81}$
GraphPrompt	54.81$_{\pm 2.40}$	3.21$_{\pm 0.83}$	52.26$_{\pm 0.59}$	31.66$_{\pm 1.21}$	22.63$_{\pm 2.71}$	56.05$_{\pm 6.79}$	113.68$_{\pm 3.52}$	44.52$_{\pm 2.18}$	20.76$_{\pm 1.64}$
GraphPrompt+	48.43$_{\pm 11.83}$	5.18$_{\pm 2.05}$	54.39$_{\pm 1.52}$	33.18$_{\pm 5.48}$	24.96$_{\pm 5.69}$	51.03$_{\pm 5.04}$	75.51$_{\pm 9.74}$	43.06$_{\pm 2.80}$	20.60$_{\pm 1.58}$
ProNoG	60.38$_{\pm 2.27}$	3.21$_{\pm 0.46}$	**56.21**$_{\pm 0.43}$	29.28$_{\pm 3.99}$	24.53$_{\pm 4.16}$	**61.81**$_{\pm 6.64}$	**11.44**$_{\pm 4.29}$	43.74$_{\pm 2.63}$	20.09$_{\pm 1.46}$

Table 20: 5-shot configuration results for H2GCN across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	**76.95**$_{\pm 1.74}$	28.60$_{\pm 1.45}$	**54.29**$_{\pm 0.30}$	25.03$_{\pm 3.16}$	29.72$_{\pm 4.74}$	57.04$_{\pm 12.34}$	**11.87**$_{\pm 0.34}$	50.05$_{\pm 1.75}$	**26.79**$_{\pm 1.28}$
All-in-One	74.69$_{\pm 4.22}$	27.26$_{\pm 3.14}$	53.81$_{\pm 0.31}$	**35.76**$_{\pm 3.93}$	27.90$_{\pm 3.52}$	61.24$_{\pm 9.64}$	105.06$_{\pm 0.65}$	51.40$_{\pm 1.11}$	22.09$_{\pm 2.93}$
EdgePrompt	70.42$_{\pm 2.20}$	**31.25**$_{\pm 3.26}$	54.04$_{\pm 0.45}$	25.97$_{\pm 2.18}$	31.11$_{\pm 4.13}$	61.08$_{\pm 11.77}$	109.00$_{\pm 0.26}$	50.15$_{\pm 2.33}$	21.01$_{\pm 6.28}$
EdgePrompt+	70.21$_{\pm 4.30}$	26.47$_{\pm 2.80}$	53.15$_{\pm 0.66}$	28.21$_{\pm 5.17}$	30.43$_{\pm 3.80}$	61.81$_{\pm 16.74}$	108.06$_{\pm 0.35}$	49.19$_{\pm 1.20}$	24.62$_{\pm 1.38}$
GPF	75.03$_{\pm 2.21}$	29.19$_{\pm 2.77}$	54.05$_{\pm 0.45}$	25.16$_{\pm 0.78}$	30.22$_{\pm 2.39}$	63.97$_{\pm 9.70}$	110.75$_{\pm 0.21}$	51.68$_{\pm 2.39}$	21.15$_{\pm 1.67}$
GPF+	75.02$_{\pm 3.15}$	21.08$_{\pm 7.62}$	53.37$_{\pm 0.52}$	26.27$_{\pm 1.48}$	**31.40**$_{\pm 2.90}$	65.32$_{\pm 10.05}$	110.76$_{\pm 0.23}$	**51.74**$_{\pm 1.21}$	20.86$_{\pm 2.22}$
GPPT	66.56$_{\pm 4.83}$	21.11$_{\pm 3.46}$	50.02$_{\pm 0.54}$	28.14$_{\pm 8.33}$	31.23$_{\pm 3.25}$	53.51$_{\pm 12.08}$	108.56$_{\pm 1.22}$	50.31$_{\pm 0.70}$	18.05$_{\pm 3.91}$
GraphPrompt	67.31$_{\pm 3.97}$	12.05$_{\pm 1.32}$	53.23$_{\pm 0.45}$	34.77$_{\pm 3.24}$	25.05$_{\pm 4.76}$	64.30$_{\pm 11.87}$	99.27$_{\pm 0.82}$	42.60$_{\pm 2.64}$	21.80$_{\pm 2.24}$
GraphPrompt+	54.23$_{\pm 3.17}$	10.61$_{\pm 1.62}$	54.21$_{\pm 0.65}$	35.16$_{\pm 3.93}$	21.82$_{\pm 3.27}$	53.38$_{\pm 10.17}$	58.94$_{\pm 18.31}$	42.93$_{\pm 2.83}$	21.22$_{\pm 1.38}$
ProNoG	71.54$_{\pm 2.25}$	26.98$_{\pm 5.03}$	53.97$_{\pm 0.83}$	26.80$_{\pm 2.67}$	28.52$_{\pm 4.12}$	**70.00**$_{\pm 5.18}$	16.30$_{\pm 9.25}$	43.10$_{\pm 2.70}$	22.05$_{\pm 3.31}$
Pretraining: ContextPred									
Full fine-tuning	76.93$_{\pm 3.61}$	26.60$_{\pm 1.98}$	**54.28**$_{\pm 0.27}$	24.80$_{\pm 3.29}$	30.56$_{\pm 3.48}$	**58.07**$_{\pm 11.74}$	**6.43**$_{\pm 0.91}$	48.90$_{\pm 0.85}$	**26.71**$_{\pm 1.96}$
All-in-One	73.98$_{\pm 3.61}$	22.26$_{\pm 2.03}$	53.45$_{\pm 0.61}$	**35.97**$_{\pm 5.00}$	27.26$_{\pm 3.07}$	56.54$_{\pm 9.81}$	98.75$_{\pm 0.82}$	47.06$_{\pm 1.08}$	26.11$_{\pm 0.61}$
EdgePrompt	67.72$_{\pm 5.34}$	**28.81**$_{\pm 2.52}$	53.49$_{\pm 0.62}$	25.61$_{\pm 0.30}$	29.81$_{\pm 2.77}$	51.77$_{\pm 7.50}$	108.77$_{\pm 0.35}$	**51.02**$_{\pm 0.71}$	24.48$_{\pm 2.13}$
EdgePrompt+	68.54$_{\pm 3.95}$	24.62$_{\pm 0.81}$	53.57$_{\pm 0.40}$	25.28$_{\pm 0.49}$	**31.99**$_{\pm 2.80}$	55.26$_{\pm 11.52}$	108.79$_{\pm 0.24}$	50.96$_{\pm 0.65}$	24.64$_{\pm 2.30}$
GPF	76.79$_{\pm 3.86}$	24.07$_{\pm 12.47}$	53.53$_{\pm 0.62}$	25.32$_{\pm 0.72}$	30.32$_{\pm 2.23}$	54.15$_{\pm 9.56}$	108.79$_{\pm 0.25}$	50.47$_{\pm 1.05}$	19.13$_{\pm 3.87}$
GPF+	77.06$_{\pm 3.04}$	23.31$_{\pm 6.14}$	53.93$_{\pm 0.44}$	25.61$_{\pm 1.78}$	30.57$_{\pm 1.83}$	55.55$_{\pm 10.41}$	109.13$_{\pm 0.67}$	50.72$_{\pm 0.54}$	16.28$_{\pm 4.95}$
GPPT	76.04$_{\pm 2.42}$	28.39$_{\pm 4.30}$	50.50$_{\pm 1.09}$	33.83$_{\pm 2.78}$	31.47$_{\pm 1.85}$	50.23$_{\pm 8.46}$	104.39$_{\pm 1.99}$	50.07$_{\pm 0.31}$	22.23$_{\pm 0.83}$
GraphPrompt	76.03$_{\pm 3.39}$	13.80$_{\pm 2.05}$	53.76$_{\pm 0.50}$	34.07$_{\pm 5.20}$	26.79$_{\pm 6.10}$	57.40$_{\pm 8.20}$	97.03$_{\pm 13.10}$	42.98$_{\pm 2.58}$	25.51$_{\pm 1.09}$
GraphPrompt+	72.54$_{\pm 5.04}$	9.72$_{\pm 2.51}$	53.67$_{\pm 0.53}$	35.11$_{\pm 3.94}$	22.63$_{\pm 2.33}$	49.66$_{\pm 4.51}$	69.12$_{\pm 14.13}$	42.45$_{\pm 2.98}$	20.22$_{\pm 2.24}$
ProNoG	**78.92**$_{\pm 2.23}$	25.87$_{\pm 4.28}$	53.35$_{\pm 0.76}$	26.26$_{\pm 1.45}$	28.42$_{\pm 3.31}$	57.37$_{\pm 5.76}$	18.82$_{\pm 10.39}$	43.47$_{\pm 2.58}$	24.14$_{\pm 0.44}$
Pretraining: DGI									
Full fine-tuning	76.60$_{\pm 2.59}$	**23.31**$_{\pm 1.26}$	53.95$_{\pm 0.07}$	27.21$_{\pm 2.81}$	30.88$_{\pm 3.63}$	57.55$_{\pm 10.73}$	**9.08**$_{\pm 0.82}$	48.86$_{\pm 1.42}$	**25.71**$_{\pm 1.50}$
All-in-One	74.30$_{\pm 2.54}$	18.66$_{\pm 1.88}$	**54.34**$_{\pm 1.00}$	**32.97**$_{\pm 4.26}$	28.71$_{\pm 3.43}$	55.15$_{\pm 14.41}$	103.84$_{\pm 0.28}$	48.74$_{\pm 2.07}$	24.50$_{\pm 1.56}$
EdgePrompt	69.31$_{\pm 4.89}$	22.77$_{\pm 2.28}$	53.02$_{\pm 0.61}$	26.02$_{\pm 2.19}$	30.68$_{\pm 2.23}$	59.79$_{\pm 9.21}$	88.47$_{\pm 2.13}$	50.26$_{\pm 2.01}$	23.45$_{\pm 0.91}$
EdgePrompt+	70.42$_{\pm 3.71}$	22.84$_{\pm 1.35}$	53.58$_{\pm 0.56}$	26.29$_{\pm 2.60}$	30.68$_{\pm 2.44}$	58.86$_{\pm 9.70}$	84.45$_{\pm 1.98}$	51.35$_{\pm 1.85}$	21.31$_{\pm 4.35}$
GPF	**77.74**$_{\pm 3.28}$	19.38$_{\pm 7.94}$	53.38$_{\pm 0.69}$	26.14$_{\pm 5.63}$	30.56$_{\pm 1.50}$	54.78$_{\pm 13.57}$	111.54$_{\pm 0.21}$	**51.53**$_{\pm 1.77}$	17.47$_{\pm 5.33}$
GPF+	77.37$_{\pm 2.95}$	17.59$_{\pm 8.06}$	53.39$_{\pm 0.47}$	26.19$_{\pm 1.78}$	30.37$_{\pm 2.98}$	54.91$_{\pm 12.02}$	111.87$_{\pm 0.75}$	51.00$_{\pm 1.14}$	16.13$_{\pm 6.05}$
GPPT	19.86$_{\pm 6.40}$	16.07$_{\pm 5.11}$	50.08$_{\pm 0.19}$	30.29$_{\pm 1.06}$	**31.27**$_{\pm 2.16}$	49.99$_{\pm 7.17}$	106.25$_{\pm 1.30}$	49.78$_{\pm 0.48}$	13.85$_{\pm 4.18}$
GraphPrompt	70.68$_{\pm 4.03}$	9.35$_{\pm 1.79}$	52.48$_{\pm 0.63}$	31.76$_{\pm 4.07}$	25.99$_{\pm 4.58}$	48.62$_{\pm 9.70}$	98.98$_{\pm 3.47}$	43.06$_{\pm 2.72}$	24.12$_{\pm 1.08}$
GraphPrompt+	62.01$_{\pm 6.59}$	6.19$_{\pm 1.11}$	52.91$_{\pm 0.34}$	31.95$_{\pm 1.70}$	22.08$_{\pm 2.45}$	51.64$_{\pm 10.04}$	48.79$_{\pm 6.48}$	42.24$_{\pm 3.01}$	21.37$_{\pm 2.03}$
ProNoG	72.75$_{\pm 6.90}$	19.85$_{\pm 3.14}$	53.22$_{\pm 0.86}$	27.61$_{\pm 2.59}$	28.88$_{\pm 3.78}$	**62.86**$_{\pm 6.78}$	42.28$_{\pm 31.30}$	42.87$_{\pm 2.53}$	24.04$_{\pm 1.32}$
Pretraining: EdgePred									
Full fine-tuning	**74.34**$_{\pm 1.94}$	**19.69**$_{\pm 1.24}$	**54.24**$_{\pm 0.52}$	25.81$_{\pm 2.70}$	**30.27**$_{\pm 4.09}$	63.58$_{\pm 9.74}$	**6.43**$_{\pm 0.91}$	48.07$_{\pm 1.01}$	**25.65**$_{\pm 1.02}$
All-in-One	25.76$_{\pm 5.97}$	2.44$_{\pm 0.88}$	51.46$_{\pm 0.52}$	**37.64**$_{\pm 4.45}$	26.48$_{\pm 4.83}$	57.79$_{\pm 11.07}$	98.75$_{\pm 0.82}$	49.54$_{\pm 1.25}$	24.56$_{\pm 1.98}$
EdgePrompt	50.52$_{\pm 13.18}$	6.20$_{\pm 2.33}$	52.53$_{\pm 0.64}$	25.06$_{\pm 0.16}$	29.11$_{\pm 2.65}$	59.54$_{\pm 11.58}$	108.77$_{\pm 0.35}$	**50.21**$_{\pm 1.88}$	23.30$_{\pm 1.56}$
EdgePrompt+	65.80$_{\pm 5.49}$	7.33$_{\pm 2.69}$	52.58$_{\pm 0.16}$	25.32$_{\pm 0.18}$	30.23$_{\pm 5.37}$	58.31$_{\pm 9.29}$	108.79$_{\pm 0.24}$	48.83$_{\pm 0.28}$	23.51$_{\pm 1.73}$
GPF	40.44$_{\pm 9.38}$	4.01$_{\pm 1.54}$	51.85$_{\pm 0.68}$	25.06$_{\pm 0.13}$	28.44$_{\pm 3.08}$	**65.54**$_{\pm 8.29}$	108.79$_{\pm 0.25}$	48.96$_{\pm 0.68}$	20.65$_{\pm 5.20}$
GPF+	56.78$_{\pm 8.79}$	6.25$_{\pm 0.74}$	52.62$_{\pm 0.85}$	25.73$_{\pm 0.72}$	27.91$_{\pm 3.65}$	61.75$_{\pm 9.93}$	110.49$_{\pm 3.71}$	48.92$_{\pm 0.83}$	23.79$_{\pm 0.49}$
GPPT	14.48$_{\pm 5.09}$	2.87$_{\pm 1.70}$	49.82$_{\pm 0.68}$	32.65$_{\pm 2.61}$	22.99$_{\pm 2.74}$	56.33$_{\pm 8.84}$	104.39$_{\pm 1.99}$	50.04$_{\pm 0.17}$	19.00$_{\pm 1.21}$
GraphPrompt	19.38$_{\pm 3.43}$	4.37$_{\pm 0.94}$	50.36$_{\pm 0.49}$	34.38$_{\pm 3.48}$	24.33$_{\pm 2.95}$	55.37$_{\pm 13.01}$	87.42$_{\pm 0.46}$	46.19$_{\pm 2.86}$	23.64$_{\pm 1.43}$
GraphPrompt+	43.80$_{\pm 7.04}$	2.72$_{\pm 0.92}$	52.10$_{\pm 0.17}$	35.28$_{\pm 3.84}$	24.33$_{\pm 3.79}$	52.39$_{\pm 10.09}$	69.12$_{\pm 14.13}$	42.81$_{\pm 2.83}$	19.22$_{\pm 2.76}$
ProNoG	17.81$_{\pm 2.60}$	7.42$_{\pm 0.55}$	52.44$_{\pm 0.63}$	27.16$_{\pm 2.49}$	22.88$_{\pm 2.94}$	59.62$_{\pm 6.56}$	18.82$_{\pm 10.39}$	44.19$_{\pm 2.23}$	21.88$_{\pm 1.23}$
Pretraining: GraphCL									
Full fine-tuning	**77.02**$_{\pm 3.00}$	23.28$_{\pm 0.62}$	**54.12**$_{\pm 0.27}$	26.53$_{\pm 2.73}$	28.62$_{\pm 4.32}$	**60.12**$_{\pm 12.54}$	**6.97**$_{\pm 0.68}$	49.17$_{\pm 1.42}$	**24.54**$_{\pm 1.88}$
All-in-One	67.14$_{\pm 2.05}$	22.55$_{\pm 2.42}$	53.40$_{\pm 0.37}$	32.39$_{\pm 5.84}$	27.91$_{\pm 1.61}$	54.73$_{\pm 11.40}$	89.82$_{\pm 1.25}$	49.53$_{\pm 2.70}$	23.03$_{\pm 0.89}$
EdgePrompt	67.51$_{\pm 5.31}$	**26.35**$_{\pm 1.29}$	52.92$_{\pm 0.69}$	28.15$_{\pm 3.73}$	30.67$_{\pm 2.49}$	50.70$_{\pm 6.55}$	106.09$_{\pm 0.47}$	**50.07**$_{\pm 0.98}$	21.51$_{\pm 2.11}$
EdgePrompt+	70.21$_{\pm 4.96}$	23.13$_{\pm 1.21}$	53.25$_{\pm 0.47}$	**34.00**$_{\pm 2.81}$	**31.18**$_{\pm 2.25}$	51.01$_{\pm 9.41}$	106.10$_{\pm 0.34}$	49.48$_{\pm 0.96}$	23.74$_{\pm 0.57}$
GPF	76.89$_{\pm 1.93}$	24.59$_{\pm 5.13}$	53.21$_{\pm 0.39}$	25.98$_{\pm 2.87}$	30.36$_{\pm 2.45}$	52.90$_{\pm 10.67}$	106.14$_{\pm 0.29}$	49.26$_{\pm 1.68}$	20.30$_{\pm 4.40}$
GPF+	76.70$_{\pm 1.71}$	19.80$_{\pm 6.62}$	53.63$_{\pm 0.27}$	26.63$_{\pm 2.40}$	30.67$_{\pm 2.01}$	56.87$_{\pm 11.51}$	106.73$_{\pm 1.15}$	49.18$_{\pm 1.24}$	19.77$_{\pm 4.29}$
GPPT	66.93$_{\pm 1.84}$	12.12$_{\pm 2.86}$	50.24$_{\pm 0.90}$	30.10$_{\pm 3.92}$	30.12$_{\pm 1.86}$	53.57$_{\pm 11.03}$	99.91$_{\pm 3.12}$	49.43$_{\pm 0.74}$	18.52$_{\pm 2.44}$
GraphPrompt	66.58$_{\pm 3.57}$	10.70$_{\pm 1.53}$	53.72$_{\pm 0.29}$	27.81$_{\pm 4.87}$	26.29$_{\pm 4.81}$	52.15$_{\pm 9.84}$	80.25$_{\pm 17.73}$	44.54$_{\pm 2.51}$	23.38$_{\pm 1.59}$
GraphPrompt+	71.09$_{\pm 2.71}$	11.22$_{\pm 1.76}$	53.90$_{\pm 0.32}$	33.28$_{\pm 3.73}$	26.11$_{\pm 4.36}$	50.38$_{\pm 4.84}$	55.93$_{\pm 22.12}$	42.45$_{\pm 2.91}$	20.48$_{\pm 3.20}$
ProNoG	63.55$_{\pm 2.31}$	18.30$_{\pm 1.73}$	53.39$_{\pm 1.24}$	26.58$_{\pm 3.38}$	29.22$_{\pm 2.55}$	58.88$_{\pm 15.89}$	48.49$_{\pm 49.84}$	46.24$_{\pm 1.69}$	21.94$_{\pm 1.22}$
Pretraining: InfoGraph									
Full fine-tuning	**76.71**$_{\pm 2.53}$	26.42$_{\pm 1.27}$	54.32$_{\pm 0.51}$	25.33$_{\pm 2.99}$	29.80$_{\pm 4.54}$	**63.28**$_{\pm 9.70}$	**7.39**$_{\pm 1.17}$	50.88$_{\pm 1.32}$	**26.40**$_{\pm 0.95}$
All-in-One	62.42$_{\pm 1.40}$	19.45$_{\pm 1.02}$	53.21$_{\pm 0.54}$	**35.90**$_{\pm 4.66}$	29.15$_{\pm 4.47}$	58.73$_{\pm 11.21}$	111.67$_{\pm 0.14}$	**52.29**$_{\pm 2.72}$	21.94$_{\pm 1.14}$
EdgePrompt	67.53$_{\pm 4.96}$	**28.10**$_{\pm 2.17}$	52.90$_{\pm 0.55}$	25.01$_{\pm 0.28}$	31.12$_{\pm 2.93}$	51.26$_{\pm 5.74}$	115.17$_{\pm 0.17}$	52.18$_{\pm 0.92}$	20.72$_{\pm 1.89}$
EdgePrompt+	69.73$_{\pm 2.62}$	25.73$_{\pm 0.75}$	53.38$_{\pm 0.20}$	27.42$_{\pm 2.96}$	30.96$_{\pm 4.46}$	61.01$_{\pm 7.53}$	115.13$_{\pm 0.17}$	52.18$_{\pm 1.33}$	24.42$_{\pm 1.45}$
GPF	74.30$_{\pm 2.74}$	26.51$_{\pm 1.72}$	52.97$_{\pm 0.59}$	25.95$_{\pm 1.66}$	31.51$_{\pm 2.59}$	51.20$_{\pm 14.11}$	115.51$_{\pm 0.18}$	50.76$_{\pm 1.68}$	19.16$_{\pm 6.14}$
GPF+	76.29$_{\pm 2.48}$	20.26$_{\pm 7.50}$	53.54$_{\pm 0.44}$	27.03$_{\pm 4.18}$	**32.29**$_{\pm 2.74}$	61.19$_{\pm 10.51}$	115.55$_{\pm 0.13}$	51.90$_{\pm 0.93}$	22.29$_{\pm 1.74}$
GPPT	65.30$_{\pm 1.47}$	13.92$_{\pm 2.28}$	50.20$_{\pm 1.41}$	27.47$_{\pm 3.06}$	30.75$_{\pm 3.24}$	50.61$_{\pm 8.93}$	113.01$_{\pm 0.73}$	50.04$_{\pm 0.39}$	16.55$_{\pm 3.39}$
GraphPrompt	62.45$_{\pm 2.89}$	8.99$_{\pm 1.38}$	53.33$_{\pm 0.75}$	31.06$_{\pm 4.79}$	25.96$_{\pm 4.87}$	57.41$_{\pm 9.91}$	111.03$_{\pm 0.11}$	44.91$_{\pm 2.45}$	24.84$_{\pm 1.60}$
GraphPrompt+	73.92$_{\pm 4.48}$	6.84$_{\pm 1.28}$	53.87$_{\pm 0.62}$	34.31$_{\pm 4.36}$	23.77$_{\pm 2.59}$	51.77$_{\pm 8.32}$	88.02$_{\pm 15.79}$	42.75$_{\pm 2.96}$	21.86$_{\pm 1.96}$
ProNoG	68.41$_{\pm 0.66}$	23.13$_{\pm 2.41}$	**54.91**$_{\pm 0.73}$	27.35$_{\pm 2.04}$	28.06$_{\pm 1.73}$	52.93$_{\pm 12.51}$	8.40$_{\pm 3.46}$	44.52$_{\pm 1.94}$	22.28$_{\pm 1.08}$

Table 21: 5-shot configuration results for FAGCN across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	76.57±1.07	27.53±1.09	54.17±0.25	29.28±1.77	**32.49**±2.97	**54.17**±8.06	**6.38**±0.63	48.50±0.86	23.80±1.65
All-in-One	76.84±2.32	**29.69**±0.97	53.96±0.40	31.95±2.81	28.35±2.32	52.20±11.40	78.06±0.55	**52.13**±3.57	22.29±2.47
EdgePrompt	68.55±6.14	24.37±3.15	53.83±0.53	32.32±4.39	24.89±5.02	53.26±9.50	91.73±0.60	49.31±1.50	23.98±1.23
EdgePrompt+	69.09±3.76	24.14±3.95	54.13±0.41	33.68±5.39	25.95±4.61	49.04±9.61	90.65±0.71	48.96±0.71	**25.06**±1.03
GPF	76.56±1.75	28.82±1.63	54.07±0.50	31.74±5.16	24.97±1.88	50.57±7.60	101.41±1.86	49.93±0.77	20.61±4.74
GPF+	**77.63**±2.50	20.60±9.03	53.89±0.38	29.42±2.91	25.68±2.08	51.34±7.27	101.28±1.60	49.88±0.56	22.04±1.67
GPPT	43.81±10.53	19.63±6.20	51.04±1.47	34.26±1.91	22.85±4.18	51.36±11.47	86.29±7.90	48.98±0.81	16.77±2.96
GraphPrompt	64.18±4.01	12.05±2.24	53.71±0.29	**35.16**±3.93	23.24±1.95	53.10±7.35	68.05±33.95	42.70±2.86	24.27±1.54
GraphPrompt+	50.86±3.52	12.87±1.71	**54.19**±0.50	**35.16**±3.92	22.01±2.46	44.58±4.10	29.22±5.33	42.83±2.80	21.41±4.02
ProNoG	73.18±3.37	25.75±4.22	54.05±0.47	26.84±2.45	26.55±2.31	49.29±5.14	20.23±2.90	42.93±2.83	24.26±1.18
Pretraining: ContextPred									
Full fine-tuning	79.93±3.11	**23.77**±1.46	**54.55**±0.54	28.79±2.12	**31.05**±3.59	53.13±7.01	**6.24**±0.38	48.08±0.93	**25.37**±0.93
All-in-One	74.81±4.43	18.47±1.98	53.28±0.73	34.21±2.45	28.80±1.38	55.33±10.88	100.97±1.51	42.72±2.96	22.50±1.88
EdgePrompt	71.53±5.19	18.42±3.79	53.71±0.60	34.19±5.67	28.23±3.63	52.65±4.44	107.05±0.53	50.06±1.39	21.97±2.51
EdgePrompt+	69.81±4.66	19.51±1.32	54.15±0.30	**35.09**±4.99	28.23±6.21	53.65±7.04	107.40±0.68	50.03±1.34	24.41±1.62
GPF	80.88±4.46	19.54±1.13	53.64±0.44	32.96±3.11	28.57±3.46	48.04±6.75	108.22±0.99	49.64±1.34	21.03±3.01
GPF+	**81.27**±3.42	15.34±5.89	53.70±0.35	33.15±3.44	30.08±3.31	48.59±8.18	108.64±1.87	49.23±0.59	20.73±2.55
GPPT	74.46±2.92	14.06±2.03	49.56±1.06	32.21±3.70	28.37±3.19	50.92±11.64	98.75±4.23	**50.38**±1.20	18.77±2.34
GraphPrompt	75.76±5.33	10.47±2.00	53.59±0.46	31.74±2.92	25.15±5.97	**55.42**±2.97	73.11±1.46	43.52±2.82	24.69±1.42
GraphPrompt+	75.67±3.84	10.15±2.14	54.17±0.48	34.51±3.32	22.42±2.53	53.14±7.50	63.28±14.33	42.71±3.14	21.74±3.89
ProNoG	80.56±3.53	14.69±3.24	53.43±1.07	28.36±3.21	27.52±2.45	53.70±7.34	15.27±5.49	43.66±2.80	24.87±1.44
Pretraining: DGI									
Full fine-tuning	**76.40**±4.64	27.18±1.88	**53.75**±0.55	30.80±4.58	32.67±3.82	57.43±12.84	**7.68**±1.86	47.71±0.90	23.36±1.42
All-in-One	74.84±5.49	25.28±0.70	53.39±0.51	**31.91**±0.85	32.30±3.66	54.60±8.16	105.82±1.53	45.37±1.75	21.34±3.44
EdgePrompt	63.47±5.05	**29.90**±1.82	53.50±0.49	28.89±7.49	34.07±2.93	52.58±9.43	108.19±0.14	50.79±0.41	21.04±1.77
EdgePrompt+	67.43±6.78	29.51±1.33	53.69±0.46	29.62±6.57	**34.22**±1.42	54.89±9.76	108.04±0.62	**50.99**±0.56	26.23±0.64
GPF	74.44±5.93	28.92±1.11	53.12±0.53	27.42±3.18	31.59±2.34	59.74±6.53	110.40±0.21	48.67±1.10	20.57±3.10
GPF+	76.10±4.66	22.97±10.48	53.32±0.48	27.16±6.98	31.33±0.90	53.20±11.71	110.65±0.65	48.94±1.17	20.76±6.73
GPPT	11.25±0.23	22.77±4.69	49.97±1.60	28.58±6.06	30.37±2.31	53.38±14.55	106.89±1.71	49.97±1.46	19.67±2.16
GraphPrompt	66.59±9.67	8.03±1.80	52.98±0.58	29.54±4.02	23.05±4.63	50.67±5.92	95.45±0.56	43.42±2.63	26.79±0.57
GraphPrompt+	68.25±7.13	9.44±2.03	53.63±0.77	29.50±2.12	24.31±3.65	45.20±6.98	68.06±17.05	43.00±2.87	19.39±2.78
ProNoG	68.56±9.04	26.96±3.12	51.96±0.59	28.46±2.32	28.42±3.04	**65.18**±11.23	21.26±16.10	46.09±1.41	**27.28**±0.44
Pretraining: EdgePred									
Full fine-tuning	**75.45**±1.19	**21.32**±0.96	**54.13**±0.31	31.25±3.14	30.53±2.47	48.59±8.32	**6.24**±0.38	47.05±1.36	23.96±1.59
All-in-One	33.44±2.90	5.33±0.69	51.71±0.43	**35.27**±2.17	**30.91**±3.50	**53.13**±6.94	100.97±1.51	46.92±1.24	20.40±1.79
EdgePrompt	19.05±4.67	4.35±1.15	51.77±0.64	32.26±4.28	25.28±7.45	49.20±7.50	107.05±0.53	47.95±0.97	21.27±1.72
EdgePrompt+	50.13±8.34	5.36±1.15	52.55±0.39	29.13±8.46	29.40±5.61	48.32±8.55	107.40±0.68	48.00±0.67	24.24±2.68
GPF	36.76±6.51	4.74±1.14	51.63±0.59	29.76±4.86	30.31±2.98	49.42±4.78	108.22±0.99	48.78±0.62	22.65±1.52
GPF+	67.73±7.31	10.89±2.25	53.62±0.28	27.54±3.16	26.94±5.15	48.23±4.38	107.94±0.49	48.53±0.73	23.16±1.25
GPPT	35.35±3.15	2.95±0.65	50.41±0.60	30.89±2.04	27.69±5.88	52.50±10.27	98.75±4.23	**50.33**±0.14	18.65±2.04
GraphPrompt	29.45±3.09	3.73±0.19	49.90±0.88	35.13±3.55	21.17±4.40	52.35±5.24	73.11±1.46	44.83±1.66	**26.49**±0.91
GraphPrompt+	55.11±3.57	3.87±0.29	52.15±0.68	33.69±3.97	24.29±4.71	48.68±4.07	63.28±14.33	42.84±2.93	18.82±2.52
ProNoG	32.42±2.28	4.43±0.52	51.15±0.63	26.02±4.40	28.77±3.31	52.86±6.46	15.27±5.49	43.62±2.57	23.98±1.51
Pretraining: GraphCL									
Full fine-tuning	77.25±2.30	27.87±1.15	53.87±0.42	29.28±1.74	**31.18**±2.91	53.60±10.29	**6.16**±0.53	48.64±1.48	20.66±1.13
All-in-One	77.65±1.73	**30.31**±0.26	53.55±0.26	29.08±4.11	28.32±2.25	54.44±8.45	94.09±2.07	**55.07**±2.91	20.89±2.23
EdgePrompt	64.00±4.96	28.59±2.48	53.64±0.38	**36.07**±3.43	26.87±2.14	48.68±6.27	102.40±0.28	49.65±1.42	21.33±1.84
EdgePrompt+	64.65±5.19	26.89±3.38	53.78±0.41	34.63±2.84	27.49±2.13	52.21±6.24	102.23±0.39	49.41±1.43	21.27±1.69
GPF	77.65±2.36	30.09±2.36	53.65±0.37	24.89±4.75	27.74±1.48	51.49±5.08	103.01±0.81	49.92±1.66	19.44±1.55
GPF+	**78.47**±1.45	19.74±8.23	53.42±0.75	30.79±2.85	28.94±1.84	52.08±7.56	103.99±2.55	49.94±1.13	18.25±3.66
GPPT	78.46±1.94	15.22±0.90	51.63±0.76	33.09±1.71	27.95±4.86	50.14±10.08	91.04±6.00	49.11±0.87	13.32±0.74
GraphPrompt	76.46±2.77	14.67±2.23	53.89±0.29	35.08±2.93	24.96±6.86	**55.70**±5.60	55.96±2.14	45.30±2.25	21.00±2.00
GraphPrompt+	76.99±3.02	13.59±1.88	**53.97**±0.45	34.89±3.12	23.21±2.92	49.98±4.64	44.72±15.83	43.12±2.63	17.13±2.85
ProNoG	76.99±1.61	27.10±3.22	53.73±0.82	28.27±6.22	26.99±3.13	50.69±8.88	15.91±5.14	45.79±1.84	**23.58**±1.66
Pretraining: InfoGraph									
Full fine-tuning	**77.04**±1.41	**26.12**±1.78	**54.30**±0.35	30.36±3.79	31.76±5.33	54.73±8.20	**6.37**±0.73	48.71±1.11	24.08±2.35
All-in-One	75.84±2.48	25.05±0.82	53.44±0.47	**37.42**±4.48	30.75±2.70	52.29±6.78	112.06±0.25	**56.80**±2.85	16.49±4.67
EdgePrompt	58.86±3.31	20.63±2.43	53.61±0.49	26.73±2.26	**34.17**±3.41	52.68±7.97	112.23±0.14	48.68±0.56	21.46±1.82
EdgePrompt+	63.23±4.78	18.86±1.45	53.92±0.41	32.48±4.99	33.45±3.62	47.92±6.98	112.11±0.07	48.71±0.36	24.64±2.63
GPF	76.20±3.34	24.28±2.10	53.61±0.45	28.60±5.29	32.47±3.25	49.68±5.26	113.86±0.24	49.95±1.97	21.85±1.81
GPF+	76.64±3.75	18.01±6.75	53.80±0.40	26.80±4.42	31.98±2.74	49.85±6.22	113.90±0.15	50.01±0.58	20.41±2.95
GPPT	73.92±2.80	17.07±2.40	51.05±0.95	32.99±3.01	31.80±3.37	48.23±4.00	111.70±0.34	49.97±0.39	20.67±1.64
GraphPrompt	71.62±4.37	7.05±0.33	53.85±0.41	33.61±3.05	25.06±5.15	**55.27**±13.58	111.13±1.60	45.09±2.69	**27.69**±0.83
GraphPrompt+	75.42±3.66	12.63±1.37	54.13±0.72	33.93±0.88	24.38±4.02	48.88±5.00	97.41±5.78	42.75±2.78	21.41±2.12
ProNoG	75.89±0.90	25.75±2.11	53.65±0.77	28.02±2.44	29.08±2.94	52.68±8.43	9.66±5.30	43.28±2.66	24.85±1.07

Table 22: 5-shot configuration results for Transformer across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	80.32±1.36	26.98±1.49	54.84±0.50	33.38±5.47	29.33±3.77	56.19±11.19	15.17±1.60	**49.88**±1.33	**26.40**±2.51
All-in-One	79.45±1.29	**30.08**±0.92	53.37±0.30	35.90±3.94	30.60±2.74	**56.58**±11.18	94.59±5.02	48.01±1.50	23.51±2.22
EdgePrompt	67.44±4.94	25.31±2.59	52.98±0.56	34.14±3.47	30.44±3.37	42.83±7.86	107.44±0.28	49.03±0.40	18.02±3.87
EdgePrompt+	68.39±6.42	23.73±2.37	52.78±0.35	34.62±3.34	30.84±3.79	50.08±6.86	106.96±0.40	49.13±0.81	22.55±2.33
GPF	**80.81**±2.33	23.54±1.59	52.87±0.51	**38.10**±3.36	30.66±3.36	46.71±4.21	111.34±0.29	49.08±0.60	17.53±4.34
GPF+	80.27±1.70	22.42±1.52	52.63±0.61	37.62±4.03	32.08±2.20	45.89±5.52	111.29±0.26	49.08±0.54	18.88±2.99
GPPT	73.02±10.93	18.80±6.48	50.54±1.66	29.98±4.56	**32.50**±3.56	52.65±11.68	108.23±1.59	48.95±1.55	12.01±0.75
GraphPrompt	72.76±1.47	12.56±1.72	52.89±0.39	35.40±4.13	27.83±6.74	48.51±3.37	103.93±4.21	44.30±3.65	21.84±1.26
GraphPrompt+	64.06±7.51	11.78±1.38	52.49±0.62	35.23±4.04	22.18±2.35	47.74±7.83	45.22±35.40	42.47±2.86	22.78±1.27
ProNoG	78.39±4.68	25.13±4.00	**55.60**±0.53	29.23±5.98	28.34±2.53	55.28±9.25	**14.45**±8.52	42.41±3.26	24.62±1.32
Pretraining: ContextPred									
Full fine-tuning	76.70±3.54	**22.98**±1.64	55.17±0.79	33.45±3.28	30.42±2.94	59.50±10.37	**8.79**±2.00	48.76±0.76	**23.86**±1.43
All-in-One	76.09±1.11	21.20±2.31	52.65±0.83	30.19±5.26	**31.89**±2.10	56.22±12.80	96.94±0.16	42.49±3.37	22.52±1.15
EdgePrompt	69.11±7.34	17.68±1.30	**57.38**±1.50	38.17±6.37	30.65±4.74	51.44±14.74	107.31±0.19	49.35±1.10	22.15±2.17
EdgePrompt+	69.56±3.97	15.45±2.48	55.21±1.29	**38.80**±5.58	30.52±3.75	56.88±13.31	107.35±0.13	49.65±0.61	21.31±2.18
GPF	78.85±0.63	19.80±1.72	56.91±0.93	37.88±5.08	29.92±3.43	55.86±12.45	107.31±0.37	**49.99**±1.16	20.15±2.74
GPF+	78.85±1.18	17.20±2.43	54.36±0.60	37.32±4.45	29.92±3.43	52.18±11.82	107.69±0.93	49.32±1.23	18.28±3.16
GPPT	76.17±1.93	12.01±4.72	50.87±1.74	31.85±2.14	30.18±2.08	53.75±10.93	101.58±3.74	49.52±1.08	15.12±1.21
GraphPrompt	77.56±1.63	9.40±1.64	51.99±0.86	28.48±5.32	25.63±6.32	57.41±11.69	77.86±1.06	42.64±2.90	20.15±1.27
GraphPrompt+	74.01±3.90	7.32±1.84	51.72±0.64	34.36±2.53	23.94±3.79	55.21±8.57	62.37±10.61	42.52±2.92	18.39±2.83
ProNoG	**79.42**±1.95	15.43±2.76	56.29±0.53	28.80±2.17	28.40±3.03	**61.66**±11.25	13.12±3.81	42.84±2.52	19.46±1.62
Pretraining: DGI									
Full fine-tuning	79.02±1.72	24.57±1.20	55.53±1.00	32.65±3.28	31.19±4.24	52.63±12.21	**6.16**±0.27	49.94±0.87	**26.24**±1.13
All-in-One	76.22±2.52	22.67±0.68	52.82±0.81	30.99±4.69	32.46±2.01	55.54±9.25	47.88±5.38	**52.54**±1.16	20.60±4.42
EdgePrompt	71.30±4.02	24.68±0.75	55.81±1.29	33.44±6.06	34.71±3.01	51.76±7.21	97.14±0.73	51.14±1.95	17.19±3.65
EdgePrompt+	70.41±2.58	23.22±0.87	54.51±0.88	32.55±7.02	**35.39**±2.10	**58.59**±8.43	97.02±0.73	51.57±2.50	21.39±3.42
GPF	78.12±2.32	24.02±0.93	55.78±2.16	**34.41**±2.47	34.95±2.84	47.75±10.83	100.32±1.07	51.39±1.16	20.19±1.58
GPF+	**79.11**±3.03	19.77±9.61	54.42±0.65	33.45±3.48	34.40±2.46	48.81±8.16	101.06±2.04	51.24±0.45	18.93±6.12
GPPT	44.42±11.79	14.15±6.94	49.85±0.57	31.25±2.40	29.29±2.94	49.54±6.47	80.94±4.37	49.19±1.42	18.80±2.40
GraphPrompt	72.20±7.04	7.58±0.94	52.55±0.93	30.34±2.69	23.83±4.18	58.25±13.10	44.47±1.01	42.95±2.87	23.86±1.42
GraphPrompt+	68.32±2.99	9.45±0.85	52.95±1.35	29.01±4.05	23.24±5.61	56.27±6.62	25.02±4.18	42.81±2.33	22.42±1.66
ProNoG	74.11±5.64	**25.12**±2.70	**56.61**±0.42	24.41±3.22	30.12±1.69	54.54±8.81	21.97±7.51	42.98±3.09	25.52±1.03
Pretraining: EdgePred									
Full fine-tuning	**64.42**±22.74	**22.87**±1.19	54.25±0.64	33.06±3.33	29.31±3.03	57.72±12.84	**7.27**±1.37	48.89±0.89	23.93±2.15
All-in-One	30.46±2.61	7.12±0.97	51.78±0.99	32.91±5.06	**29.33**±3.21	54.80±9.00	96.94±0.16	48.18±1.55	21.65±2.70
EdgePrompt	38.53±6.09	5.09±2.43	50.95±0.49	38.12±3.98	28.58±1.44	51.96±10.20	109.83±3.84	50.37±0.55	20.85±2.36
EdgePrompt+	53.82±10.12	6.39±1.57	50.99±0.87	**38.63**±5.32	29.32±1.99	55.84±10.01	107.35±0.13	50.14±0.80	23.98±1.85
GPF	35.39±3.48	6.80±2.49	51.21±0.53	37.71±4.32	28.98±4.22	53.60±9.89	107.31±0.37	50.36±1.17	19.86±3.59
GPF+	54.48±2.35	11.65±2.02	52.23±0.67	37.16±3.62	29.08±3.68	54.96±10.10	107.30±0.12	**50.54**±0.79	21.39±3.50
GPPT	33.90±1.11	2.28±0.48	49.95±0.32	32.44±1.81	28.58±2.55	49.32±14.50	101.58±3.74	50.00±0.65	17.99±2.28
GraphPrompt	25.28±1.29	3.99±0.80	49.43±0.61	33.13±4.11	25.09±5.28	51.99±11.19	77.86±1.06	44.69±1.91	**25.76**±0.93
GraphPrompt+	41.25±7.00	3.77±0.96	49.76±0.94	34.65±4.09	25.70±3.95	51.49±9.32	62.37±10.61	42.42±3.02	20.10±2.98
ProNoG	29.26±4.06	5.19±0.63	**54.29**±0.78	29.27±4.21	28.31±1.85	**59.08**±10.42	13.12±3.81	43.88±2.41	23.26±1.38
Pretraining: GraphCL									
Full fine-tuning	75.10±3.59	25.07±1.05	54.67±0.72	33.61±4.20	28.50±2.88	**58.10**±7.87	**7.50**±2.25	48.50±0.91	25.75±0.60
All-in-One	70.74±2.21	24.10±1.55	53.71±0.40	32.24±4.27	29.44±1.65	55.64±9.73	91.21±0.21	**52.43**±1.16	26.89±1.30
EdgePrompt	68.03±4.96	26.03±0.47	53.52±0.74	35.73±5.43	**30.92**±2.56	51.31±7.76	109.62±0.12	48.25±2.05	22.68±1.81
EdgePrompt+	68.38±4.80	24.90±1.39	52.76±1.11	34.09±4.47	30.75±2.55	52.20±7.35	109.56±0.09	48.56±1.23	23.75±1.52
GPF	73.69±3.19	**27.73**±1.12	51.73±0.83	**38.67**±4.85	29.46±2.19	55.48±4.92	109.83±0.14	48.06±1.96	21.01±2.81
GPF+	73.61±1.84	22.42±5.75	52.20±0.60	37.32±4.72	29.46±2.19	54.24±9.96	109.82±0.10	48.31±1.19	20.88±1.88
GPPT	62.52±3.17	18.26±2.34	50.37±1.42	33.62±2.87	28.75±3.37	51.05±12.57	106.96±2.55	49.72±0.76	19.20±2.37
GraphPrompt	67.13±1.39	12.78±1.43	53.06±0.48	29.03±2.40	23.12±4.22	53.82±6.30	92.81±0.64	43.95±2.66	**28.16**±2.70
GraphPrompt+	**77.22**±2.58	11.86±1.45	53.04±0.60	34.58±4.48	23.98±4.25	50.65±4.48	50.38±18.61	42.67±2.79	20.28±3.22
ProNoG	65.97±1.90	20.70±1.48	**57.01**±0.60	31.21±4.15	27.83±3.28	52.22±9.46	18.78±24.14	46.45±1.94	25.18±1.66
Pretraining: InfoGraph									
Full fine-tuning	**76.99**±3.75	**25.09**±1.34	55.39±0.82	33.20±3.90	29.88±3.57	57.19±9.37	**6.38**±0.50	50.22±0.76	21.62±2.50
All-in-One	63.63±6.80	22.26±2.23	52.66±0.73	32.21±2.72	30.64±2.61	56.43±10.90	105.34±0.15	**57.03**±2.42	11.84±1.71
EdgePrompt	62.26±5.73	18.30±2.00	55.84±1.79	36.85±5.34	31.63±1.77	45.48±6.15	114.05±0.18	51.00±1.34	21.66±2.89
EdgePrompt+	69.15±6.24	18.53±1.83	54.48±1.05	37.83±5.02	32.85±1.21	51.83±13.17	113.68±0.25	51.07±1.17	21.76±2.65
GPF	75.90±2.80	19.33±1.15	55.47±1.00	**37.95**±3.33	31.00±3.17	53.86±6.31	115.28±0.12	50.46±1.55	13.99±4.67
GPF+	76.62±3.88	19.17±1.27	52.68±0.91	36.91±2.83	32.82±1.01	52.37±6.54	115.23±0.08	49.59±1.00	16.85±4.97
GPPT	62.64±0.94	10.61±2.03	49.89±0.38	32.62±2.12	**33.72**±2.27	50.80±9.56	111.89±0.94	50.02±0.23	13.40±1.97
GraphPrompt	57.86±2.41	7.12±0.95	52.74±0.91	31.30±1.56	25.22±4.90	59.67±10.75	111.78±2.61	43.75±2.33	18.84±0.74
GraphPrompt+	68.66±4.54	8.67±2.42	53.50±0.99	35.35±4.36	25.22±4.90	50.82±6.93	81.15±21.78	43.16±3.18	**22.24**±3.28
ProNoG	63.12±1.25	19.94±1.67	**56.68**±0.54	29.13±2.09	30.57±3.07	**59.87**±9.97	8.83±3.31	43.81±2.48	14.77±0.60

Table 23: 5-shot configuration results for NodeFormer across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	69.96±5.09	5.62±2.50	**59.91**±3.28	24.92±0.19	18.91±2.48	49.89±15.96	50.29±32.58	50.21±0.32	13.81±2.50
All-in-One	68.58±2.85	15.36±2.28	51.80±3.86	31.15±4.74	28.83±5.16	**74.63**±14.46	103.55±0.62	44.77±2.44	11.29±0.60
EdgePrompt	65.36±6.62	4.96±2.55	55.29±3.99	24.94±0.04	23.40±2.92	44.04±10.54	113.00±0.80	50.71±0.24	16.70±2.93
EdgePrompt+	64.60±5.09	4.65±2.13	57.16±3.77	24.96±0.00	23.06±1.91	59.60±17.39	113.03±0.87	50.66±0.31	15.78±4.58
GPF	68.06±3.49	4.25±2.83	52.43±3.15	24.15±1.51	20.91±2.15	53.44±17.96	113.07±0.44	50.74±0.30	10.21±0.67
GPF+	68.38±2.94	3.17±2.24	54.44±1.83	25.23±1.12	20.76±2.83	51.53±18.98	113.00±0.70	**50.81**±0.30	12.69±4.84
GPPT	70.32±5.29	4.07±1.17	50.85±4.52	35.03±2.05	28.00±4.54	60.08±21.39	108.25±0.61	50.20±0.60	14.13±3.49
GraphPrompt	62.81±6.46	5.53±1.06	52.18±4.51	35.13±3.97	23.66±4.01	53.44±10.44	104.92±0.68	42.79±2.70	**25.07**±2.40
GraphPrompt+	45.99±12.24	3.29±0.90	53.43±3.57	**35.27**±3.92	22.25±2.23	53.72±16.38	82.24±4.45	43.23±3.09	23.12±2.64
ProNoG	**72.53**±5.52	**25.31**±3.82	53.51±0.99	30.92±5.63	**29.02**±3.76	58.32±18.03	**8.78**±3.85	42.93±2.69	24.46±1.36
Pretraining: ContextPred									
Full fine-tuning	70.83±4.17	**15.57**±8.04	**59.55**±3.81	24.91±0.13	25.33±4.16	49.66±10.58	51.23±32.45	50.38±0.46	16.33±4.41
All-in-One	**78.90**±2.98	2.82±2.59	52.32±2.02	30.32±2.11	28.89±2.88	52.49±8.79	109.95±0.31	**51.21**±2.13	19.19±1.58
EdgePrompt	69.29±8.48	3.96±3.56	54.07±1.56	25.06±0.07	27.42±3.71	**57.27**±15.86	113.38±1.08	49.29±0.37	14.22±3.89
EdgePrompt+	68.84±8.03	6.72±4.88	54.55±5.50	25.13±0.22	26.83±3.98	50.58±8.78	113.34±1.16	49.34±0.21	14.21±4.08
GPF	77.56±3.66	6.47±3.68	50.78±2.54	24.99±0.13	22.67±2.07	54.03±11.41	113.85±1.04	49.91±0.23	13.73±2.78
GPF+	78.83±2.29	8.08±4.54	52.85±2.68	25.32±1.13	28.24±4.44	53.20±11.82	113.85±1.05	49.97±0.24	13.74±4.29
GPPT	74.95±3.46	3.34±3.59	55.35±2.88	33.20±2.03	**29.38**±3.84	48.78±10.45	110.43±1.34	49.98±1.31	19.56±1.23
GraphPrompt	76.85±1.65	3.07±0.96	54.38±3.72	35.04±3.54	24.15±3.96	43.45±3.06	109.38±2.70	43.15±2.63	20.59±2.11
GraphPrompt+	75.09±3.53	4.55±1.40	54.96±5.03	**35.37**±4.17	24.58±3.75	50.06±4.41	99.49±0.45	43.00±2.77	19.44±3.45
ProNoG	75.50±0.58	13.45±0.90	56.41±1.14	27.56±7.81	27.11±1.51	43.39±5.76	**15.12**±1.72	42.77±2.29	**21.59**±0.96
Pretraining: DGI									
Full fine-tuning	54.81±14.81	15.44±4.76	52.30±2.43	24.96±0.12	22.67±2.71	52.70±25.45	50.85±32.68	49.77±0.14	9.83±0.57
All-in-One	**71.79**±7.63	**20.54**±1.66	50.85±4.36	29.88±1.24	**32.24**±2.93	**58.95**±31.37	107.01±0.13	45.05±2.32	9.97±0.42
EdgePrompt	36.92±6.69	5.38±4.62	52.53±2.16	24.97±0.07	23.72±3.32	53.20±10.45	111.54±0.07	50.79±0.38	9.69±0.35
EdgePrompt+	45.97±13.89	5.79±4.68	53.12±2.11	25.01±0.05	23.69±3.23	52.89±29.47	111.55±0.12	50.53±0.92	9.85±0.00
GPF	23.76±12.05	3.58±3.14	51.68±2.73	24.87±0.55	21.33±2.86	43.19±27.00	111.54±0.07	**51.32**±0.78	9.60±0.47
GPF+	27.73±13.66	9.00±4.71	52.38±2.72	26.73±3.41	25.25±1.92	54.93±26.49	111.54±0.07	51.22±0.65	9.85±0.11
GPPT	68.07±7.73	7.83±4.43	50.38±1.79	34.75±2.29	30.75±2.27	49.16±29.09	107.80±0.07	49.92±1.09	10.54±0.42
GraphPrompt	61.40±10.07	3.62±1.12	49.13±2.46	35.08±3.91	22.02±2.26	49.61±17.32	105.05±0.24	42.73±2.79	14.96±2.83
GraphPrompt+	70.16±6.55	8.71±2.04	51.72±2.76	**35.61**±4.09	23.16±2.90	53.56±4.39	94.72±2.95	42.79±2.75	**15.55**±3.24
ProNoG	59.36±3.49	12.64±3.35	**56.01**±1.35	24.32±4.51	25.84±3.90	56.28±13.46	**15.99**±2.06	42.92±2.99	14.16±1.07
Pretraining: EdgePred									
Full fine-tuning	**39.37**±12.26	**18.04**±2.53	56.63±4.11	24.94±0.93	23.16±5.01	42.81±16.12	55.63±46.90	49.38±0.86	10.49±0.74
All-in-One	17.97±4.74	4.69±2.49	52.50±4.59	26.94±2.93	25.40±4.16	41.52±20.37	109.88±0.24	**54.68**±2.01	19.99±2.09
EdgePrompt	8.18±2.31	3.22±2.80	50.84±1.36	26.00±0.86	20.68±0.71	51.34±10.44	112.87±0.11	51.85±0.69	9.86±0.58
EdgePrompt+	12.20±3.81	1.56±0.94	51.08±0.80	26.15±1.33	20.69±1.04	45.61±11.27	112.82±0.13	51.99±0.93	9.28±0.51
GPF	14.43±4.21	1.72±1.85	49.98±0.69	24.67±0.48	22.58±3.10	49.18±16.53	113.38±0.09	52.42±1.42	11.90±4.20
GPF+	17.12±6.05	3.32±4.78	50.66±3.05	24.39±1.10	25.19±3.49	47.32±11.69	113.37±0.09	51.92±1.48	9.41±0.48
GPPT	16.69±7.41	3.59±4.42	50.59±1.81	32.89±2.79	**27.05**±4.52	49.99±11.95	109.84±0.07	50.35±0.74	9.92±0.00
GraphPrompt	20.10±2.87	2.85±0.81	51.07±1.76	**33.88**±2.93	22.71±2.03	47.92±8.45	108.16±0.13	43.23±2.23	**20.29**±1.88
GraphPrompt+	19.36±4.74	3.01±1.15	53.47±1.13	30.85±5.46	23.05±2.65	54.79±8.91	99.49±0.45	43.15±2.86	15.55±3.30
ProNoG	18.77±3.05	3.80±0.58	**56.98**±1.20	28.27±2.78	25.19±2.25	**57.91**±10.31	**15.04**±1.74	43.41±2.57	18.81±1.87
Pretraining: GraphCL									
Full fine-tuning	**70.22**±6.65	9.73±5.77	52.24±2.07	24.97±0.04	28.06±3.46	52.15±17.67	49.08±33.56	49.21±1.39	9.87±0.03
All-in-One	62.77±3.02	**20.05**±1.52	50.74±2.19	26.05±1.72	**29.98**±3.96	**61.74**±22.39	111.31±0.06	**52.89**±1.86	9.38±0.34
EdgePrompt	44.88±7.81	9.49±0.90	51.95±2.93	25.37±0.92	27.92±3.93	53.41±10.88	114.98±0.03	49.77±1.10	9.57±0.47
EdgePrompt+	50.24±7.04	10.32±1.66	50.65±3.11	24.96±0.00	28.22±4.54	52.98±12.98	114.94±0.06	49.79±1.11	9.52±0.42
GPF	58.84±3.88	11.64±3.24	52.01±1.57	25.09±2.04	23.88±4.95	53.34±16.81	115.11±0.03	49.46±1.00	10.07±0.17
GPF+	59.90±4.42	11.95±3.55	48.19±4.77	25.45±1.04	27.50±4.81	51.03±13.40	115.11±0.03	49.81±1.21	10.07±0.17
GPPT	64.03±2.66	10.11±1.49	51.32±1.38	35.01±2.04	28.99±4.09	49.37±21.86	111.79±0.04	49.71±0.66	9.94±0.05
GraphPrompt	59.55±2.85	7.01±1.33	49.42±2.12	35.20±3.95	24.25±4.02	52.58±12.91	111.29±0.26	47.38±1.93	**15.06**±3.38
GraphPrompt+	67.64±4.19	3.58±0.94	51.25±2.98	**35.28**±3.95	22.06±2.43	49.96±7.70	100.15±0.50	44.65±2.66	14.76±3.13
ProNoG	66.09±2.18	19.12±1.71	**53.10**±0.56	30.31±6.74	27.06±2.60	48.37±10.58	**14.77**±1.79	48.06±0.75	13.89±1.77
Pretraining: InfoGraph									
Full fine-tuning	61.85±9.16	2.67±1.13	52.36±3.63	24.99±0.14	24.09±6.43	45.07±14.86	52.46±16.50	50.52±0.65	10.03±0.41
All-in-One	64.34±7.35	3.24±2.54	52.29±0.62	32.09±5.38	**25.75**±3.46	**60.29**±11.79	110.44±0.13	**55.36**±2.45	9.67±0.55
EdgePrompt	48.11±10.27	6.54±5.76	50.98±0.62	24.99±0.30	21.79±2.93	47.13±15.34	113.99±0.07	50.20±1.18	9.40±0.46
EdgePrompt+	49.63±8.45	6.01±6.12	50.97±0.62	25.06±0.29	21.94±3.62	50.84±11.13	114.04±0.03	50.13±1.21	9.75±0.38
GPF	60.87±4.20	2.98±3.40	52.40±2.72	25.08±0.23	21.41±2.40	47.09±13.77	114.34±0.02	50.30±0.87	9.74±0.38
GPF+	**65.42**±5.01	3.02±2.38	**53.87**±2.31	23.79±1.98	23.13±2.71	49.70±12.50	114.34±0.02	50.89±1.75	10.20±1.18
GPPT	55.51±8.95	3.84±3.86	49.71±0.39	34.36±2.54	25.69±2.16	52.15±13.21	111.00±0.07	49.98±0.87	9.75±0.38
GraphPrompt	62.75±3.93	2.49±0.97	52.19±2.42	**35.30**±3.96	22.91±2.39	56.27±5.75	110.14±0.22	43.44±2.72	15.01±3.06
GraphPrompt+	59.42±6.99	2.39±0.97	53.52±1.77	35.28±3.88	22.50±4.17	54.29±9.44	100.83±1.11	42.82±2.68	**15.42**±3.08
ProNoG	48.39±3.19	**13.38**±2.21	51.35±1.32	29.42±6.46	22.37±3.22	47.50±8.84	**14.70**±1.65	44.34±2.16	14.83±1.39

Table 24: 100-shot configuration results for GCN across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	87.75±0.63	37.84±0.46	**55.10**±0.29	40.03±3.77	33.79±1.16	45.35±6.30	40.11±48.15	50.34±1.30	25.61±8.79
All-in-One	87.92±0.57	32.11±3.41	51.26±0.47	39.92±0.51	36.78±1.18	52.01±6.25	88.08±2.16	**57.51**±1.51	22.97±0.81
EdgePrompt	82.32±0.87	**44.05**±0.73	54.08±0.28	39.34±0.94	**39.68**±1.17	**55.17**±6.38	80.98±0.22	52.32±0.63	25.20±3.19
EdgePrompt+	84.17±1.17	39.40±0.73	53.87±0.35	**45.52**±2.13	38.22±1.25	52.94±12.37	80.43±0.86	52.47±0.64	28.78±1.41
GPF	**89.11**±0.33	43.64±0.61	54.12±0.32	39.62±1.91	39.49±0.97	46.63±10.31	104.80±0.57	52.07±0.95	24.05±3.06
GPF+	87.36±0.43	41.38±0.54	53.85±0.26	41.80±2.24	37.88±1.32	49.11±7.94	104.42±0.43	52.53±0.34	27.87±0.53
GPPT	88.21±0.67	25.28±2.97	51.38±1.38	33.06±5.94	33.13±4.04	48.69±9.40	113.64±1.78	50.02±1.92	19.18±1.16
GraphPrompt	81.10±0.31	19.58±0.96	53.54±0.46	38.66±0.99	28.75±0.51	48.13±13.08	20.46±0.37	44.45±1.41	**30.47**±1.12
GraphPrompt+	85.25±1.30	19.29±0.81	54.12±0.30	37.92±1.09	26.13±1.10	48.06±8.81	18.33±1.33	42.67±1.16	25.51±2.42
ProNoG	86.26±0.30	39.08±0.51	54.62±0.82	32.99±1.92	31.21±1.30	48.70±9.25	**7.92**±2.02	42.85±1.41	26.59±0.42
Pretraining: ContextPred									
Full fine-tuning	87.57±0.50	37.75±0.48	**54.94**±0.29	40.71±1.70	35.00±0.86	49.46±12.30	43.34±45.36	52.18±0.84	24.33±7.46
All-in-One	86.01±0.51	16.42±0.70	52.00±0.45	39.97±0.64	37.05±0.79	52.68±7.88	98.92±2.70	50.83±1.41	28.81±0.55
EdgePrompt	81.43±1.78	43.08±0.65	53.31±0.66	40.48±1.39	**38.78**±1.83	**57.36**±5.48	93.07±0.43	52.21±0.70	28.11±3.09
EdgePrompt+	83.66±2.43	39.22±1.09	53.62±0.33	**46.94**±2.55	37.63±1.30	50.54±6.03	92.62±0.56	**52.30**±1.56	**30.73**±1.11
GPF	**88.33**±0.42	**43.20**±0.94	52.89±0.44	41.59±1.07	38.59±0.94	45.04±11.15	105.54±0.41	52.02±0.71	28.88±0.90
GPF+	86.60±0.41	41.93±0.76	53.44±0.49	41.32±1.78	37.55±0.77	46.54±5.98	105.23±0.35	51.89±1.14	30.46±0.75
GPPT	86.83±0.57	29.30±3.73	51.09±0.78	35.70±1.01	36.26±1.88	55.22±9.88	113.94±1.87	50.25±0.61	20.76±1.20
GraphPrompt	86.12±0.32	20.01±0.62	52.48±0.53	38.10±2.25	29.29±1.50	50.75±6.57	20.60±0.50	47.17±1.26	29.99±0.36
GraphPrompt+	85.62±0.53	18.90±2.23	52.94±0.68	37.95±1.71	28.29±2.27	56.93±7.07	19.70±0.19	44.58±1.21	27.20±1.88
ProNoG	85.76±0.70	36.49±0.47	53.49±0.61	35.16±1.26	30.47±2.28	53.60±11.50	**10.31**±1.91	44.75±1.25	26.28±0.71
Pretraining: DGI									
Full fine-tuning	87.67±0.60	33.24±0.45	**55.08**±0.63	42.28±1.40	35.40±1.16	44.49±8.67	**5.36**±0.20	44.94±2.04	29.26±1.05
All-in-One	86.54±0.39	31.59±2.09	51.52±0.66	40.23±1.17	38.05±0.92	51.94±12.93	80.60±4.12	**53.77**±0.86	26.91±2.43
EdgePrompt	80.73±2.98	30.50±2.95	54.45±0.44	38.66±0.84	**41.06**±0.87	**58.48**±9.58	86.04±0.88	51.04±0.67	25.23±3.45
EdgePrompt+	83.61±1.70	32.85±1.23	53.87±0.72	**45.62**±1.61	40.43±0.99	52.82±15.03	85.22±0.73	51.20±0.51	30.93±1.30
GPF	**88.30**±0.36	30.63±2.93	52.94±0.47	40.78±2.68	40.14±1.14	51.72±10.83	97.76±0.72	51.02±0.57	25.51±2.80
GPF+	86.60±0.53	30.08±2.71	53.75±0.54	40.00±1.42	38.27±0.57	50.73±10.30	97.16±0.69	51.20±0.61	29.29±1.61
GPPT	84.54±0.78	36.01±0.92	49.86±0.16	34.51±3.83	36.68±0.87	50.20±12.22	104.84±8.45	50.18±1.08	23.22±1.43
GraphPrompt	81.67±0.30	19.03±1.38	52.84±0.38	35.27±1.62	28.49±1.41	52.44±12.14	18.38±0.08	44.37±1.01	**31.80**±0.56
GraphPrompt+	86.25±0.41	16.37±2.01	53.31±0.28	38.33±2.93	29.66±2.28	54.70±9.31	18.58±0.28	44.86±1.20	27.48±2.88
ProNoG	79.23±0.57	**37.97**±0.35	52.86±0.83	34.66±1.71	31.22±2.70	51.99±8.68	9.63±1.89	45.04±0.72	27.46±0.46
Pretraining: EdgePred									
Full fine-tuning	**87.17**±0.40	**38.18**±0.78	**54.68**±1.16	41.62±1.42	35.39±0.93	42.09±7.79	**5.39**±0.19	45.94±2.84	29.26±1.05
All-in-One	60.67±0.78	12.22±1.04	50.90±0.52	43.34±0.86	37.52±1.53	48.80±8.89	98.92±2.70	**60.96**±1.64	24.97±1.76
EdgePrompt	73.82±3.73	28.97±1.33	53.82±1.17	39.67±1.11	37.15±0.73	47.54±5.69	93.07±0.43	52.65±0.89	27.04±2.37
EdgePrompt+	83.56±1.24	32.70±1.34	53.49±0.41	**46.96**±2.17	**38.59**±0.79	48.51±7.03	92.62±0.56	52.88±1.01	30.34±1.00
GPF	74.90±2.71	26.77±2.01	51.57±1.31	40.53±1.90	37.48±0.99	40.02±12.53	105.54±0.41	52.29±0.97	23.70±3.66
GPF+	86.29±0.86	29.57±1.91	53.46±0.49	41.57±0.48	38.47±1.46	41.14±6.39	105.23±0.35	50.33±2.36	29.05±0.66
GPPT	32.33±2.73	3.30±0.28	51.21±0.93	35.62±1.76	30.96±1.34	**53.76**±10.85	113.94±1.87	49.99±0.23	23.22±1.43
GraphPrompt	37.21±1.49	10.73±0.52	50.68±0.62	34.99±1.22	29.27±3.01	44.59±6.71	39.54±42.82	46.17±1.42	**31.86**±0.39
GraphPrompt+	81.55±1.35	9.64±2.64	53.15±1.40	37.90±2.31	28.46±1.59	46.83±7.11	19.70±0.19	43.33±1.45	23.95±0.66
ProNoG	33.02±0.48	14.54±0.54	52.17±0.86	37.47±1.76	31.23±1.25	50.77±12.47	10.31±1.91	41.53±1.20	26.13±0.66
Pretraining: GraphCL									
Full fine-tuning	87.66±0.70	37.30±0.83	**55.23**±0.42	39.90±1.91	34.80±1.03	41.43±6.45	5.67±0.11	50.82±1.52	**29.76**±0.49
All-in-One	86.78±0.55	14.33±1.18	51.72±0.64	42.28±1.31	36.43±0.76	49.27±9.30	111.92±1.17	**56.66**±1.66	28.01±0.73
EdgePrompt	82.92±0.52	**37.89**±2.92	53.30±0.32	39.27±0.67	**39.07**±0.92	50.49±8.49	69.30±0.53	50.91±0.98	28.10±0.32
EdgePrompt+	84.06±2.19	33.56±1.03	53.62±0.33	**46.28**±2.27	38.31±1.91	46.85±8.94	63.53±1.01	52.15±1.07	28.76±0.79
GPF	**88.75**±0.50	35.56±1.64	53.25±0.19	41.47±2.44	37.57±1.22	42.24±8.04	114.92±1.24	51.02±1.36	28.01±0.30
GPF+	87.56±0.70	35.36±2.34	53.57±0.31	39.80±1.95	36.78±1.18	46.90±7.83	113.95±0.94	51.22±1.10	29.28±0.44
GPPT	87.82±1.03	29.86±0.64	50.24±0.94	32.84±1.97	37.12±0.79	50.11±12.69	116.83±0.24	50.09±0.28	19.34±0.98
GraphPrompt	84.32±0.50	20.28±0.23	53.15±0.45	36.46±1.48	26.15±1.98	41.40±5.76	102.03±0.35	45.75±2.03	29.10±0.81
GraphPrompt+	87.38±0.60	20.32±0.22	53.75±0.45	37.47±0.66	29.34±1.14	48.54±6.20	26.21±9.61	44.92±1.39	27.46±0.65
ProNoG	86.81±0.28	30.46±0.49	54.45±0.78	36.18±2.12	29.17±1.38	**51.84**±11.29	**5.37**±1.14	43.71±1.36	25.76±0.68
Pretraining: InfoGraph									
Full fine-tuning	87.19±0.75	**38.25**±1.11	**54.83**±0.34	41.09±1.77	34.59±1.22	44.38±10.86	**5.26**±0.07	50.42±1.75	29.32±0.57
All-in-One	74.86±0.55	19.65±8.10	51.88±0.78	43.34±0.86	37.91±0.73	50.51±6.17	112.04±1.15	**60.96**±1.64	24.97±1.76
EdgePrompt	79.45±1.07	36.47±0.57	52.76±0.37	39.67±1.11	**39.74**±0.82	**53.91**±3.31	96.44±1.28	52.65±0.89	27.04±2.37
EdgePrompt+	83.60±1.80	36.33±0.10	53.48±0.49	**46.96**±2.17	38.87±0.70	49.54±6.28	95.50±0.52	52.88±1.01	30.34±1.00
GPF	87.38±0.27	34.92±0.68	52.42±0.22	39.90±1.09	39.09±0.60	42.52±9.00	115.53±0.86	52.54±0.97	25.84±3.26
GPF+	**87.50**±0.50	33.72±1.15	53.41±0.25	39.87±1.64	37.82±0.41	42.71±6.92	114.90±0.98	52.77±1.00	30.25±0.62
GPPT	73.22±1.01	5.51±0.34	51.17±1.34	35.62±1.76	36.20±1.01	49.54±7.24	116.78±0.27	49.99±0.23	20.04±0.70
GraphPrompt	74.63±0.75	14.20±1.05	52.02±0.34	36.03±0.69	30.07±2.14	49.27±8.11	99.37±0.56	46.60±2.07	**31.97**±0.46
GraphPrompt+	85.65±0.86	15.07±2.42	52.51±0.23	36.99±1.39	27.80±1.31	47.66±7.65	20.30±1.06	44.15±1.35	30.09±3.23
ProNoG	76.77±1.03	20.33±0.70	53.68±0.63	34.76±1.49	34.39±1.37	50.74±10.10	8.04±1.46	41.86±1.30	27.28±0.45

Table 25: 100-shot configuration results for GAT across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	87.29±0.75	39.88±0.64	54.85±0.26	34.84±1.67	36.11±1.05	51.75±10.96	24.94±6.90	52.87±1.01	**38.93**±1.24
All-in-One	**87.79**±0.37	34.79±0.99	51.62±0.56	36.86±2.17	32.04±0.92	53.89±9.06	97.19±0.61	**57.64**±0.56	18.14±8.00
EdgePrompt	75.92±5.10	**44.89**±0.53	52.91±0.25	26.48±1.74	34.30±1.54	52.46±6.44	98.86±0.75	50.69±0.59	18.40±3.79
EdgePrompt+	77.56±3.95	40.84±1.07	53.58±0.33	**39.47**±1.10	**36.80**±0.74	**57.46**±3.94	98.30±0.89	52.22±1.35	33.82±3.02
GPF	87.10±1.30	44.84±0.61	53.46±0.21	28.81±2.34	34.33±0.68	52.77±9.96	109.48±0.43	50.72±0.94	22.65±4.37
GPF+	86.41±1.29	42.64±1.06	53.61±0.19	33.59±0.90	36.70±1.85	51.80±8.26	109.45±0.57	51.12±0.60	23.51±6.44
GPPT	87.74±0.62	24.18±2.39	50.21±1.26	27.67±4.43	33.47±1.92	47.18±8.04	113.77±2.16	49.75±0.71	20.50±2.86
GraphPrompt	83.95±0.76	20.56±0.95	53.50±0.18	38.53±0.86	29.08±1.53	51.11±9.68	24.26±0.52	42.94±1.50	34.40±0.45
GraphPrompt+	84.41±1.07	20.49±1.51	53.54±0.18	37.92±1.09	28.34±2.18	43.40±10.14	23.65±3.69	42.40±1.08	35.46±0.90
ProNoG	87.40±0.47	38.38±0.62	**55.08**±0.52	30.20±0.99	30.79±1.87	50.51±6.83	**11.32**±1.69	41.43±1.28	31.31±0.75
Pretraining: ContextPred									
Full fine-tuning	**87.76**±0.76	39.36±0.59	55.27±0.38	34.89±2.30	**36.20**±1.28	57.48±15.44	20.94±1.19	52.56±1.13	35.54±3.52
All-in-One	85.97±0.65	35.07±0.64	50.69±0.46	34.03±1.14	31.04±1.06	53.32±8.03	113.94±0.71	50.40±2.18	35.75±2.58
EdgePrompt	77.43±1.44	43.16±0.77	52.77±0.32	39.39±0.94	32.79±0.80	53.22±14.90	110.30±0.48	51.85±0.82	33.69±3.79
EdgePrompt+	84.31±1.01	39.85±0.67	53.65±0.51	**40.00**±1.06	**36.20**±1.39	50.89±12.96	110.13±0.84	**52.73**±1.14	**38.82**±1.27
GPF	87.22±0.43	**43.93**±0.60	53.15±0.30	30.84±0.96	32.68±0.43	54.77±15.41	116.00±0.49	51.77±1.21	34.22±4.13
GPF+	86.71±1.14	41.15±2.30	53.94±0.24	32.91±1.97	35.82±0.54	52.96±10.66	116.19±0.53	51.98±0.92	35.07±3.84
GPPT	86.06±0.85	34.53±3.68	50.06±0.14	35.06±1.33	31.52±2.23	47.41±7.74	116.95±0.14	49.33±2.95	27.21±0.56
GraphPrompt	82.49±0.82	20.82±0.38	52.85±0.43	37.80±2.05	27.69±1.43	**59.50**±15.78	108.35±0.41	45.01±1.58	34.84±0.34
GraphPrompt+	84.36±1.39	21.57±0.81	53.11±0.40	38.63±1.16	27.62±2.41	49.58±6.54	105.19±1.39	45.41±1.68	35.55±0.84
ProNoG	82.30±1.14	35.78±0.60	**55.35**±0.63	26.63±2.45	29.85±0.80	51.56±14.48	**10.54**±0.90	44.25±1.29	30.97±0.24
Pretraining: DGI									
Full fine-tuning	**87.23**±0.96	36.22±0.54	**55.25**±0.47	32.13±1.86	38.40±0.91	53.75±10.24	27.83±4.37	**51.43**±1.32	38.82±1.41
All-in-One	82.67±0.88	32.65±2.17	52.00±0.46	34.25±6.58	37.93±1.01	55.58±16.84	107.98±0.46	50.05±1.11	22.30±1.17
EdgePrompt	76.24±1.81	35.20±3.42	53.13±0.15	30.53±6.29	38.62±2.16	42.50±16.42	78.37±0.85	51.37±1.60	25.10±6.77
EdgePrompt+	80.97±2.64	36.17±1.10	53.85±0.33	**40.71**±0.46	37.75±2.07	50.77±8.32	71.91±0.24	50.79±1.64	**39.20**±1.01
GPF	82.89±3.11	34.12±4.80	52.74±0.64	27.92±3.64	38.65±1.07	47.37±18.05	113.88±1.09	50.89±0.85	29.24±3.69
GPF+	82.20±0.74	32.78±2.71	53.50±0.16	33.57±2.88	**39.10**±0.94	42.83±12.00	113.91±1.17	50.54±0.25	32.42±2.34
GPPT	69.92±8.95	36.06±0.60	50.11±1.52	32.03±1.50	35.36±2.06	54.03±5.04	116.55±0.47	50.13±0.22	24.92±2.82
GraphPrompt	67.62±0.27	15.49±2.25	52.99±0.58	31.80±0.49	31.11±4.07	52.20±10.85	92.94±0.57	40.37±0.99	38.30±0.52
GraphPrompt+	76.37±2.79	17.59±0.84	53.47±0.56	38.41±3.04	30.73±3.79	52.01±17.13	84.89±2.52	41.02±0.79	38.07±0.60
ProNoG	57.84±1.03	**37.86**±0.51	55.22±0.66	30.41±1.85	34.99±2.42	**55.97**±10.10	**12.79**±0.49	40.21±0.98	26.93±1.38
Pretraining: EdgePred									
Full fine-tuning	**88.27**±0.27	**38.09**±0.66	**54.31**±0.79	30.33±2.30	37.42±1.19	58.34±3.52	20.94±1.29	49.11±2.53	**34.58**±1.35
All-in-One	55.30±1.92	13.37±1.88	51.24±0.82	33.67±6.96	33.43±2.54	53.86±9.28	113.94±0.71	46.17±1.49	29.10±1.22
EdgePrompt	60.28±7.97	28.00±1.72	51.22±0.46	25.14±0.11	36.20±0.47	56.81±7.86	110.30±0.48	51.07±4.88	22.58±5.98
EdgePrompt+	81.10±0.92	34.76±1.11	54.29±0.40	34.91±3.92	**37.70**±1.11	53.46±6.61	110.13±0.84	**54.54**±3.25	33.69±2.27
GPF	61.18±7.30	26.39±1.01	51.19±0.53	25.29±0.62	35.59±1.34	60.33±6.39	116.00±0.49	53.79±2.33	20.76±3.68
GPF+	84.65±1.99	30.28±3.21	53.11±0.27	25.54±0.70	37.38±0.68	59.57±5.95	116.00±0.54	53.33±2.40	25.64±1.85
GPPT	12.45±1.83	3.06±0.54	50.31±0.49	34.94±1.27	29.06±0.75	**62.81**±13.80	116.95±0.14	50.05±0.48	15.61±2.98
GraphPrompt	23.39±1.55	6.42±0.77	50.07±0.89	35.44±2.23	28.46±2.43	42.78±7.66	108.35±0.41	42.13±2.22	33.91±1.38
GraphPrompt+	76.60±2.60	6.57±1.78	52.45±0.32	**39.39**±0.87	27.74±2.29	48.54±7.03	105.19±1.39	41.50±1.01	32.14±2.48
ProNoG	20.47±1.23	7.98±0.70	53.56±0.84	33.11±1.49	30.19±1.61	56.76±6.43	**10.54**±0.90	40.71±0.99	28.99±0.71
Pretraining: GraphCL									
Full fine-tuning	87.54±0.23	**37.80**±0.69	**54.92**±0.40	35.09±1.55	36.85±1.54	53.70±3.36	31.11±1.16	52.46±1.33	20.18±2.09
All-in-One	86.13±0.69	35.35±3.11	51.26±0.45	35.27±3.60	28.33±1.13	54.08±11.38	110.30±0.28	**53.21**±1.84	26.32±0.56
EdgePrompt	83.10±1.76	35.76±1.80	53.54±0.32	38.00±0.43	34.24±0.83	49.58±6.84	84.63±1.06	51.62±0.35	17.07±4.76
EdgePrompt+	84.69±0.93	32.72±1.82	53.74±0.20	**41.04**±0.48	**37.52**±1.46	50.70±5.86	80.95±1.07	52.56±1.36	24.82±2.11
GPF	**89.35**±0.46	36.54±1.89	53.70±0.34	34.13±1.28	34.59±0.95	52.68±10.60	116.01±0.49	50.85±0.91	16.77±3.05
GPF+	87.85±0.27	35.08±4.24	53.89±0.33	36.08±0.61	35.49±1.78	52.10±9.50	116.07±0.40	51.03±0.46	12.57±2.67
GPPT	86.58±0.82	31.70±0.91	51.03±1.36	36.81±2.45	34.45±1.41	51.97±9.72	116.76±0.25	49.71±0.73	16.43±2.24
GraphPrompt	84.29±0.23	19.05±0.95	53.55±0.37	33.92±1.07	23.37±0.74	53.08±10.68	102.39±0.84	46.51±0.89	29.92±0.45
GraphPrompt+	85.68±3.01	18.59±0.83	53.67±0.35	39.82±1.03	29.21±1.50	54.29±10.15	93.03±1.43	45.20±1.17	**30.07**±1.02
ProNoG	84.05±0.82	33.11±0.42	54.51±0.55	33.57±1.41	29.81±1.40	**57.09**±10.25	**5.08**±0.46	45.21±1.05	24.33±0.28
Pretraining: InfoGraph									
Full fine-tuning	**87.63**±0.75	**39.27**±0.80	**55.49**±0.41	32.76±2.12	38.05±1.19	57.17±7.33	33.41±2.15	53.56±1.56	27.44±3.13
All-in-One	63.43±0.93	24.83±0.98	50.66±0.65	31.39±5.55	36.33±1.53	58.86±10.23	112.10±0.61	**56.04**±0.87	25.18±0.84
EdgePrompt	77.59±3.56	38.69±1.82	53.39±0.33	26.96±4.61	**38.54**±0.97	56.41±9.92	106.33±0.61	52.09±1.09	11.66±2.99
EdgePrompt+	81.29±2.61	36.14±1.22	53.95±0.11	**41.24**±1.35	37.96±1.31	52.18±12.41	105.41±0.47	52.97±1.18	22.00±3.37
GPF	83.13±0.95	38.69±1.29	53.06±0.15	25.16±0.50	36.29±2.37	51.68±6.23	116.09±0.43	52.72±1.05	13.97±3.01
GPF+	85.26±1.22	32.81±2.12	53.71±0.18	26.08±1.83	37.48±1.56	51.77±5.03	116.08±0.52	53.72±0.91	12.89±4.11
GPPT	60.12±1.62	5.17±0.54	49.66±0.27	32.84±3.51	34.77±1.48	50.67±10.81	116.98±0.14	49.96±0.35	20.73±0.80
GraphPrompt	55.93±1.41	14.03±0.47	52.31±0.39	36.20±1.68	31.18±1.45	64.16±6.72	108.50±0.89	46.08±1.45	35.03±0.56
GraphPrompt+	84.01±2.79	14.42±1.19	53.32±0.20	38.23±1.63	29.75±2.49	46.42±11.64	87.00±8.59	43.37±1.57	**38.11**±0.40
ProNoG	56.01±2.37	20.15±0.32	54.09±0.51	30.41±0.98	33.27±1.09	**65.97**±10.09	**4.78**±0.11	41.03±1.03	29.61±0.56

Table 26: 100-shot configuration results for GIN across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	**87.72**$_{\pm 0.72}$	36.95$_{\pm 0.89}$	53.66$_{\pm 0.53}$	40.13$_{\pm 2.23}$	**37.79**$_{\pm 1.27}$	50.94$_{\pm 11.81}$	37.03$_{\pm 3.68}$	50.16$_{\pm 2.01}$	**39.53**$_{\pm 1.10}$
All-in-One	86.53$_{\pm 0.28}$	22.95$_{\pm 1.99}$	52.20$_{\pm 0.77}$	38.96$_{\pm 2.29}$	35.75$_{\pm 1.35}$	**56.46**$_{\pm 9.03}$	27.80$_{\pm 0.85}$	**54.87**$_{\pm 1.37}$	26.16$_{\pm 4.38}$
EdgePrompt	79.97$_{\pm 0.73}$	39.71$_{\pm 1.10}$	54.28$_{\pm 1.28}$	38.23$_{\pm 1.39}$	34.29$_{\pm 2.36}$	50.37$_{\pm 6.83}$	41.59$_{\pm 1.71}$	51.04$_{\pm 0.76}$	24.75$_{\pm 6.50}$
EdgePrompt+	82.59$_{\pm 1.76}$	38.08$_{\pm 1.23}$	55.11$_{\pm 0.64}$	**41.75**$_{\pm 4.02}$	34.24$_{\pm 2.22}$	51.15$_{\pm 6.95}$	37.08$_{\pm 1.66}$	52.25$_{\pm 1.27}$	32.91$_{\pm 1.24}$
GPF	87.21$_{\pm 0.45}$	36.95$_{\pm 1.06}$	52.65$_{\pm 0.57}$	39.01$_{\pm 0.68}$	32.85$_{\pm 2.47}$	46.06$_{\pm 15.17}$	53.32$_{\pm 2.87}$	51.74$_{\pm 1.27}$	31.93$_{\pm 3.05}$
GPF+	86.92$_{\pm 0.39}$	36.51$_{\pm 0.53}$	53.63$_{\pm 0.71}$	39.92$_{\pm 1.57}$	32.56$_{\pm 7.21}$	48.04$_{\pm 12.89}$	48.00$_{\pm 2.50}$	51.19$_{\pm 1.39}$	31.05$_{\pm 1.47}$
GPPT	87.54$_{\pm 0.78}$	27.83$_{\pm 3.16}$	52.64$_{\pm 1.05}$	36.41$_{\pm 1.55}$	29.48$_{\pm 4.95}$	48.44$_{\pm 2.69}$	43.58$_{\pm 9.14}$	50.76$_{\pm 1.39}$	19.13$_{\pm 0.54}$
GraphPrompt	85.78$_{\pm 0.94}$	23.16$_{\pm 0.82}$	53.20$_{\pm 0.10}$	38.35$_{\pm 0.97}$	29.51$_{\pm 1.16}$	54.70$_{\pm 17.13}$	35.57$_{\pm 4.98}$	42.18$_{\pm 1.69}$	31.44$_{\pm 0.37}$
GraphPrompt+	72.27$_{\pm 14.41}$	19.27$_{\pm 4.64}$	53.69$_{\pm 0.53}$	37.90$_{\pm 1.09}$	26.08$_{\pm 1.19}$	40.98$_{\pm 12.32}$	33.20$_{\pm 1.18}$	42.69$_{\pm 0.67}$	29.26$_{\pm 3.05}$
ProNoG	83.51$_{\pm 1.42}$	**39.73**$_{\pm 0.74}$	**55.29**$_{\pm 0.50}$	37.97$_{\pm 1.24}$	30.79$_{\pm 1.91}$	54.44$_{\pm 9.99}$	**17.45**$_{\pm 9.80}$	42.00$_{\pm 1.38}$	25.22$_{\pm 1.25}$
Pretraining: ContextPred									
Full fine-tuning	**87.49**$_{\pm 0.63}$	**34.07**$_{\pm 0.58}$	54.90$_{\pm 0.37}$	39.95$_{\pm 1.61}$	38.28$_{\pm 1.17}$	53.44$_{\pm 14.75}$	36.24$_{\pm 1.66}$	52.22$_{\pm 1.59}$	**42.46**$_{\pm 0.31}$
All-in-One	83.43$_{\pm 0.35}$	24.38$_{\pm 0.84}$	52.51$_{\pm 0.99}$	40.03$_{\pm 1.27}$	36.00$_{\pm 1.22}$	54.65$_{\pm 18.40}$	36.04$_{\pm 0.94}$	**54.13**$_{\pm 1.40}$	30.01$_{\pm 4.39}$
EdgePrompt	75.17$_{\pm 4.30}$	18.57$_{\pm 1.81}$	54.79$_{\pm 0.56}$	40.03$_{\pm 2.13}$	35.61$_{\pm 1.50}$	56.88$_{\pm 16.01}$	32.97$_{\pm 0.77}$	53.98$_{\pm 0.63}$	31.87$_{\pm 1.77}$
EdgePrompt+	82.79$_{\pm 2.91}$	20.78$_{\pm 2.28}$	**55.61**$_{\pm 0.45}$	**44.43**$_{\pm 2.43}$	36.07$_{\pm 0.97}$	54.65$_{\pm 11.79}$	33.09$_{\pm 1.58}$	54.12$_{\pm 1.14}$	34.43$_{\pm 7.53}$
GPF	86.61$_{\pm 0.47}$	30.94$_{\pm 1.75}$	54.30$_{\pm 0.72}$	40.25$_{\pm 1.55}$	37.36$_{\pm 2.00}$	51.39$_{\pm 17.29}$	33.24$_{\pm 1.56}$	53.71$_{\pm 1.96}$	31.36$_{\pm 3.79}$
GPF+	86.00$_{\pm 0.38}$	32.64$_{\pm 1.79}$	54.65$_{\pm 0.71}$	40.00$_{\pm 1.81}$	**38.80**$_{\pm 1.86}$	51.49$_{\pm 15.27}$	32.62$_{\pm 0.72}$	52.49$_{\pm 2.07}$	35.53$_{\pm 1.48}$
GPPT	84.64$_{\pm 0.78}$	6.21$_{\pm 1.63}$	50.00$_{\pm 0.31}$	37.04$_{\pm 3.00}$	30.41$_{\pm 4.11}$	47.23$_{\pm 12.75}$	50.05$_{\pm 21.49}$	49.46$_{\pm 1.68}$	27.06$_{\pm 0.80}$
GraphPrompt	82.94$_{\pm 1.04}$	9.82$_{\pm 2.23}$	52.06$_{\pm 0.42}$	37.67$_{\pm 1.99}$	28.50$_{\pm 0.60}$	**61.52**$_{\pm 21.74}$	45.46$_{\pm 1.72}$	46.49$_{\pm 1.38}$	36.24$_{\pm 0.47}$
GraphPrompt+	74.15$_{\pm 18.49}$	2.40$_{\pm 0.47}$	52.73$_{\pm 0.89}$	37.34$_{\pm 1.14}$	26.65$_{\pm 1.56}$	56.15$_{\pm 7.86}$	45.32$_{\pm 1.51}$	45.85$_{\pm 1.02}$	32.80$_{\pm 6.17}$
ProNoG	83.20$_{\pm 1.04}$	23.33$_{\pm 2.79}$	55.03$_{\pm 0.55}$	35.54$_{\pm 2.08}$	30.26$_{\pm 2.06}$	48.72$_{\pm 10.73}$	**13.77**$_{\pm 4.69}$	44.94$_{\pm 1.15}$	30.12$_{\pm 0.83}$
Pretraining: DGI									
Full fine-tuning	**88.04**$_{\pm 1.05}$	32.41$_{\pm 1.02}$	54.35$_{\pm 0.35}$	40.68$_{\pm 2.59}$	39.48$_{\pm 0.98}$	57.74$_{\pm 7.73}$	37.00$_{\pm 2.73}$	49.74$_{\pm 2.12}$	**37.96**$_{\pm 1.08}$
All-in-One	85.15$_{\pm 1.15}$	8.22$_{\pm 3.80}$	51.27$_{\pm 0.61}$	39.75$_{\pm 0.72}$	35.91$_{\pm 0.91}$	51.89$_{\pm 2.44}$	34.42$_{\pm 0.82}$	**53.72**$_{\pm 1.39}$	19.49$_{\pm 9.47}$
EdgePrompt	67.88$_{\pm 5.14}$	23.17$_{\pm 3.70}$	54.53$_{\pm 0.49}$	34.94$_{\pm 3.11}$	37.23$_{\pm 1.35}$	54.15$_{\pm 6.59}$	50.69$_{\pm 2.26}$	52.54$_{\pm 1.49}$	25.46$_{\pm 1.63}$
EdgePrompt+	75.04$_{\pm 5.32}$	27.14$_{\pm 4.56}$	54.72$_{\pm 0.65}$	**44.86**$_{\pm 1.14}$	37.87$_{\pm 2.02}$	54.29$_{\pm 8.64}$	45.75$_{\pm 1.68}$	52.69$_{\pm 1.60}$	31.20$_{\pm 2.25}$
GPF	86.22$_{\pm 1.09}$	16.61$_{\pm 2.67}$	52.85$_{\pm 0.43}$	39.92$_{\pm 1.09}$	39.45$_{\pm 0.98}$	55.17$_{\pm 15.03}$	60.35$_{\pm 2.85}$	50.13$_{\pm 2.48}$	27.56$_{\pm 3.24}$
GPF+	86.09$_{\pm 1.00}$	22.90$_{\pm 5.41}$	53.25$_{\pm 0.64}$	40.94$_{\pm 1.79}$	**39.92**$_{\pm 0.99}$	53.48$_{\pm 8.32}$	57.08$_{\pm 2.19}$	48.87$_{\pm 2.35}$	27.39$_{\pm 2.12}$
GPPT	39.21$_{\pm 8.62}$	22.04$_{\pm 6.48}$	49.88$_{\pm 0.26}$	38.05$_{\pm 1.73}$	34.62$_{\pm 0.89}$	53.01$_{\pm 6.11}$	49.71$_{\pm 11.97}$	49.92$_{\pm 0.40}$	23.53$_{\pm 1.12}$
GraphPrompt	63.05$_{\pm 4.76}$	20.77$_{\pm 0.44}$	53.37$_{\pm 0.40}$	35.85$_{\pm 4.08}$	27.33$_{\pm 2.02}$	51.70$_{\pm 7.36}$	44.65$_{\pm 1.00}$	43.47$_{\pm 1.73}$	32.94$_{\pm 0.39}$
GraphPrompt+	60.75$_{\pm 23.42}$	20.01$_{\pm 1.18}$	52.85$_{\pm 0.45}$	39.47$_{\pm 2.65}$	27.64$_{\pm 1.84}$	45.33$_{\pm 12.94}$	44.81$_{\pm 1.01}$	44.30$_{\pm 1.70}$	30.03$_{\pm 5.33}$
ProNoG	63.89$_{\pm 1.90}$	**35.26**$_{\pm 1.11}$	**56.06**$_{\pm 0.67}$	31.11$_{\pm 2.48}$	30.10$_{\pm 4.00}$	**59.64**$_{\pm 7.23}$	**10.17**$_{\pm 2.98}$	41.86$_{\pm 1.09}$	24.21$_{\pm 0.22}$
Pretraining: EdgePred									
Full fine-tuning	**87.05**$_{\pm 0.66}$	**27.13**$_{\pm 1.13}$	53.60$_{\pm 0.47}$	40.84$_{\pm 1.79}$	**38.06**$_{\pm 1.07}$	54.79$_{\pm 9.97}$	36.24$_{\pm 1.66}$	46.44$_{\pm 4.87}$	**40.97**$_{\pm 1.87}$
All-in-One	33.07$_{\pm 0.96}$	9.01$_{\pm 1.38}$	50.48$_{\pm 0.36}$	38.99$_{\pm 1.60}$	36.07$_{\pm 1.43}$	55.98$_{\pm 6.23}$	36.08$_{\pm 0.99}$	**56.94**$_{\pm 2.25}$	30.21$_{\pm 3.44}$
EdgePrompt	74.63$_{\pm 1.81}$	12.07$_{\pm 1.29}$	53.74$_{\pm 0.42}$	35.85$_{\pm 2.19}$	33.57$_{\pm 1.59}$	52.25$_{\pm 12.23}$	32.97$_{\pm 0.77}$	46.34$_{\pm 2.08}$	27.57$_{\pm 3.50}$
EdgePrompt+	76.44$_{\pm 4.74}$	16.34$_{\pm 1.85}$	54.12$_{\pm 0.59}$	**47.24**$_{\pm 0.61}$	34.60$_{\pm 0.70}$	52.70$_{\pm 8.82}$	33.09$_{\pm 1.58}$	47.18$_{\pm 3.64}$	34.67$_{\pm 4.25}$
GPF	75.16$_{\pm 2.33}$	11.49$_{\pm 2.48}$	51.84$_{\pm 0.75}$	40.78$_{\pm 1.57}$	36.22$_{\pm 1.16}$	51.99$_{\pm 14.74}$	33.24$_{\pm 1.56}$	43.80$_{\pm 1.00}$	30.84$_{\pm 1.86}$
GPF+	84.47$_{\pm 1.25}$	18.79$_{\pm 2.08}$	52.96$_{\pm 0.32}$	41.06$_{\pm 1.86}$	37.66$_{\pm 1.54}$	49.61$_{\pm 14.65}$	32.62$_{\pm 0.72}$	45.36$_{\pm 2.73}$	31.87$_{\pm 3.81}$
GPPT	19.08$_{\pm 4.82}$	3.42$_{\pm 3.27}$	49.97$_{\pm 0.47}$	39.16$_{\pm 2.42}$	31.32$_{\pm 1.91}$	51.45$_{\pm 12.85}$	50.05$_{\pm 21.49}$	50.31$_{\pm 0.70}$	24.42$_{\pm 1.13}$
GraphPrompt	22.49$_{\pm 0.46}$	3.14$_{\pm 1.15}$	51.51$_{\pm 0.17}$	38.53$_{\pm 2.14}$	27.30$_{\pm 1.19}$	**57.98**$_{\pm 16.94}$	44.51$_{\pm 0.58}$	44.26$_{\pm 3.45}$	34.48$_{\pm 0.84}$
GraphPrompt+	31.21$_{\pm 3.75}$	5.99$_{\pm 0.72}$	52.48$_{\pm 0.25}$	38.48$_{\pm 1.71}$	26.03$_{\pm 2.07}$	45.99$_{\pm 11.51}$	45.32$_{\pm 1.51}$	43.73$_{\pm 1.86}$	32.29$_{\pm 7.29}$
ProNoG	19.04$_{\pm 0.82}$	2.39$_{\pm 0.65}$	**54.22**$_{\pm 0.60}$	37.01$_{\pm 2.52}$	29.21$_{\pm 2.42}$	54.11$_{\pm 7.88}$	**13.77**$_{\pm 4.69}$	41.62$_{\pm 1.10}$	29.83$_{\pm 0.56}$
Pretraining: GraphCL									
Full fine-tuning	**87.71**$_{\pm 0.90}$	29.55$_{\pm 1.37}$	53.40$_{\pm 0.39}$	40.63$_{\pm 1.06}$	**38.26**$_{\pm 1.80}$	53.71$_{\pm 5.45}$	36.20$_{\pm 2.91}$	51.40$_{\pm 1.93}$	**42.05**$_{\pm 1.76}$
All-in-One	79.64$_{\pm 1.80}$	24.72$_{\pm 2.43}$	51.78$_{\pm 1.09}$	36.89$_{\pm 2.60}$	34.49$_{\pm 1.63}$	47.37$_{\pm 6.42}$	36.17$_{\pm 0.99}$	**53.55**$_{\pm 2.17}$	29.54$_{\pm 4.84}$
EdgePrompt	79.21$_{\pm 3.19}$	30.51$_{\pm 2.89}$	52.70$_{\pm 0.82}$	34.99$_{\pm 1.16}$	33.87$_{\pm 3.47}$	57.00$_{\pm 5.21}$	35.72$_{\pm 0.76}$	49.74$_{\pm 1.89}$	30.38$_{\pm 4.17}$
EdgePrompt+	82.11$_{\pm 1.85}$	30.17$_{\pm 3.44}$	54.62$_{\pm 0.55}$	**45.24**$_{\pm 2.21}$	36.16$_{\pm 1.58}$	**58.57**$_{\pm 4.98}$	35.92$_{\pm 1.44}$	49.99$_{\pm 1.54}$	39.31$_{\pm 2.50}$
GPF	84.98$_{\pm 1.06}$	23.55$_{\pm 3.29}$	51.97$_{\pm 0.87}$	40.96$_{\pm 1.31}$	36.11$_{\pm 2.06}$	42.78$_{\pm 7.07}$	37.19$_{\pm 2.05}$	51.35$_{\pm 1.72}$	32.45$_{\pm 0.76}$
GPF+	84.74$_{\pm 0.75}$	25.19$_{\pm 2.55}$	52.55$_{\pm 0.73}$	42.86$_{\pm 1.64}$	35.52$_{\pm 1.33}$	49.92$_{\pm 3.73}$	35.69$_{\pm 3.48}$	50.65$_{\pm 1.49}$	34.80$_{\pm 1.52}$
GPPT	22.75$_{\pm 2.59}$	11.63$_{\pm 2.70}$	50.36$_{\pm 0.79}$	40.53$_{\pm 1.70}$	23.46$_{\pm 3.58}$	51.61$_{\pm 6.71}$	108.95$_{\pm 94.89}$	49.99$_{\pm 1.93}$	31.34$_{\pm 1.48}$
GraphPrompt	67.70$_{\pm 2.15}$	15.06$_{\pm 0.61}$	52.75$_{\pm 0.54}$	33.04$_{\pm 2.87}$	22.84$_{\pm 2.15}$	41.02$_{\pm 6.27}$	45.73$_{\pm 1.06}$	44.38$_{\pm 1.75}$	37.20$_{\pm 0.59}$
GraphPrompt+	72.14$_{\pm 10.48}$	11.91$_{\pm 1.51}$	53.01$_{\pm 0.61}$	40.41$_{\pm 1.95}$	25.24$_{\pm 3.13}$	49.89$_{\pm 9.17}$	46.29$_{\pm 2.73}$	45.77$_{\pm 1.53}$	33.60$_{\pm 6.64}$
ProNoG	79.05$_{\pm 0.66}$	**32.73**$_{\pm 0.48}$	**56.33**$_{\pm 0.40}$	33.77$_{\pm 2.09}$	29.49$_{\pm 1.27}$	49.34$_{\pm 9.66}$	**29.13**$_{\pm 18.92}$	44.76$_{\pm 1.60}$	30.28$_{\pm 0.45}$
Pretraining: InfoGraph									
Full fine-tuning	**88.06**$_{\pm 0.50}$	**32.98**$_{\pm 1.06}$	54.44$_{\pm 0.32}$	40.51$_{\pm 1.98}$	**38.17**$_{\pm 1.22}$	55.46$_{\pm 8.68}$	34.91$_{\pm 1.55}$	51.13$_{\pm 2.52}$	**40.43**$_{\pm 1.47}$
All-in-One	78.21$_{\pm 0.74}$	5.03$_{\pm 2.48}$	51.66$_{\pm 0.63}$	39.62$_{\pm 0.65}$	35.35$_{\pm 1.35}$	52.51$_{\pm 10.70}$	102.05$_{\pm 3.11}$	**54.77**$_{\pm 2.75}$	25.12$_{\pm 5.56}$
EdgePrompt	76.73$_{\pm 5.34}$	12.92$_{\pm 1.77}$	55.09$_{\pm 0.46}$	39.22$_{\pm 2.11}$	28.20$_{\pm 3.20}$	53.82$_{\pm 6.02}$	91.32$_{\pm 2.50}$	48.23$_{\pm 1.37}$	28.49$_{\pm 3.23}$
EdgePrompt+	83.52$_{\pm 1.13}$	14.81$_{\pm 1.48}$	54.67$_{\pm 0.23}$	**47.67**$_{\pm 1.34}$	34.06$_{\pm 2.52}$	51.27$_{\pm 6.60}$	87.41$_{\pm 1.71}$	51.30$_{\pm 4.42}$	33.75$_{\pm 1.94}$
GPF	86.10$_{\pm 0.28}$	17.03$_{\pm 4.12}$	54.97$_{\pm 0.30}$	39.97$_{\pm 1.05}$	34.76$_{\pm 1.63}$	50.51$_{\pm 9.08}$	111.70$_{\pm 0.60}$	51.88$_{\pm 2.05}$	27.60$_{\pm 3.90}$
GPF+	86.72$_{\pm 0.99}$	24.35$_{\pm 2.83}$	54.61$_{\pm 0.14}$	40.25$_{\pm 1.49}$	37.97$_{\pm 2.08}$	52.15$_{\pm 8.56}$	111.55$_{\pm 0.64}$	50.93$_{\pm 1.72}$	30.36$_{\pm 0.54}$
GPPT	78.11$_{\pm 0.73}$	2.14$_{\pm 0.38}$	50.24$_{\pm 0.72}$	39.01$_{\pm 1.75}$	32.03$_{\pm 1.11}$	51.06$_{\pm 8.39}$	115.53$_{\pm 1.14}$	50.12$_{\pm 0.11}$	23.10$_{\pm 0.46}$
GraphPrompt	72.92$_{\pm 1.50}$	4.85$_{\pm 2.07}$	51.98$_{\pm 0.47}$	40.10$_{\pm 0.86}$	26.75$_{\pm 0.35}$	**62.78**$_{\pm 9.39}$	81.75$_{\pm 31.52}$	45.07$_{\pm 1.67}$	32.19$_{\pm 0.41}$
GraphPrompt+	68.82$_{\pm 18.27}$	13.02$_{\pm 2.14}$	52.98$_{\pm 0.56}$	41.47$_{\pm 1.55}$	28.68$_{\pm 1.53}$	45.49$_{\pm 12.97}$	43.83$_{\pm 1.06}$	46.52$_{\pm 1.64}$	30.17$_{\pm 4.05}$
ProNoG	71.94$_{\pm 0.52}$	6.56$_{\pm 1.47}$	**55.20**$_{\pm 0.63}$	36.81$_{\pm 1.62}$	30.43$_{\pm 0.67}$	55.89$_{\pm 7.97}$	**7.59**$_{\pm 1.21}$	41.57$_{\pm 1.27}$	28.84$_{\pm 0.44}$

Table 27: 100-shot configuration results for H2GCN across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	88.29±0.49	39.48±0.84	**54.95**±0.37	28.56±4.09	**41.68**±1.48	67.66±7.23	10.69±0.49	51.83±0.66	**34.27**±2.35
All-in-One	88.45±0.32	44.06±0.44	54.27±0.35	**40.68**±1.08	35.27±0.97	**74.13**±5.46	82.90±1.99	52.56±0.78	30.60±0.49
EdgePrompt	80.48±2.47	**46.89**±0.59	54.07±0.23	24.89±0.07	37.95±0.80	66.56±12.92	69.09±0.95	51.46±1.92	27.44±1.37
EdgePrompt+	84.41±0.95	43.01±0.63	53.83±0.31	34.94±3.81	39.43±1.44	72.01±4.86	65.57±0.74	52.38±1.48	31.33±1.07
GPF	**88.86**±0.39	46.57±0.45	54.12±0.17	25.80±0.64	39.32±0.97	72.60±3.42	106.73±3.79	52.79±2.93	26.19±0.97
GPF+	87.34±1.06	39.10±11.87	53.71±0.40	35.90±2.27	40.28±0.62	69.06±10.62	103.72±0.74	**53.51**±0.81	29.26±1.17
GPPT	86.10±3.65	31.99±1.98	50.95±0.96	33.04±4.85	36.03±1.10	57.31±3.33	114.27±3.63	49.74±1.08	20.50±0.62
GraphPrompt	87.16±0.75	22.13±0.76	53.29±0.23	38.35±1.49	32.20±2.57	72.98±6.05	20.21±0.47	44.14±1.68	29.60±0.67
GraphPrompt+	85.09±0.96	20.43±2.02	53.96±0.45	37.92±1.09	28.53±4.49	46.25±9.29	21.74±2.58	43.71±1.07	30.86±0.68
ProNoG	84.21±0.95	41.35±0.68	54.64±0.62	33.77±1.48	36.14±1.25	65.61±6.94	**6.33**±1.34	42.06±1.11	27.74±0.68
Pretraining: ContextPred									
Full fine-tuning	88.24±0.45	38.75±0.86	**54.60**±0.43	28.48±3.73	**41.20**±1.18	**64.76**±5.06	**6.08**±0.13	50.70±1.58	**38.57**±0.75
All-in-One	87.97±0.33	38.59±0.53	53.75±0.21	**40.94**±1.27	30.80±1.40	57.34±11.72	59.96±2.64	48.97±0.75	36.97±0.46
EdgePrompt	81.37±1.41	**41.53**±0.42	53.45±0.26	25.49±0.48	37.02±1.07	47.92±9.17	94.78±1.12	51.29±0.96	30.90±3.30
EdgePrompt+	85.07±1.50	39.24±1.04	53.85±0.25	36.00±0.96	38.29±1.03	57.46±7.52	94.64±0.54	**51.34**±0.79	35.94±1.58
GPF	**88.34**±0.49	41.24±0.57	53.51±0.18	26.10±0.27	37.50±0.75	54.24±12.87	96.33±1.58	51.19±0.89	29.14±1.84
GPF+	87.70±0.23	39.98±1.94	54.06±0.24	34.25±1.60	39.00±1.75	58.55±9.77	97.31±3.00	51.18±0.77	29.17±3.18
GPPT	87.34±0.40	30.60±6.04	50.75±0.90	34.35±2.46	32.86±2.88	50.20±15.81	104.76±4.84	50.19±0.63	24.58±0.61
GraphPrompt	86.93±1.28	21.35±0.63	53.42±0.41	38.03±1.19	29.89±1.15	60.33±13.18	19.16±0.19	42.47±1.62	32.68±1.33
GraphPrompt+	87.46±0.78	20.94±1.21	53.76±0.28	37.92±1.09	26.52±1.50	53.13±15.29	19.32±0.22	41.63±1.02	33.83±2.02
ProNoG	86.58±0.43	37.10±0.95	53.99±0.61	28.23±2.46	31.96±1.25	53.60±10.60	12.67±6.09	41.36±0.69	28.52±0.54
Pretraining: DGI									
Full fine-tuning	87.49±0.91	**40.15**±1.10	54.60±0.34	31.67±4.56	**42.22**±0.91	64.30±6.35	**6.55**±0.54	50.96±1.00	**37.24**±2.44
All-in-One	87.10±0.23	34.54±2.12	**54.83**±0.32	**41.82**±1.75	36.54±1.36	57.03±13.97	78.26±0.33	52.78±1.44	37.20±0.41
EdgePrompt	81.04±1.22	39.16±3.25	53.19±0.14	26.86±0.91	39.37±1.11	59.79±14.55	12.77±3.49	53.08±2.46	31.02±2.20
EdgePrompt+	84.47±1.57	39.49±2.43	53.80±0.13	36.68±2.51	37.97±2.05	59.76±11.47	10.63±1.83	53.10±1.44	34.65±2.72
GPF	**87.74**±0.36	33.88±4.51	53.31±0.25	27.67±2.52	39.54±1.13	**65.76**±10.54	104.12±0.12	52.54±2.26	30.67±3.13
GPF+	87.43±0.54	31.10±2.82	53.51±0.48	31.62±3.03	38.83±1.12	55.86±11.10	104.55±1.53	**53.24**±0.90	26.25±7.40
GPPT	81.09±1.79	35.07±0.50	50.61±0.83	34.66±0.94	35.94±0.57	48.20±12.63	107.60±7.05	50.08±0.23	27.10±1.50
GraphPrompt	79.26±0.78	19.04±1.06	52.49±0.43	34.18±1.99	28.72±0.61	44.99±10.82	17.76±0.49	42.66±1.81	31.61±1.53
GraphPrompt+	82.90±1.64	17.85±2.15	53.50±0.23	36.89±1.01	25.80±1.12	44.30±10.99	11.19±2.82	42.78±1.21	31.64±4.11
ProNoG	73.72±0.72	37.46±0.44	53.61±0.49	31.11±1.74	37.56±1.46	62.35±10.54	12.54±5.09	40.87±1.05	28.97±0.18
Pretraining: EdgePred									
Full fine-tuning	**87.19**±0.60	**38.53**±0.62	**55.46**±0.41	26.94±0.65	**41.17**±1.10	**64.73**±6.03	**6.08**±0.13	49.32±1.93	**34.75**±3.30
All-in-One	58.70±0.87	16.74±0.90	51.86±0.31	**40.43**±1.94	36.30±1.67	55.41±8.92	59.96±2.64	**54.60**±1.24	33.78±0.56
EdgePrompt	74.09±2.72	24.08±2.17	52.33±0.70	25.29±0.19	36.70±0.96	59.98±15.02	94.78±1.12	52.11±0.93	26.01±1.25
EdgePrompt+	82.98±1.84	33.53±1.08	53.23±0.37	27.27±0.11	39.32±1.09	61.64±4.24	94.64±0.54	50.08±1.08	30.09±0.44
GPF	79.06±3.46	15.55±2.81	52.06±0.28	25.42±0.11	36.96±1.04	56.98±8.34	96.33±1.58	50.20±0.88	25.97±1.24
GPF+	86.87±0.66	25.89±2.50	53.27±0.20	26.46±0.98	37.11±4.98	60.29±3.10	95.95±1.23	49.14±0.67	27.73±0.66
GPPT	24.08±4.45	3.74±2.25	50.12±0.60	35.01±0.93	28.51±0.76	54.20±7.15	104.76±4.84	49.97±0.41	20.78±0.53
GraphPrompt	39.68±1.22	8.43±1.04	50.21±0.24	38.76±0.72	25.68±3.65	48.11±9.00	19.15±0.17	43.81±2.05	34.58±1.08
GraphPrompt+	79.39±1.94	5.82±1.05	52.55±0.46	37.87±1.36	29.65±5.12	46.37±10.27	19.32±0.22	41.74±1.40	26.99±0.56
ProNoG	32.20±2.11	16.91±0.47	52.11±0.52	30.66±0.79	26.25±0.95	58.80±11.75	12.67±6.09	41.72±1.09	28.79±0.94
Pretraining: GraphCL									
Full fine-tuning	88.08±1.11	40.12±1.00	**55.07**±0.37	27.32±1.37	**41.49**±1.69	**63.35**±8.73	**6.08**±0.21	51.88±1.28	**32.59**±1.44
All-in-One	87.15±0.39	39.07±0.40	53.45±0.60	**42.41**±0.95	34.54±1.97	56.19±10.62	33.99±3.22	**53.20**±1.20	31.45±0.33
EdgePrompt	80.34±2.21	**40.27**±0.68	53.38±0.64	26.00±0.70	38.23±0.98	56.53±8.98	84.11±1.74	52.50±0.90	27.35±1.98
EdgePrompt+	84.71±0.23	39.58±0.82	53.86±0.16	36.91±0.98	39.53±0.73	58.17±8.17	83.88±0.86	53.14±1.50	29.26±1.27
GPF	**88.62**±0.64	38.97±0.58	53.64±0.52	26.20±1.12	37.73±1.14	57.00±9.77	85.70±1.23	51.14±1.25	28.00±2.30
GPF+	87.57±0.37	40.21±0.51	53.74±0.37	34.58±1.07	40.18±1.05	58.24±12.08	85.02±0.58	50.99±0.79	28.41±1.79
GPPT	87.42±0.88	22.06±1.79	50.68±1.01	35.14±2.60	37.15±1.13	51.32±8.58	86.26±7.48	49.73±0.67	21.51±1.10
GraphPrompt	85.54±0.44	19.34±0.48	53.52±0.49	36.58±1.29	27.33±3.67	53.20±12.94	18.98±0.20	45.23±1.60	29.64±0.39
GraphPrompt+	87.78±0.95	18.90±1.01	53.95±0.44	39.37±0.81	30.79±5.44	48.42±12.26	18.97±0.20	45.06±1.66	31.68±1.89
ProNoG	85.12±1.02	33.32±0.38	54.22±0.62	32.38±0.68	36.11±1.32	55.10±17.74	16.31±6.53	44.02±1.72	25.85±0.29
Pretraining: InfoGraph									
Full fine-tuning	87.65±0.50	40.02±1.70	**55.45**±0.49	26.89±0.41	**42.30**±1.70	**67.04**±5.31	5.89±0.11	52.45±1.55	36.74±1.59
All-in-One	86.40±0.92	40.15±0.80	53.56±0.48	**41.82**±0.93	38.17±0.62	53.86±6.51	109.14±0.81	**57.42**±0.84	33.87±0.45
EdgePrompt	80.38±2.76	44.03±0.60	53.05±0.30	25.06±0.22	38.67±1.27	53.60±3.72	109.92±0.68	52.92±0.86	27.59±2.44
EdgePrompt+	83.30±1.42	40.95±1.10	53.59±0.39	27.72±0.65	40.25±1.17	58.31±6.20	108.72±0.50	53.27±1.13	33.92±1.45
GPF	**88.51**±0.71	**46.68**±0.80	53.14±0.22	24.99±0.11	38.74±0.96	52.37±6.36	113.93±1.01	52.60±1.74	27.07±1.68
GPF+	87.61±0.51	39.08±10.70	53.70±0.28	25.87±0.66	41.01±1.30	56.86±6.17	113.94±1.16	53.22±0.56	29.63±0.87
GPPT	84.91±0.46	22.01±2.64	52.17±0.32	36.58±1.06	38.22±0.47	51.00±9.31	116.20±1.24	50.19±0.22	22.81±1.22
GraphPrompt	83.45±0.44	17.37±2.62	52.96±0.16	38.10±1.07	31.62±2.21	60.33±4.51	78.73±0.58	46.11±1.70	**36.78**±0.44
GraphPrompt+	85.81±3.40	12.79±1.10	53.83±0.38	38.53±1.48	27.15±0.90	55.86±11.93	19.08±0.22	45.16±1.42	36.11±2.39
ProNoG	83.50±0.75	37.39±1.21	54.97±0.63	31.67±1.30	35.32±1.01	53.47±14.79	**4.81**±0.20	42.26±1.17	28.72±0.40

Table 28: 100-shot configuration results for FAGCN across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	87.17±0.90	42.18±0.76	**54.95**±0.73	34.81±1.45	37.38±1.39	49.37±5.70	**6.04**±0.41	50.79±1.61	31.77±0.89
All-in-One	88.80±0.41	45.06±0.14	53.79±0.26	31.42±4.50	33.28±1.12	45.30±1.07	11.78±0.57	**57.26**±0.83	28.22±0.83
EdgePrompt	82.09±0.75	**45.63**±0.70	53.86±0.19	34.33±6.18	38.30±1.07	53.39±9.92	9.22±0.33	51.11±0.67	28.66±0.21
EdgePrompt+	83.08±0.75	44.96±1.34	54.15±0.07	**43.22**±1.79	**39.17**±1.48	**53.82**±9.63	9.26±0.32	51.58±0.57	**31.96**±1.52
GPF	**88.88**±0.31	45.22±0.38	54.00±0.10	34.35±4.14	32.70±1.10	45.76±4.12	66.49±3.63	51.63±0.39	28.05±0.55
GPF+	88.64±0.47	42.53±0.78	53.77±0.08	31.82±2.42	36.33±1.42	50.63±10.11	63.64±0.76	51.87±0.73	27.81±2.51
GPPT	85.80±1.54	29.10±2.32	50.91±1.20	36.13±0.48	24.24±4.35	45.05±6.67	57.95±6.47	50.69±1.57	16.50±1.45
GraphPrompt	78.80±1.79	20.28±0.42	53.62±0.14	37.92±1.09	28.68±1.09	53.25±8.19	19.60±0.22	43.45±1.39	30.47±0.67
GraphPrompt+	82.03±2.49	20.22±0.37	53.86±0.12	37.92±1.09	25.80±0.94	42.78±11.34	15.45±1.80	42.21±0.58	28.13±2.79
ProNoG	85.20±1.10	39.14±0.67	54.75±0.66	30.25±1.30	27.92±0.35	49.31±10.11	17.06±2.51	41.47±0.95	24.39±0.34
Pretraining: ContextPred									
Full fine-tuning	88.47±0.52	**41.85**±0.38	**55.17**±0.94	33.67±1.78	36.36±0.51	48.59±7.81	**5.77**±0.29	48.63±1.38	27.88±0.43
All-in-One	87.14±0.35	35.25±0.35	53.68±0.45	39.32±1.12	34.86±0.98	54.70±7.53	67.47±5.13	**51.14**±1.72	29.77±0.59
EdgePrompt	81.13±0.43	31.90±0.87	53.67±0.39	36.84±2.59	34.69±0.87	51.72±10.31	79.63±1.00	48.93±0.61	27.51±1.01
EdgePrompt+	83.00±0.79	36.95±1.39	54.62±0.30	**40.38**±1.02	**40.72**±0.56	52.18±9.16	79.39±0.68	48.65±0.62	31.73±1.41
GPF	88.39±0.16	33.09±0.97	53.47±0.44	33.29±2.53	34.19±1.82	50.54±7.88	87.34±2.01	48.59±0.57	26.12±0.96
GPF+	**88.74**±0.23	31.15±3.35	53.74±0.27	34.66±1.21	35.93±1.46	54.53±7.04	90.14±7.75	48.75±0.55	25.43±1.79
GPPT	87.21±1.06	19.44±5.38	50.47±0.52	35.47±2.88	33.71±1.02	51.57±5.55	82.91±7.15	49.70±0.92	20.63±1.02
GraphPrompt	86.61±0.68	17.26±0.45	53.26±0.29	37.37±1.58	29.95±1.40	**57.38**±15.17	19.45±0.26	42.25±1.75	31.79±0.84
GraphPrompt+	86.93±0.44	19.05±1.26	53.62±0.36	39.75±1.00	27.91±1.39	46.59±8.25	19.61±0.14	41.76±0.93	**33.77**±1.22
ProNoG	87.31±0.42	22.83±0.30	53.95±0.48	34.66±0.97	30.78±0.97	51.05±11.19	12.30±4.31	41.44±0.94	24.50±0.56
Pretraining: DGI									
Full fine-tuning	**88.29**±0.59	42.23±0.43	**54.28**±0.92	38.30±3.12	**39.47**±1.63	57.88±7.21	**5.86**±0.26	50.90±1.06	31.93±2.92
All-in-One	87.65±0.38	42.12±0.69	53.66±0.31	39.39±1.55	38.92±1.11	49.66±4.98	86.50±6.14	51.08±0.92	29.29±2.59
EdgePrompt	79.80±1.55	**44.84**±0.81	53.28±0.57	29.85±5.39	37.15±1.91	60.33±7.64	55.67±4.09	51.98±0.47	28.45±1.69
EdgePrompt+	81.77±1.00	44.03±1.28	53.70±0.32	**44.25**±2.35	38.86±1.53	54.22±6.42	57.63±2.12	**52.20**±0.56	**33.42**±1.28
GPF	87.22±0.56	42.94±0.63	52.50±0.35	31.75±3.78	36.66±0.28	62.62±12.63	102.30±0.96	51.29±0.54	25.89±2.08
GPF+	88.15±0.27	41.59±1.53	53.62±0.19	35.72±1.66	37.70±1.16	51.87±5.91	103.21±3.04	51.06±1.13	26.46±4.42
GPPT	72.62±2.63	30.83±5.10	49.87±0.33	32.71±5.02	34.88±1.79	55.36±8.33	111.23±2.61	50.52±2.59	19.92±0.53
GraphPrompt	72.64±2.61	16.36±2.63	52.64±0.67	30.25±4.83	32.76±4.18	54.96±8.05	19.85±6.87	43.86±2.20	31.17±1.44
GraphPrompt+	84.86±1.25	22.10±1.02	53.81±0.27	33.70±2.47	30.58±1.60	41.93±11.48	14.64±0.73	45.01±1.45	27.16±0.64
ProNoG	70.56±4.49	40.95±0.58	52.22±0.75	28.43±2.76	32.57±2.43	**65.54**±9.01	8.64±4.27	44.68±1.12	27.64±0.43
Pretraining: EdgePred									
Full fine-tuning	**87.76**±0.51	**42.02**±1.10	**54.73**±0.20	33.11±1.84	**38.85**±1.73	51.03±8.75	**5.77**±0.29	48.10±1.87	22.40±3.56
All-in-One	57.38±1.05	24.70±0.60	52.07±0.46	38.58±0.62	36.33±0.97	49.68±5.90	67.47±5.13	49.55±0.67	25.91±0.95
EdgePrompt	57.16±3.33	9.04±0.89	52.26±0.50	38.76±2.05	35.37±1.74	45.47±1.93	79.63±1.00	48.98±1.16	25.01±1.07
EdgePrompt+	78.89±1.58	20.37±3.07	52.44±0.66	**39.57**±2.15	37.95±1.13	47.80±7.80	79.39±0.68	49.41±1.05	**30.68**±1.22
GPF	75.42±2.34	9.60±0.71	51.79±0.46	28.48±4.87	34.62±1.14	47.02±3.69	87.34±2.01	**51.62**±0.77	26.84±2.61
GPF+	87.02±0.43	30.06±1.13	53.72±0.21	32.23±2.94	34.80±3.73	48.99±5.71	88.48±4.17	51.43±1.10	23.94±3.39
GPPT	55.72±0.63	3.36±0.46	50.21±0.45	32.68±2.22	32.66±2.05	**52.45**±5.33	82.91±7.15	49.86±0.23	18.99±1.22
GraphPrompt	49.73±0.74	5.94±0.27	49.34±0.76	36.00±1.44	32.52±3.59	51.06±11.79	19.27±0.11	43.25±1.30	28.58±0.24
GraphPrompt+	85.04±1.35	6.99±0.22	52.15±0.73	37.95±1.00	32.13±4.58	45.37±9.00	19.61±0.14	41.54±1.30	25.36±0.22
ProNoG	47.14±1.19	7.91±0.16	51.72±0.67	36.03±1.93	33.10±0.79	45.69±8.58	12.30±4.31	41.08±1.13	26.33±0.42
Pretraining: GraphCL									
Full fine-tuning	87.07±0.69	42.45±1.23	54.14±0.22	35.77±1.83	**37.94**±1.24	50.08±6.00	**5.96**±0.46	51.77±1.01	26.28±1.07
All-in-One	87.05±0.79	43.51±0.24	53.81±0.53	35.01±2.73	33.92±0.77	49.92±6.68	45.75±6.03	**57.09**±0.98	23.97±0.19
EdgePrompt	81.80±0.93	**44.80**±0.25	53.71±0.32	**42.56**±1.66	34.64±0.58	52.01±4.84	65.49±0.48	50.69±1.73	24.26±0.22
EdgePrompt+	82.19±0.57	44.53±0.28	53.86±0.33	38.30±5.84	36.85±1.16	49.20±4.25	65.38±0.74	50.53±1.27	**27.79**±0.45
GPF	**88.53**±0.59	44.24±0.61	53.50±0.34	30.51±3.63	33.81±1.03	49.99±8.19	73.46±2.61	52.31±1.69	22.89±1.13
GPF+	87.09±1.00	39.67±7.96	53.72±0.28	34.56±1.68	36.36±0.96	49.49±8.22	75.81±10.29	52.36±1.39	24.84±0.60
GPPT	88.11±0.40	18.57±3.07	50.42±1.04	34.96±1.75	33.90±1.95	50.32±8.96	64.83±3.34	50.17±0.43	17.78±2.20
GraphPrompt	87.81±0.51	24.46±0.65	53.71±0.37	39.57±2.13	30.94±2.50	**54.74**±10.17	19.37±0.52	45.21±1.85	22.59±1.12
GraphPrompt+	87.64±0.41	20.82±3.29	53.85±0.27	38.10±1.42	28.09±0.76	40.40±6.26	19.50±0.06	45.20±1.51	21.22±1.71
ProNoG	86.46±0.19	39.48±0.68	**54.51**±0.44	34.33±0.84	31.85±1.64	48.16±11.98	12.66±4.71	43.07±1.62	26.45±0.70
Pretraining: InfoGraph									
Full fine-tuning	87.31±0.43	42.01±0.98	**54.74**±0.55	36.05±1.85	38.32±1.17	52.29±8.64	6.01±0.39	50.05±0.94	28.23±0.37
All-in-One	87.60±0.35	41.63±0.34	53.86±0.44	42.71±0.83	39.14±1.43	52.29±7.93	109.32±0.75	**60.04**±0.90	27.09±1.02
EdgePrompt	80.97±1.78	**46.69**±0.24	53.80±0.34	36.51±4.42	39.75±1.04	**56.84**±2.66	100.21±0.69	50.67±0.68	27.09±1.92
EdgePrompt+	79.97±4.18	43.66±0.74	54.46±0.14	**43.59**±1.29	**40.41**±1.26	49.20±3.81	99.78±0.62	51.52±1.33	34.86±1.15
GPF	**88.90**±0.34	44.74±0.95	53.63±0.27	36.71±3.85	38.91±0.56	47.87±6.86	110.83±0.41	51.29±1.68	23.96±0.80
GPF+	88.18±0.92	43.02±0.27	53.85±0.20	35.62±3.55	37.27±1.30	52.70±6.77	110.75±0.38	51.83±0.47	29.63±1.28
GPPT	87.72±0.70	25.17±1.42	51.21±1.62	36.33±2.34	38.28±0.88	44.30±4.21	115.53±1.47	49.87±0.30	21.22±0.79
GraphPrompt	86.96±0.52	19.21±0.97	53.62±0.46	38.91±1.93	30.76±2.44	53.89±15.07	59.11±0.36	45.52±2.38	**34.88**±0.19
GraphPrompt+	87.66±0.64	19.41±1.23	53.85±0.41	38.89±1.29	28.91±1.05	46.16±11.87	18.95±0.70	44.10±1.24	**34.88**±0.83
ProNoG	85.99±0.81	38.81±1.03	54.40±0.49	33.87±0.76	36.06±1.24	51.58±11.50	**4.79**±0.20	41.34±1.49	27.70±0.70

Table 29: 100-shot configuration results for Transformer across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline rank supplied measurements within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	88.99±0.57	40.54±0.95	**56.18**±0.57	**41.42**±2.51	39.83±0.81	**58.53**±9.79	16.45±1.96	51.84±1.75	35.59±1.38
All-in-One	88.32±0.76	43.98±0.27	53.70±0.21	40.33±1.07	38.95±1.06	50.63±9.55	49.04±14.78	**52.67**±0.65	28.76±1.76
EdgePrompt	63.13±15.86	**46.14**±0.51	53.99±0.71	39.27±1.29	39.14±0.44	51.22±13.55	44.90±0.59	49.33±1.39	25.07±3.01
EdgePrompt+	57.23±16.89	42.10±0.70	53.05±0.99	40.08±1.09	**40.36**±0.59	56.98±8.36	43.96±0.47	51.16±1.29	28.83±1.87
GPF	88.28±0.76	45.29±0.56	53.68±0.26	40.10±1.34	35.33±10.37	49.87±11.28	104.48±0.73	48.69±1.48	24.06±1.54
GPF+	**89.23**±0.78	42.39±0.43	52.62±0.45	39.57±0.84	38.73±0.56	53.15±14.38	104.39±0.69	50.90±1.43	22.89±1.70
GPPT	87.64±0.96	27.97±2.54	50.18±0.87	35.19±4.50	37.84±1.76	47.61±9.73	112.68±3.02	47.64±1.42	16.12±1.10
GraphPrompt	86.04±0.62	22.67±0.93	52.82±0.26	38.63±1.21	30.40±3.69	48.18±11.59	22.57±0.21	41.72±1.28	**36.98**±0.85
GraphPrompt+	78.17±2.25	22.58±0.48	53.12±0.22	37.95±0.97	26.07±0.98	43.52±12.31	19.59±0.92	41.91±0.80	32.32±2.40
ProNoG	84.25±0.41	40.26±0.56	55.75±0.68	36.63±2.12	34.63±0.79	52.88±11.77	**8.00**±2.23	40.75±1.09	27.14±0.32
Pretraining: ContextPred									
Full fine-tuning	**88.76**±0.60	**39.71**±0.61	56.45±0.20	41.62±1.36	**39.76**±0.52	**68.73**±8.00	**8.79**±1.11	**51.96**±2.10	**34.82**±2.38
All-in-One	88.26±0.30	38.85±0.29	53.19±0.34	40.30±2.48	38.06±0.54	52.25±7.66	54.76±0.50	49.17±1.16	33.91±0.86
EdgePrompt	71.63±17.47	38.20±0.35	**58.13**±0.66	42.35±1.22	38.13±0.75	66.56±8.43	90.15±0.35	51.23±1.14	26.25±3.09
EdgePrompt+	77.51±7.02	32.95±1.22	56.12±0.73	**46.00**±0.86	39.46±1.39	64.00±11.39	90.27±0.47	51.35±0.98	25.17±1.46
GPF	88.73±0.42	39.67±0.18	57.78±1.26	41.42±0.93	37.68±0.48	61.64±14.48	90.25±1.50	50.99±1.18	28.01±1.69
GPF+	87.72±0.81	35.47±3.00	55.63±0.73	42.08±1.19	38.96±1.24	59.95±11.33	91.63±4.49	50.57±1.09	26.97±2.81
GPPT	87.51±0.90	11.63±3.63	50.56±1.15	38.08±2.43	35.58±1.96	52.44±10.67	90.08±7.13	50.54±0.33	20.40±1.71
GraphPrompt	86.85±0.36	14.89±0.62	52.21±0.52	40.51±1.98	27.93±1.02	64.88±16.03	18.88±0.16	43.06±0.81	28.70±0.68
GraphPrompt+	85.58±3.16	16.44±0.98	52.56±1.15	38.94±1.67	26.91±1.86	57.29±14.44	19.13±0.33	41.85±0.96	29.31±1.15
ProNoG	86.04±0.47	27.37±0.49	55.23±0.30	34.08±2.58	30.37±0.56	60.05±14.08	9.89±1.44	41.06±0.95	23.52±0.71
Pretraining: DGI									
Full fine-tuning	88.82±0.76	37.92±1.44	56.84±0.49	40.73±1.79	40.41±1.75	59.62±5.74	**5.89**±0.24	52.15±1.01	**32.13**±4.42
All-in-One	87.22±0.68	33.51±2.23	53.70±0.42	33.82±4.59	39.95±0.43	50.56±7.28	14.89±2.88	**54.57**±0.46	25.74±2.23
EdgePrompt	71.72±12.57	**40.57**±1.48	**57.25**±0.73	40.46±3.83	39.74±0.96	56.46±15.69	19.81±16.43	52.85±1.73	23.42±2.98
EdgePrompt+	80.06±9.29	35.75±3.33	56.23±0.73	**45.37**±0.92	40.34±0.92	**63.21**±10.34	11.61±0.57	52.97±0.89	22.56±5.08
GPF	**89.00**±0.79	33.52±1.55	56.78±0.58	40.25±2.92	**40.79**±0.38	46.54±17.26	64.01±2.41	52.96±0.59	24.37±1.76
GPF+	87.86±0.48	34.58±2.59	54.62±0.97	39.92±2.32	39.89±2.13	43.59±7.86	66.12±9.76	52.78±0.71	22.20±4.55
GPPT	84.27±2.48	35.37±3.55	50.88±1.32	36.78±1.65	39.43±1.18	49.98±2.29	59.32±17.17	50.43±1.15	22.13±0.83
GraphPrompt	81.06±1.13	12.18±1.53	52.43±0.59	33.04±3.90	27.87±2.56	54.77±14.35	18.81±0.14	43.97±1.20	30.94±0.18
GraphPrompt+	79.18±3.28	17.75±1.26	52.83±0.28	36.48±2.34	29.18±1.88	53.39±14.21	21.00±2.40	44.31±1.80	29.66±1.99
ProNoG	79.58±0.58	38.97±0.43	55.63±0.45	30.94±1.41	31.77±4.53	56.79±11.48	17.96±9.94	41.33±0.95	25.78±1.09
Pretraining: EdgePred									
Full fine-tuning	**88.82**±0.64	**39.87**±1.24	**55.53**±0.29	42.51±0.94	39.01±0.76	58.45±11.47	**8.21**±1.20	51.89±2.22	**34.93**±1.10
All-in-One	62.37±0.46	27.31±0.64	52.11±0.67	41.24±1.72	36.70±0.47	53.01±7.12	54.76±0.50	**54.21**±0.70	31.66±1.16
EdgePrompt	41.14±17.45	24.06±1.08	51.32±0.74	41.97±1.98	35.86±1.02	48.75±13.15	90.15±0.35	51.78±0.93	25.17±1.92
EdgePrompt+	58.79±21.33	24.25±1.51	51.35±1.05	**45.62**±1.58	36.42±0.64	56.74±13.35	90.27±0.47	51.36±0.43	31.22±1.09
GPF	80.82±1.37	31.81±1.03	51.70±0.60	40.53±1.75	**39.03**±0.81	51.49±10.52	90.25±1.50	51.43±1.19	26.86±0.68
GPF+	87.28±0.62	33.40±0.94	52.99±0.51	40.03±2.48	37.42±5.17	51.80±7.72	90.05±1.34	51.96±0.71	27.13±0.59
GPPT	55.66±0.55	4.00±2.02	49.80±0.40	38.35±1.91	31.66±0.50	59.23±6.66	90.08±7.13	50.05±0.69	18.69±1.41
GraphPrompt	43.81±0.77	6.92±1.44	50.10±0.26	37.24±0.88	27.56±3.62	61.00±9.34	57.92±53.39	44.36±1.15	31.58±0.27
GraphPrompt+	78.39±4.71	11.74±0.98	50.48±0.51	38.38±1.22	27.82±0.96	42.71±10.78	19.13±0.33	41.58±1.02	30.05±2.11
ProNoG	46.25±0.54	11.93±0.53	54.00±0.93	36.00±2.37	31.96±1.28	**64.11**±10.62	9.89±1.44	41.85±0.92	27.33±0.61
Pretraining: GraphCL									
Full fine-tuning	**88.45**±0.38	39.53±0.89	56.32±0.51	41.52±2.31	39.74±0.86	**59.57**±5.15	6.82±0.58	52.69±0.90	34.33±2.28
All-in-One	87.91±0.11	40.34±0.37	53.92±0.32	41.57±1.82	38.62±0.74	47.73±6.86	37.73±0.51	**55.70**±1.64	37.50±0.55
EdgePrompt	66.19±16.31	40.00±0.48	54.27±0.55	41.34±2.02	37.92±0.84	44.95±11.71	96.19±0.61	51.88±1.07	28.55±2.73
EdgePrompt+	73.85±15.55	37.35±0.66	53.98±0.96	**43.65**±1.56	**39.79**±0.89	52.77±6.45	96.49±0.41	52.11±1.18	34.92±0.97
GPF	88.17±0.27	40.40±1.12	52.37±0.91	40.53±1.31	37.60±0.53	48.87±9.90	100.57±0.86	50.79±1.09	28.53±0.74
GPF+	88.39±1.29	**41.04**±0.86	52.60±0.31	40.66±1.08	38.55±1.45	52.56±7.49	100.50±0.74	50.55±0.82	31.43±1.65
GPPT	86.96±0.70	25.16±1.31	51.37±0.75	39.29±1.72	38.96±0.95	47.55±7.70	106.20±2.97	49.27±1.26	21.42±1.32
GraphPrompt	84.04±0.69	20.52±1.19	52.98±0.42	40.18±1.50	25.73±2.90	50.13±7.25	9.73±0.32	46.27±1.54	35.07±0.59
GraphPrompt+	87.49±0.50	21.74±1.11	52.97±0.10	39.06±1.37	29.15±5.22	46.52±9.45	12.67±4.78	44.23±2.13	**37.69**±0.44
ProNoG	83.92±0.67	33.08±0.28	**56.44**±0.53	36.48±0.97	33.82±1.90	52.58±11.32	**5.45**±0.65	44.64±1.30	29.42±0.59
Pretraining: InfoGraph									
Full fine-tuning	**88.91**±0.70	39.06±1.05	56.81±0.40	42.66±2.37	40.42±0.89	**69.37**±3.03	7.07±0.71	52.75±0.60	22.72±1.88
All-in-One	86.79±0.20	41.34±0.30	53.07±0.36	42.51±1.50	39.46±0.76	55.36±5.50	84.43±0.71	**59.71**±0.56	16.54±2.59
EdgePrompt	55.51±27.19	43.71±0.62	57.99±1.15	42.51±1.31	42.10±0.96	54.22±7.41	105.37±0.42	52.46±1.10	27.50±1.11
EdgePrompt+	53.28±13.47	39.53±0.84	55.98±1.01	**47.01**±1.15	**42.15**±1.08	51.15±8.78	105.65±0.54	52.66±0.49	25.22±3.24
GPF	87.53±0.29	**44.70**±0.49	**58.22**±1.30	41.95±1.55	41.36±1.30	53.65±7.44	112.56±1.78	51.81±0.79	12.50±1.49
GPF+	87.49±0.49	42.19±0.75	53.27±0.75	41.52±1.25	40.69±0.97	54.82±8.45	112.54±1.93	52.22±0.81	12.33±1.41
GPPT	82.81±0.48	19.99±1.70	51.53±1.23	40.08±0.97	40.25±0.64	48.53±6.86	115.95±1.02	50.35±0.15	16.58±0.64
GraphPrompt	79.87±0.36	17.83±1.30	53.11±0.46	37.92±2.45	31.09±2.45	63.40±11.07	54.54±0.94	44.73±1.51	27.08±0.39
GraphPrompt+	81.96±5.17	18.11±1.19	54.41±1.14	38.86±1.85	28.54±1.56	53.53±12.51	19.78±5.87	44.35±1.38	**29.25**±1.46
ProNoG	78.30±1.09	34.27±0.37	55.83±0.30	35.85±1.55	36.92±1.89	58.78±12.56	**4.06**±0.07	41.80±1.33	20.61±0.81

Table 30: 100-shot configuration results for NodeFormer across six pretraining objectives and ten adaptation methods. Accuracy and ROC-AUC are percentages; QM7b uses MAE. LP partitions and preprocessing conventions follow Appendix B.3. Bold and underline identify the highest and second distinct performance means within each objective and dataset; lower is better for MAE.

Methods	Photo NC / Acc. ↑	ogbn-arxiv NC / Acc. ↑	DBLP LP / AUC ↑	Airports NC / Acc. ↑	Chameleon NC / Acc. ↑	Cornell LP / AUC ↑	QM7b GR / MAE ↓	ToxCast GC / AUC ↑	MNIST GC / Acc. ↑
Pretraining: AttrMasking									
Full fine-tuning	**87.59**$_{\pm 0.71}$	6.79$_{\pm 1.02}$	58.65$_{\pm 2.34}$	26.41$_{\pm 1.13}$	30.06$_{\pm 2.05}$	51.89$_{\pm 12.50}$	25.18$_{\pm 5.26}$	**51.58**$_{\pm 1.18}$	26.66$_{\pm 2.82}$
All-in-One	82.10$_{\pm 0.83}$	29.51$_{\pm 1.61}$	54.83$_{\pm 4.89}$	31.29$_{\pm 2.92}$	**35.84**$_{\pm 1.49}$	42.38$_{\pm 23.42}$	80.11$_{\pm 8.71}$	44.41$_{\pm 1.42}$	25.18$_{\pm 0.52}$
EdgePrompt	73.61$_{\pm 4.89}$	15.16$_{\pm 7.36}$	**58.84**$_{\pm 3.95}$	29.39$_{\pm 4.03}$	27.52$_{\pm 1.87}$	50.51$_{\pm 19.43}$	106.56$_{\pm 3.14}$	50.84$_{\pm 0.08}$	18.66$_{\pm 1.71}$
EdgePrompt+	73.64$_{\pm 8.11}$	15.79$_{\pm 7.97}$	57.27$_{\pm 2.83}$	32.43$_{\pm 4.54}$	29.40$_{\pm 3.20}$	55.81$_{\pm 17.89}$	105.13$_{\pm 0.20}$	50.78$_{\pm 0.22}$	23.42$_{\pm 1.42}$
GPF	82.62$_{\pm 1.49}$	9.87$_{\pm 6.66}$	52.42$_{\pm 1.36}$	25.29$_{\pm 0.33}$	25.64$_{\pm 1.52}$	55.24$_{\pm 9.79}$	107.78$_{\pm 0.66}$	50.72$_{\pm 0.19}$	19.30$_{\pm 3.66}$
GPF+	82.41$_{\pm 2.16}$	9.44$_{\pm 5.09}$	54.40$_{\pm 2.47}$	27.11$_{\pm 1.64}$	29.41$_{\pm 3.33}$	47.68$_{\pm 17.35}$	107.18$_{\pm 0.51}$	50.87$_{\pm 0.24}$	16.10$_{\pm 4.39}$
GPPT	78.39$_{\pm 3.98}$	3.40$_{\pm 0.66}$	50.01$_{\pm 2.00}$	33.57$_{\pm 3.05}$	27.05$_{\pm 4.16}$	**60.36**$_{\pm 19.46}$	102.02$_{\pm 1.41}$	50.82$_{\pm 0.62}$	15.64$_{\pm 0.86}$
GraphPrompt	73.09$_{\pm 2.13}$	10.73$_{\pm 1.05}$	58.65$_{\pm 3.19}$	37.92$_{\pm 1.09}$	26.78$_{\pm 1.71}$	55.27$_{\pm 11.92}$	20.70$_{\pm 2.19}$	40.47$_{\pm 1.21}$	**29.14**$_{\pm 0.68}$
GraphPrompt+	73.94$_{\pm 3.22}$	5.63$_{\pm 1.63}$	55.88$_{\pm 2.83}$	**39.52**$_{\pm 0.94}$	30.06$_{\pm 2.30}$	53.15$_{\pm 17.98}$	8.10$_{\pm 4.09}$	40.60$_{\pm 1.32}$	28.91$_{\pm 2.47}$
ProNoG	82.84$_{\pm 0.91}$	**36.83**$_{\pm 0.67}$	54.23$_{\pm 0.62}$	36.76$_{\pm 1.14}$	30.89$_{\pm 2.68}$	49.95$_{\pm 11.24}$	**4.94**$_{\pm 0.54}$	40.26$_{\pm 0.90}$	24.00$_{\pm 0.40}$
Pretraining: ContextPred									
Full fine-tuning	**88.41**$_{\pm 1.39}$	30.46$_{\pm 15.09}$	**58.28**$_{\pm 3.73}$	29.39$_{\pm 3.09}$	34.23$_{\pm 0.81}$	49.04$_{\pm 11.34}$	24.04$_{\pm 5.46}$	50.89$_{\pm 0.41}$	21.49$_{\pm 3.51}$
All-in-One	87.18$_{\pm 1.26}$	30.91$_{\pm 0.98}$	52.13$_{\pm 3.09}$	34.00$_{\pm 1.65}$	**36.99**$_{\pm 1.34}$	56.74$_{\pm 10.53}$	101.71$_{\pm 0.87}$	**56.20**$_{\pm 0.79}$	**31.72**$_{\pm 0.68}$
EdgePrompt	75.11$_{\pm 5.85}$	4.87$_{\pm 3.68}$	49.56$_{\pm 4.42}$	27.70$_{\pm 4.50}$	29.35$_{\pm 4.06}$	57.74$_{\pm 7.54}$	108.43$_{\pm 0.04}$	49.39$_{\pm 0.05}$	20.14$_{\pm 1.83}$
EdgePrompt+	66.25$_{\pm 4.81}$	9.85$_{\pm 2.94}$	53.42$_{\pm 4.01}$	24.53$_{\pm 4.76}$	30.30$_{\pm 5.44}$	54.34$_{\pm 10.17}$	108.34$_{\pm 0.08}$	49.49$_{\pm 0.15}$	20.12$_{\pm 4.04}$
GPF	86.33$_{\pm 1.25}$	**34.48**$_{\pm 2.96}$	51.68$_{\pm 1.54}$	25.42$_{\pm 1.17}$	32.65$_{\pm 1.41}$	55.29$_{\pm 8.68}$	111.72$_{\pm 2.98}$	50.67$_{\pm 0.24}$	17.28$_{\pm 1.68}$
GPF+	86.88$_{\pm 0.90}$	33.69$_{\pm 2.21}$	51.80$_{\pm 2.20}$	26.10$_{\pm 1.62}$	32.54$_{\pm 1.55}$	50.68$_{\pm 3.44}$	110.37$_{\pm 0.04}$	50.69$_{\pm 0.35}$	21.27$_{\pm 4.32}$
GPPT	83.71$_{\pm 2.54}$	1.09$_{\pm 0.67}$	53.84$_{\pm 3.53}$	33.27$_{\pm 1.77}$	35.26$_{\pm 1.54}$	**66.83**$_{\pm 9.07}$	107.01$_{\pm 0.04}$	49.34$_{\pm 1.05}$	20.11$_{\pm 2.18}$
GraphPrompt	80.75$_{\pm 1.78}$	6.93$_{\pm 1.01}$	56.72$_{\pm 4.09}$	38.00$_{\pm 1.88}$	29.06$_{\pm 2.05}$	41.24$_{\pm 10.11}$	45.29$_{\pm 0.26}$	40.56$_{\pm 1.19}$	26.85$_{\pm 0.30}$
GraphPrompt+	84.34$_{\pm 1.06}$	17.18$_{\pm 2.68}$	57.49$_{\pm 2.09}$	**38.71**$_{\pm 1.80}$	30.88$_{\pm 3.79}$	50.37$_{\pm 10.14}$	36.92$_{\pm 0.63}$	42.00$_{\pm 2.12}$	28.58$_{\pm 1.21}$
ProNoG	82.84$_{\pm 0.93}$	31.24$_{\pm 0.70}$	55.90$_{\pm 0.65}$	36.99$_{\pm 1.16}$	28.24$_{\pm 0.85}$	43.93$_{\pm 3.68}$	**13.57**$_{\pm 0.67}$	40.22$_{\pm 0.85}$	22.52$_{\pm 0.42}$
Pretraining: DGI									
Full fine-tuning	88.03$_{\pm 1.32}$	38.51$_{\pm 0.88}$	51.10$_{\pm 1.60}$	31.97$_{\pm 4.97}$	34.92$_{\pm 1.27}$	49.58$_{\pm 9.00}$	33.40$_{\pm 13.94}$	51.24$_{\pm 0.49}$	14.07$_{\pm 6.25}$
All-in-One	**88.25**$_{\pm 0.30}$	**40.19**$_{\pm 0.40}$	49.61$_{\pm 3.29}$	32.53$_{\pm 0.75}$	**37.65**$_{\pm 0.71}$	49.54$_{\pm 13.69}$	90.50$_{\pm 0.26}$	**54.00**$_{\pm 0.57}$	10.99$_{\pm 0.16}$
EdgePrompt	29.11$_{\pm 17.34}$	8.20$_{\pm 5.94}$	52.61$_{\pm 3.24}$	26.66$_{\pm 5.82}$	29.15$_{\pm 4.41}$	54.15$_{\pm 27.81}$	107.48$_{\pm 0.04}$	51.81$_{\pm 0.58}$	10.72$_{\pm 0.74}$
EdgePrompt+	30.66$_{\pm 19.40}$	8.03$_{\pm 3.16}$	51.33$_{\pm 4.33}$	27.06$_{\pm 4.31}$	29.56$_{\pm 3.89}$	40.98$_{\pm 32.28}$	107.50$_{\pm 0.05}$	51.79$_{\pm 0.60}$	11.26$_{\pm 0.00}$
GPF	71.53$_{\pm 14.95}$	32.57$_{\pm 3.76}$	51.15$_{\pm 1.92}$	25.09$_{\pm 0.14}$	30.77$_{\pm 5.51}$	48.30$_{\pm 20.54}$	107.79$_{\pm 0.03}$	52.44$_{\pm 0.21}$	10.44$_{\pm 0.75}$
GPF+	57.07$_{\pm 25.72}$	29.31$_{\pm 1.69}$	51.78$_{\pm 1.82}$	27.29$_{\pm 4.91}$	34.33$_{\pm 3.29}$	44.83$_{\pm 28.65}$	107.79$_{\pm 0.03}$	51.87$_{\pm 1.23}$	9.98$_{\pm 0.87}$
GPPT	54.41$_{\pm 14.94}$	4.47$_{\pm 3.29}$	51.19$_{\pm 1.73}$	34.30$_{\pm 2.34}$	33.98$_{\pm 1.79}$	49.29$_{\pm 13.88}$	103.05$_{\pm 0.01}$	50.69$_{\pm 1.60}$	9.85$_{\pm 1.13}$
GraphPrompt	75.44$_{\pm 1.20}$	10.15$_{\pm 0.30}$	49.28$_{\pm 1.51}$	**38.81**$_{\pm 0.90}$	28.90$_{\pm 2.08}$	46.11$_{\pm 17.65}$	19.71$_{\pm 0.52}$	42.26$_{\pm 1.38}$	**16.92**$_{\pm 0.61}$
GraphPrompt+	80.85$_{\pm 3.62}$	20.56$_{\pm 1.46}$	50.40$_{\pm 2.45}$	36.86$_{\pm 1.59}$	31.47$_{\pm 3.49}$	**55.36**$_{\pm 8.78}$	31.04$_{\pm 1.18}$	43.02$_{\pm 1.45}$	16.76$_{\pm 1.01}$
ProNoG	62.37$_{\pm 2.86}$	18.41$_{\pm 0.77}$	**55.13**$_{\pm 1.29}$	27.87$_{\pm 3.02}$	30.68$_{\pm 1.17}$	54.36$_{\pm 12.76}$	**14.30**$_{\pm 1.16}$	40.63$_{\pm 1.08}$	16.47$_{\pm 0.44}$
Pretraining: EdgePred									
Full fine-tuning	**85.45**$_{\pm 0.43}$	**38.07**$_{\pm 1.61}$	55.65$_{\pm 2.55}$	25.52$_{\pm 1.98}$	32.70$_{\pm 3.66}$	49.51$_{\pm 16.90}$	36.65$_{\pm 20.82}$	52.35$_{\pm 1.07}$	10.22$_{\pm 0.93}$
All-in-One	31.96$_{\pm 3.65}$	23.40$_{\pm 2.71}$	50.91$_{\pm 4.74}$	36.58$_{\pm 2.83}$	**35.49**$_{\pm 1.22}$	47.87$_{\pm 18.53}$	101.71$_{\pm 0.87}$	**56.28**$_{\pm 0.38}$	**25.67**$_{\pm 0.32}$
EdgePrompt	9.44$_{\pm 0.69}$	1.33$_{\pm 1.80}$	51.93$_{\pm 0.61}$	31.39$_{\pm 4.71}$	25.63$_{\pm 5.09}$	52.44$_{\pm 10.36}$	108.43$_{\pm 0.04}$	52.94$_{\pm 0.34}$	9.88$_{\pm 0.03}$
EdgePrompt+	25.62$_{\pm 7.18}$	6.99$_{\pm 4.08}$	51.44$_{\pm 1.75}$	31.85$_{\pm 2.08}$	29.26$_{\pm 5.60}$	51.70$_{\pm 23.20}$	108.34$_{\pm 0.08}$	53.10$_{\pm 0.58}$	22.29$_{\pm 2.45}$
GPF	22.76$_{\pm 4.12}$	20.00$_{\pm 5.20}$	52.65$_{\pm 1.11}$	25.29$_{\pm 0.66}$	28.81$_{\pm 2.88}$	55.65$_{\pm 13.12}$	111.49$_{\pm 3.11}$	52.73$_{\pm 0.63}$	13.47$_{\pm 4.92}$
GPF+	24.24$_{\pm 6.29}$	22.97$_{\pm 5.65}$	50.99$_{\pm 1.60}$	27.32$_{\pm 1.84}$	29.79$_{\pm 1.08}$	51.77$_{\pm 12.77}$	110.37$_{\pm 0.04}$	52.80$_{\pm 0.61}$	12.02$_{\pm 2.15}$
GPPT	20.10$_{\pm 8.70}$	3.82$_{\pm 5.01}$	49.94$_{\pm 0.11}$	34.33$_{\pm 3.10}$	25.17$_{\pm 3.72}$	55.18$_{\pm 9.57}$	106.76$_{\pm 0.04}$	50.27$_{\pm 0.48}$	10.56$_{\pm 1.17}$
GraphPrompt	19.63$_{\pm 1.15}$	5.65$_{\pm 1.86}$	51.03$_{\pm 1.35}$	**38.03**$_{\pm 0.92}$	30.48$_{\pm 2.42}$	48.68$_{\pm 11.58}$	57.83$_{\pm 33.08}$	42.76$_{\pm 1.49}$	22.28$_{\pm 0.71}$
GraphPrompt+	33.18$_{\pm 3.52}$	13.12$_{\pm 1.35}$	52.91$_{\pm 2.25}$	34.78$_{\pm 3.25}$	29.06$_{\pm 4.15}$	50.13$_{\pm 8.83}$	36.92$_{\pm 0.63}$	43.69$_{\pm 1.70}$	19.27$_{\pm 0.98}$
ProNoG	39.30$_{\pm 0.54}$	7.42$_{\pm 4.25}$	**56.13**$_{\pm 0.72}$	29.57$_{\pm 2.99}$	29.71$_{\pm 2.90}$	**61.66**$_{\pm 12.23}$	**13.56**$_{\pm 0.74}$	41.30$_{\pm 0.97}$	22.87$_{\pm 0.53}$
Pretraining: GraphCL									
Full fine-tuning	**83.00**$_{\pm 3.61}$	21.15$_{\pm 2.22}$	**55.44**$_{\pm 6.74}$	25.72$_{\pm 1.38}$	35.26$_{\pm 1.28}$	38.43$_{\pm 18.70}$	62.89$_{\pm 24.16}$	52.42$_{\pm 1.17}$	**18.70**$_{\pm 7.91}$
All-in-One	77.80$_{\pm 2.32}$	30.50$_{\pm 1.93}$	53.05$_{\pm 1.34}$	25.16$_{\pm 2.32}$	**37.86**$_{\pm 1.00}$	47.56$_{\pm 18.66}$	107.23$_{\pm 0.05}$	**53.59**$_{\pm 1.20}$	9.72$_{\pm 0.37}$
EdgePrompt	57.40$_{\pm 4.60}$	17.41$_{\pm 0.83}$	53.35$_{\pm 2.92}$	24.94$_{\pm 0.00}$	35.99$_{\pm 1.58}$	52.44$_{\pm 6.82}$	113.76$_{\pm 3.05}$	51.43$_{\pm 0.80}$	10.09$_{\pm 0.66}$
EdgePrompt+	62.93$_{\pm 5.84}$	21.06$_{\pm 0.99}$	50.52$_{\pm 3.93}$	24.91$_{\pm 0.06}$	36.44$_{\pm 0.81}$	51.49$_{\pm 10.11}$	114.87$_{\pm 3.04}$	51.44$_{\pm 1.02}$	10.14$_{\pm 0.63}$
GPF	67.20$_{\pm 11.05}$	12.97$_{\pm 3.33}$	52.06$_{\pm 2.56}$	24.84$_{\pm 0.70}$	34.30$_{\pm 3.11}$	**56.62**$_{\pm 5.71}$	117.10$_{\pm 0.03}$	51.44$_{\pm 1.23}$	12.30$_{\pm 5.35}$
GPF+	70.50$_{\pm 4.83}$	11.54$_{\pm 2.19}$	49.14$_{\pm 3.21}$	25.90$_{\pm 1.48}$	35.03$_{\pm 2.65}$	52.70$_{\pm 15.87}$	112.17$_{\pm 0.07}$	52.12$_{\pm 0.65}$	10.33$_{\pm 1.02}$
GPPT	73.44$_{\pm 2.94}$	11.00$_{\pm 2.61}$	51.38$_{\pm 3.69}$	34.23$_{\pm 2.50}$	33.76$_{\pm 0.92}$	42.58$_{\pm 15.81}$	110.55$_{\pm 0.01}$	50.49$_{\pm 1.67}$	9.93$_{\pm 0.28}$
GraphPrompt	74.78$_{\pm 2.90}$	14.29$_{\pm 1.24}$	49.32$_{\pm 1.11}$	38.28$_{\pm 1.46}$	30.30$_{\pm 2.31}$	45.66$_{\pm 9.03}$	84.05$_{\pm 0.13}$	45.59$_{\pm 1.46}$	16.25$_{\pm 1.82}$
GraphPrompt+	80.75$_{\pm 1.92}$	15.11$_{\pm 0.95}$	51.67$_{\pm 2.65}$	**38.33**$_{\pm 1.19}$	31.40$_{\pm 4.99}$	52.68$_{\pm 13.15}$	33.00$_{\pm 2.84}$	45.36$_{\pm 1.32}$	17.86$_{\pm 0.88}$
ProNoG	80.55$_{\pm 1.23}$	**30.51**$_{\pm 0.36}$	53.92$_{\pm 0.61}$	36.56$_{\pm 2.49}$	31.81$_{\pm 0.66}$	46.53$_{\pm 12.63}$	**13.55**$_{\pm 0.37}$	45.69$_{\pm 1.34}$	16.18$_{\pm 0.19}$
Pretraining: InfoGraph									
Full fine-tuning	**84.48**$_{\pm 2.48}$	17.56$_{\pm 2.34}$	**54.18**$_{\pm 1.73}$	25.87$_{\pm 1.43}$	33.21$_{\pm 1.53}$	52.56$_{\pm 6.79}$	45.94$_{\pm 33.32}$	51.80$_{\pm 1.17}$	16.67$_{\pm 6.26}$
All-in-One	81.94$_{\pm 1.02}$	7.03$_{\pm 2.32}$	50.31$_{\pm 1.54}$	**39.04**$_{\pm 1.81}$	**36.87**$_{\pm 1.60}$	52.41$_{\pm 7.80}$	103.81$_{\pm 0.32}$	**56.55**$_{\pm 1.09}$	9.90$_{\pm 0.00}$
EdgePrompt	51.74$_{\pm 2.80}$	1.88$_{\pm 1.41}$	52.48$_{\pm 0.21}$	33.37$_{\pm 1.41}$	28.52$_{\pm 1.37}$	55.55$_{\pm 13.57}$	111.26$_{\pm 3.24}$	51.20$_{\pm 0.97}$	9.84$_{\pm 0.02}$
EdgePrompt+	59.12$_{\pm 2.03}$	4.55$_{\pm 2.11}$	52.54$_{\pm 0.25}$	31.04$_{\pm 1.94}$	27.37$_{\pm 0.48}$	53.01$_{\pm 12.80}$	109.86$_{\pm 0.06}$	51.28$_{\pm 0.99}$	12.61$_{\pm 4.02}$
GPF	75.99$_{\pm 0.82}$	7.51$_{\pm 4.53}$	52.97$_{\pm 1.06}$	25.54$_{\pm 0.39}$	30.40$_{\pm 3.87}$	47.85$_{\pm 12.44}$	117.07$_{\pm 0.02}$	51.80$_{\pm 1.19}$	11.87$_{\pm 3.79}$
GPF+	65.04$_{\pm 14.95}$	5.37$_{\pm 3.47}$	53.68$_{\pm 0.77}$	27.09$_{\pm 1.77}$	32.99$_{\pm 1.44}$	51.61$_{\pm 3.60}$	111.25$_{\pm 0.12}$	52.48$_{\pm 0.96}$	11.85$_{\pm 3.75}$
GPPT	68.86$_{\pm 5.47}$	5.55$_{\pm 6.02}$	49.86$_{\pm 0.10}$	35.04$_{\pm 1.60}$	25.66$_{\pm 5.35}$	**57.40**$_{\pm 19.11}$	109.31$_{\pm 0.03}$	50.42$_{\pm 0.79}$	9.99$_{\pm 0.39}$
GraphPrompt	73.49$_{\pm 1.59}$	4.72$_{\pm 1.91}$	51.34$_{\pm 1.53}$	37.06$_{\pm 1.48}$	31.01$_{\pm 4.88}$	52.46$_{\pm 10.09}$	68.60$_{\pm 0.10}$	44.03$_{\pm 1.43}$	18.62$_{\pm 1.32}$
GraphPrompt+	77.54$_{\pm 3.58}$	3.85$_{\pm 2.06}$	53.29$_{\pm 0.88}$	37.77$_{\pm 1.53}$	27.66$_{\pm 3.09}$	44.30$_{\pm 10.50}$	40.78$_{\pm 2.47}$	44.33$_{\pm 1.39}$	**19.67**$_{\pm 1.83}$
ProNoG	63.42$_{\pm 1.19}$	**24.29**$_{\pm 0.65}$	52.18$_{\pm 0.53}$	37.37$_{\pm 1.41}$	30.29$_{\pm 0.91}$	43.01$_{\pm 10.96}$	**13.41**$_{\pm 0.20}$	43.17$_{\pm 1.32}$	17.71$_{\pm 0.41}$