PAGPL: Privacy-Aware Graph Prompt Learning Scheme via Adaptive Perturbation-Estimated Topology Recovery

Ju Jia[1,2], Jiansen Song[1], Jingxuan Yu[1], Jiabao Guo[3*], Xiaoshuang Jia[4*], Di Wu[5], Yali Yuan[1], Guang Cheng[1]

[1]School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China
[2]Engineering Research Center of Blockchain Application, Supervision and Management (Southeast University), Ministry of Education, China
[3]School of Computer and Information, Hefei University of Technology, Hefei 230009, China
[4]School of Information Resource Management, Renmin University of China, Beijing 100872, China
[5]School of Computing, Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia
{jiaju,sjs01,yujingxuan24,yaliyuan,chengguang}@seu.edu.cn, garbo_guo@hfut.edu.cn, jiaxs1219@ruc.edu.cn, d.wu@latrobe.edu.au

Abstract

Graph prompt learning (GPL) serves as a crucial framework for mitigating the knowledge transfer by reconciling the substantial mismatch between pre-training models and downstream tasks. However, prevalent GPL paradigm fail to accommodate graph data affected by privacy-induced noise. Specifically, 1) GPL typically relies on the stability of original graph structures for the design of effective prompt templates; 2) the construction of prompts lacks explicit guidance to suppress noise introduced by privacy perturbations; 3) prompt optimization on single disturbed graphs can easily lead to overfitting to noise patterns. To address these issues, we propose a novel privacy-aware graph prompt learning (PAGPL) scheme, which alleviates spurious clues caused by privacy noise injection. Initially, an adaptive structure-wise Bayesian estimation is applied to reconstruct the privacy-perturbed graphs. Subsequently, to suppress the impact of residual perturbation, a noise-resilient prompt generation is employed to filter unreliable structural and signals. Ultimately, we incorporate a multi-view-based progressive privacy consistency to promote the robustness of prompts against the semantic misalignment while improving the task-specific consistency. The experimental results reveal that our scheme outperforms state-of-the-art (SOTA) GPL approaches with a 10%–60% improvement in accuracy under various real-world privacy-perturbed scenarios.

Code — https://github.com/Senspecial/PAGPL

Introduction

The knowledge transfer (Fang et al. 2025b) from pre-training models to downstream tasks poses challenges due to discrepancies in data characteristics, feature distributions, and task-specific objectives (Li et al. 2024). This necessitates the development of fine-tuning or adaptation strategies to bridge the gap between desired source domains and real-world target domains. In the traditional GNN pre-training with fine-tuning paradigm (Jin et al. 2020; Sun et al. 2024), these strategies often struggle to generalize across mismatched semantic spaces and non-stationary graph distributions (Sun et al. 2023b; Liu et al. 2023a). Therefore, graph prompt learning (GPL) (Sun et al. 2023b; Liu et al. 2023b; Sun et al. 2023a; Fang et al. 2023, 2022; Sun et al. 2022; Yu et al. 2024; Jia et al. 2025c) offers an innovative solution that to this challenge. GPL adopts a two-stage framework where GNNs are first pre-trained (Lu et al. 2021; Yu et al. 2025) to capture general graph patterns, followed by the integration of the prompts customized for downstream tasks (Ge et al. 2023; Chen et al. 2025; Zhang et al. 2025). GPL achieves efficiency by freezing the base GNN encoder and introducing light-weight prompts into downstream tasks. Simultaneously, it ensures representational reliability, as the prompts can approximate a wide range of desired structural transformations. This paradigm reduces the reliance on extensive parameter fine-tuning and improves transfer efficiency.

Privacy preservation in graph representation learning is increasingly critical, as advanced privacy inference attacks have demonstrated the ability to extract sensitive attributes from learned graph embeddings (Grover and Leskovec

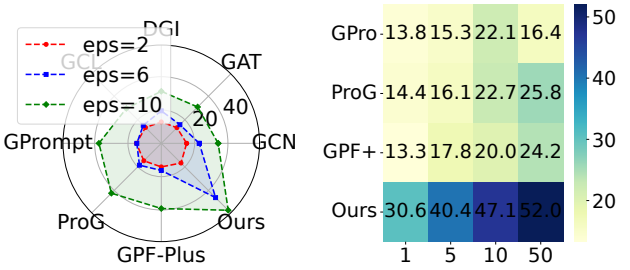

Figure 1: Performance (%) under privacy perturbation and few-shot settings. Left (Radar chart): our PAGPL shows superior robustness, with the excellent performance across varying privacy budgets. Right (Heatmap): our PAGPL maintains a salient advantage in few-shot scenarios.

*Xiaoshuang Jia and Jiabao Guo are corresponding authors.

2016; Perozzi, Al-Rfou, and Skiena 2014; Tang et al. 2015; Jia et al. 2023, 2025b, 2026). These privacy threats are not confined to traditional graph learning settings but can easily generalize to GPL scenarios. Existing privacy-preserving solutions for graph learning primarily fall into two categories: graph federated learning (GFL) and graph differential privacy (GDP). While GFL protects raw data by distributing training, its infrastructure demands and large-scale assumptions render it unsuitable for the few-shot (Satorras and Estrach 2018; Wang et al. 2020; Zhang, Ding, and Li 2025; Fang, Easwaran, and Genest 2025), downstream-oriented GPL paradigm. In contrast, GDP offers a more flexible alternative by injecting randomized noise directly into the data, which aligns well with the GPL setting where individual users handle small-scale inputs and demand local control over privacy guarantees.

In response, we reveal that the feasibility of GPL will significantly descend when privacy-induced noise is implemented. As shown in Figure 1, previous approaches suffers an accuracy drop of over 30%–60% compared to clean settings. Specifically, there are three shortcomings that lead to such infeasibility: 1) conventional GPL often considers stable and harmless graph structures, which makes the design of prompt highly sensitive to perturbations in topology or connectivity; 2) current prompt approaches lack dedicated designs to identify or mitigate noise-induced patterns. As a result, they often treat spurious signals caused by privacy perturbations as valid input features, which leads to the generation of misleading prompts; 3) prompt tuning is performed on perturbed task instances without accounting for the true underlying semantics. Therefore, the generated prompts tend to overfit to artifacts induced by privacy noise, which fails to bridge the pre-trained model with the actual downstream tasks.

To address the above issues, we propose a novel privacy-aware graph prompt learning (PAGPL) , which aims to resolve the mismatch between privacy-induced topological distortions and high-quality graph representations in real-world applications. First, an adaptive structure-wise Bayesian estimation is utilized to reconstruct the graph samples. Building upon these reconstructed structures, we design a learnable prompt that dynamically perceives and responds to noise patterns introduced by privacy perturbations. This component selectively filters out irrelevant or noisy signals, which preserves task-relevant representations and enhances the stability of prompt construction in perturbed environments. Ultimately, to further enhance the robustness of the learnable prompts, we develop a progressive privacy consistency into the optimization process, which encourages semantic alignment across multiple noisy graph views. Notably, our experiments demonstrate that PAGPL effectively handles both self supervision scarcity and external privacy-induced noise, which can produce accurate and reliable graph representations. The main contribution of this paper can be summarized as follows:

1) We propose the PAGPL, a novel privacy-aware GPL scheme that is designed to generate semantically aligned and resilient prompts by adapting to the underlying distorted structures. To the best of our knowledge, this is the first work that investigates the feasibility and stability of GPL under privacy-perturbed scenarios.
2) We reveal that existing GPL paradigm lacks capability to distinguish between task-relevant informative patterns and privacy-induced spurious signals, which leads to the construction of misleading prompts. The proposed scheme leverages topology-oriented modeling to mitigate the structural uncertainty introduced by data perturbations and employs light-weight prompts jointly optimized for task relevance and privacy consistency.
3) We conduct extensive experiments on five real-world benchmark datasets to evaluate the generalization capability of privacy-aware GPL. The results demonstrate that our scheme improves the accuracy of GPL by 10% – 60% compared to SOTA baselines.

Related Work

Privacy-Preserving Graph Data

Privacy-preserving graph data is typically generated by perturbing original graphs through element modifications and feature permutations (Fu et al. 2023; Fang et al. 2025a; Jia et al. 2025a). These data are widely used in graph representation learning and have recently been extended to GPL scenarios under privacy constraints. Traditional methods such as federated learning (Ye et al. 2025; Li et al. 2020; Zhang et al. 2021) and graph anonymization (Liu and Terzi 2008; Yazdanjue et al. 2025) are not well-suited for downstream GPL tasks because federated learning requires large-scale coordination and continuous communication (Nguyen et al. 2021), while graph anonymization is designed for static data publishing. In contrast, local differential privacy (LDP) (Shen et al. 2025; Dwork et al. 2006) offers a more suitable privacy paradigm for downstream GPL tasks. It allows each user to independently perturb their local graph data before any model interaction, which removes the need for centralized coordination or distributed training protocols. The inherent decentralization and local controllability of LDP make it well-suited for integration with light-weight, few-shot adaptation settings in GPL.

Graph Prompt Learning

GPL aims to develop efficient prompt modules with sufficient expressive power. At the same time, it is required that the pre-trained GNN remain frozen. This allows the model to retain its learned knowledge while simultaneously optimizing its performance for specific downstream tasks. GPrompt (Liu et al. 2023b) unified pre-training and downstream tasks via subgraph similarity, while ProG (Sun et al. 2023a) advanced multi-task prompting through token, structure, and insertion design. GPF (Fang et al. 2023) is also a universal prompt tuning method that injects learnable prompts into the input feature space of graphs. However, such designs neglect the practical difficulty posed by noise, under which structural noise and feature distortion can significantly impair the reliability and generalization of prompts. In contrast, our PAGPL explicitly incorporates prompts that maintain robustness and adaptability under privacy perturbation.

Bayesian Estimation in Graphs

Bayesian estimation (Zyphur and Oswald 2015; Seliem et al. 2025) provides a coherent framework for parameter inference and decision-making under uncertainty and emphasizes its advantages in decision-making, posterior prediction, and few-shot modeling. Bayesian estimation has been proposed (Wang et al. 2021) to incorporate prior knowledge and to model edge weights with uncertainty analysis. Blink (Zhu, Tan, and Xiao 2023) employed Bayesian inference to reconstruct the graph structures that using noisy degrees as prior and perturbed adjacency lists as evidence to compute posterior link probabilities. In adaptive structure-wise recovery, Bayesian estimation is applied to recover latent graph structures from privacy-perturbed data by learning a Beta prior over link probabilities.

Preliminaries

Notation

Given a privacy-perturbed graph $G = (V, X)$, where V denotes the node set with $|V|$ nodes, and each node $v \in V$ is associated with a feature vector $X_v \in \mathbb{R}^{1\times d}$, where d is the feature dimension. The edge set E contains $|E|$ edges, which can be represented by the adjacency matrix $A \in \mathbb{R}^{|V|\times|V|}$, where $A_{u,v} = 1$ if and only if $(u, v) \in E$. Let $X \in \mathbb{R}^{|V|\times d}$ denote the node feature matrix, and let $H \in \mathbb{R}^{|V|\times d'}$ denote the node embedding matrix, where d' is the embedding dimension. In addition, we denote a set of graphs as $G = \{G_1, G_2, \ldots, G_N\}$.

Multi-Perspective Prompts

Our prompt design is grounded in subgraph-based representations, from which different tasks can benefit from task-specific aggregation strategies. A task instance can be a node $v \in V$, an edge $e = (u, v) \in E$, or an entire graph G. Each instance is associated with a contextual subgraph $S_x = (V(S_x), E(S_x))$, which serves as the structural signal for prompt construction and task-specific representation learning.

Node-level prompts. For node-level tasks where $x = v \in V$, the subgraph S_v is defined as the δ-hop neighborhood around v:

$$V(S_v) = \{u \in V \mid d(u, v) \le \delta\}, \tag{1}$$
$$E(S_v) = \{(i, j) \in E \mid i, j \in V(S_v)\}, \tag{2}$$

where $d(u, v)$ is the shortest path distance in graph G, and $\delta \in \mathbb{N}$ is a predefined neighborhood radius.

Edge-level prompts. For edge-level tasks where $x = e = (u, v) \in E$, the subgraph S_e aggregates the δ-hop neighborhoods of both endpoints:

$$V(S_e) = \{w \in V \mid d(w, u) \le \delta \text{ or } d(w, v) \le \delta\}, \tag{3}$$
$$E(S_e) = \{(i, j) \in E \mid i, j \in V(S_e)\}. \tag{4}$$

This subgraph encodes the interaction context of the edge, and is suitable for link prediction.

Graph-level prompts. For graph-level tasks where $x = G$, the associated subgraph is the entire graph itself:

$$S_G = G. \tag{5}$$

This formulation naturally supports graph classification that require access to global structural properties.

Methodology

Overview of Our Scheme

The architecture and workflow of the PAGPL are illustrated in Figure 2. The ultimate goal of the proposed PAGPL is to achieve notably robust and semantically aligned prompt through adaptive perturbation-estimated topology recovery. It can be divided into the following three main stages: 1) approximate reconstruction of graph structures. The approximate reconstruction of graph structures via adaptive Bayesian estimation aims to enhance task performance by restoring data utility under privacy constraints; 2) prompt construction of residual noise perception. Residual noise-sensitive prompt construction enhances the robustness of the model to privacy perturbations by extracting informative signals; 3) progressive prompt optimization with privacy consistency. By constructing multi-views with different perturbations for contrastive learning, the prompt is encouraged to maintain consistency across multiple views.

Approximate Reconstruction of Graph Structures

For a node v_i in the graph G, its corresponding adjacency links have been processed for privacy protection. Our first step is to estimate the probability of the existence of each link, using it as the prior probability. We use a learnable $\boldsymbol{\beta}$-model to estimate the prior probabilities. The parameter $\boldsymbol{\beta}$ vector is initialized as: $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_n] \in \mathbb{R}^n$. Each $\boldsymbol{\beta}_i$ represents the importance or connectivity propensity of node v_i. The link probability p_{ij} between nodes v_i and v_j is modeled as:

$$p_{ij} = \frac{\exp(\boldsymbol{\beta}_i + \boldsymbol{\beta}_j)}{1 + \exp(\boldsymbol{\beta}_i + \boldsymbol{\beta}_j)}. \tag{6}$$

Node embeddings are extracted via a frozen, pre-trained GNN. Based on these embeddings, we compute a similarity matrix $\mathbf{S} = \text{sim}(\mathbf{H}^{(i)}, \mathbf{H}^{(j)})$, which captures pairwise similarities between nodes and serves as prior structural information. To interpret these similarity scores probabilistically, a Sigmoid function is applied to map $\mathbf{S}$ into a prior probability matrix $\hat{\mathbf{P}}$. Each entry in $\hat{\mathbf{P}}$ represents the estimated likelihood of an edge existing between two nodes, and is computed as $\hat{\mathbf{P}} = \sigma(\mathbf{S}) = \frac{1}{1+\exp(-\mathbf{S})}$.

In conformity with the observation matrix, the maximum likelihood estimation (MLE) predicts the link probability p_{ij} and obtains a prior probability matrix $\mathbf{P}$. As a result, we can optimize the value of $\boldsymbol{\beta}$ by the MLE, which is described as a logarithmic equation:

$$\begin{aligned}\mathcal{L}(\beta) = \sum_{i,j} \Bigg(& \hat{p}_{ij} \left(-\log\left(1 + \exp\left(-\left(\boldsymbol{\beta}_i + \boldsymbol{\beta}_j\right)\right)\right)\right) \\ & + \left(1 - \hat{p}_{ij}\right) \left(-\log\left(1 + \exp\left(\boldsymbol{\beta}_i + \boldsymbol{\beta}_j\right)\right)\right) \Bigg).\end{aligned} \tag{7}$$

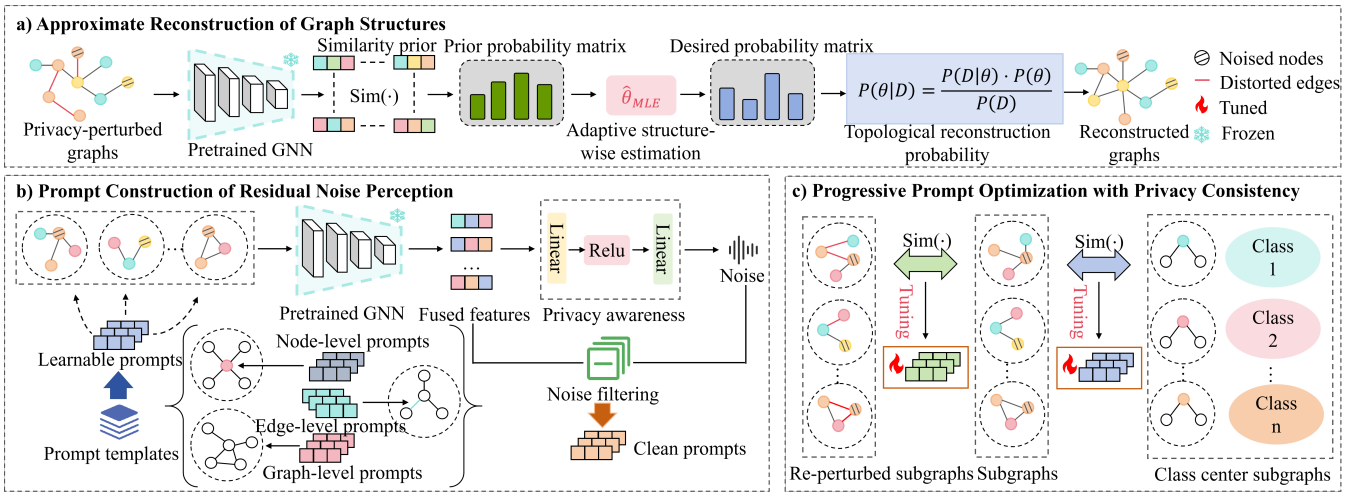

Figure 2: The overall framework of the proposed PAGPL.

Once the prior probability matrix $\mathbf{P}$ is computed using the learnable $\boldsymbol{\beta}$-model, we proceed to refine these estimates by incorporating the noisy observed adjacency matrix as additional evidence. For each potential link between nodes v_i and v_j, the observed noisy adjacency matrix provides two bits $\tilde{A}_{ij}$ and $\tilde{A}_{ji}$, indicating the presence or absence of the link. Due to the privacy protection mechanism, these bits may not directly correspond to the true existence of the edge. As is well known, the existence of an edge can be represented as whether it flips or not. We define the probability of edge existence as p_f, and depends on the privacy budget ϵ used in the privacy-perturbed mechanism we can obtain $p_f = \frac{1}{1+\exp(\epsilon)}$, this parameter captures the extent of noise introduced by the privacy mechanism.

The likelihood of observing the noisy adjacency bits $\tilde{A}_{ij}$ and $\tilde{A}_{ji}$ is calculated under two key hypotheses:

$$q_{ij} = \begin{cases} p_f^2, & (\tilde{A}_{ij}, \tilde{A}_{ji}) = (0,0), \\ p_f(1-p_f), & (\tilde{A}_{ij}, \tilde{A}_{ji}) = (0,1) \text{ or } (1,0), \\ (1-p_f)^2, & (\tilde{A}_{ij}, \tilde{A}_{ji}) = (1,1), \end{cases} \tag{8}$$

$$q'_{ij} = \begin{cases} (1-p_f)^2, & (\tilde{A}_{ij}, \tilde{A}_{ji}) = (0,0), \\ p_f(1-p_f), & (\tilde{A}_{ij}, \tilde{A}_{ji}) = (0,1) \text{ or } (1,0), \\ p_f^2, & (\tilde{A}_{ij}, \tilde{A}_{ji}) = (1,1). \end{cases} \tag{9}$$

where q_{ij} represents the likelihood of observing the noisy adjacency bits $\tilde{A}_{ij}$ and $\tilde{A}_{ji}$ assuming that an edge exists between nodes v_i and v_j, and q'_{ij} represents the likelihood of observing the noisy adjacency bits $\tilde{A}_{ij}$ and $\tilde{A}_{ji}$ assuming that no edge exists between nodes v_i and v_j. In Bayesian inference, q_{ij} and q'_{ij} are used to update the prior probabilities of edges into posterior probabilities:

$$P_{ij} = \frac{q_{ij} \cdot p_{ij}}{q_{ij} \cdot p_{ij} + q'_{ij} \cdot (1 - p_{ij})}. \tag{10}$$

The posterior probability matrix of graph G is applied to reconstruct a refined graph G_t, either by retaining probabilities as edge weights or selecting top-k edges.

Prompt Construction of Residual Noise Perception

Our prompt template design is based on transforming task instances into subgraph similarity learning problems, which enables the generation of diverse prompts and the construction of a universal graph prompt template. Let $\boldsymbol{p}$ denote a learnable prompt vector, we treat each node v in the graph as a context subgraph with a δ-hop neighborhood centered around itself, which is combined with the prompt vector. Then, a prompt-assisted readout operation is applied to obtain the combined subgraph representation $\boldsymbol{s}_v$. To further mitigate the influence of noisy or unstable features, we introduce a denoising module f_θ to refine the prompt-guided subgraph representations $\boldsymbol{r}_v = \boldsymbol{p} \odot \boldsymbol{h}_v - f_\theta(\boldsymbol{p} \odot \boldsymbol{h}_v)$. Here a MLP is used to estimate the noise component in the prompt-weighted features:

$$f_\theta = \text{MLP}(\boldsymbol{p} \odot \boldsymbol{h}_v), \tag{11}$$

where $\odot$ denotes the element-wise multiplication, the core idea is to treat the raw prompt-weighted node features as potentially noisy and apply a learnable residual correction mechanism to suppress perturbation-induced biases. This residual denoising formulation allows the model to preserve useful local information while adaptively removing unstable signals caused by perturbation or overfitting to noisy neighborhoods. The denoised representations are then aggregated to obtain the final subgraph-level representation:

$$\boldsymbol{s}_x = \text{READOUT}\left(\{\boldsymbol{r}_v : v \in V(S_x)\}\right). \tag{12}$$

Progressive Prompt Optimization with Privacy Consistency

To optimize the learnable prompts, we propose a joint optimization framework that simultaneously refines the prompt representations and enforces privacy consistency constraints.

Based on the template of subgraph similarity, we use the class prototype as the target for comparison. For a given task t with a labeled training set T_t, we define the prompt

optimization loss based on subgraph similarity between the prompt-augmented representation of an instance and the prototype of its class. Specifically, each class $c \in \mathcal{Y}$ is associated with a prototype vector $\tilde{\boldsymbol{s}}_{t,c}$, which serves as a semantic anchor representing the core pattern of that class in the embedding space. The loss is given by:

$$\mathcal{L}_{\text{task}}(\boldsymbol{p}) = -\sum_{(x_i,y_i)\in\mathcal{T}_t} \log \frac{\exp\left(\text{sim}(\boldsymbol{s}_{t,x_i}, \tilde{\boldsymbol{s}}_{t,y_i})/\tau\right)}{\sum_{c\in\mathcal{Y}} \exp\left(\text{sim}(\boldsymbol{s}_{t,x_i}, \tilde{\boldsymbol{s}}_{t,c})/\tau\right)}, \tag{13}$$

where τ represents a temperature hyperparameter that adjusts the smoothness of the output distribution.

Then we propose a multi-view privacy consistency strategy that assists prompt optimization, which enforces consistency of prompt representations across different perturbed views of the same node. Let $S_x^{(1)}$ and $S_x^{(2)}$ denote two perturbed subgraphs centered at node v via different stochastic privacy-perturbed mechanisms. As we all know, since random responses randomly flip edges, the graph-structured samples obtained are obviously different even under the same privacy budgets. Although these views differ in structures or features, they originate from the same source node and should encode similar semantic meaning. Therefore, the prompt representations derived from them should be aligned in the latent space.

$$\mathcal{L}_{\text{cons}} = -\frac{1}{N}\sum_{i=1}^{N} \log \left(\frac{\exp\left(\text{sim}(\boldsymbol{r}_i^{(1)}, \boldsymbol{r}_i^{(2)})/\tau\right)}{\sum_{\substack{j=1 \\ j\neq i}}^{N} \exp\left(\text{sim}(\boldsymbol{r}_i^{(1)}, \boldsymbol{r}_j^{(2)})/\tau\right)} \right). \tag{14}$$

To optimize these objectives, a joint learning scheme is adopted. The multi-view privacy consistency augmentations are fixed during each optimization phase, and the prompt parameters are updated based on both the task-specific loss and a contrastive consistency loss. Optionally, the views can be periodically refreshed through stochastic graph perturbations or sampled from a distribution consistent with the underlying privacy model. The two objectives are then unified into a single loss function for end-to-end optimization:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{task}} + \lambda \cdot \mathcal{L}_{\text{cons}}, \tag{15}$$

where λ controls the balance between downstream task utility and privacy consistency.

During the tuning of the downstream task, only the prompts are optimized without involving parameter tuning of the pre-trained model. The primary objective is to enhance task-specific performance by optimizing the learnable prompts without altering the underlying model weights. By freezing the parameters of the pre-trained GNN, the model achieves a significant reduction in computational resource consumption. This design improves both learning and inference efficiency, while preserving generalization capability across datasets.

Experimental Results

In this section, our experimental evaluation is designed to answer the following research questions (RQs):

- **RQ1:** How much better does our scheme perform compared to other approaches under varying privacy budgets?
- **RQ2:** Does the proposed scheme maintain generalizable performance across different data scales?
- **RQ3:** What is the adaptability of our PAGPL across varying graph encoders?
- **RQ4:** How do the main components of our scheme impact the performance?
- **RQ5:** How effective are the constructed prompts in aligning with downstream tasks?

Experimental Setup

Datasets. We adopt five real-world datasets for evaluation: Cora (Sen et al. 2008), Citeseer (Sen et al. 2008), Pubmed (Sen et al. 2008), ENZYMES (Dobson and Doig 2003), and DD (Dobson and Doig 2003). These datasets are widely used benchmarks in GNN research, which covers diverse domains and graph structures. The experiments are performed to evaluate the effectiveness of the proposed scheme in three critical aspects, classification accuracy, F1 Score and AUC.

Baselines. We assess the performance of our scheme against state-of-the-art approaches from three primary categories as outlined below.

(1) *End-to-end supervised GNNs*: These methods involve directly training a GNN model on a given task and using it to produce the final prediction without additional processing. we utilize three well-known GNN models, including graph attention network (GAT) (Veličković et al. 2018), graph convolutional network (GCN) (Kipf and Welling 2017), and graph sample and aggregate (GraphSAGE) (Hamilton, Ying, and Leskovec 2017).

(2) *Pre-training with fine-tuning*: Deep graph infomax (DGI) (Veličković et al. 2019), and graph contrastive learning (GraphCL) (You et al. 2020) follow the "pre-train and fine-tune" paradigm.

(3) *Graph prompt learning* : GPrompt (Liu et al. 2023b), GPF (Fang et al. 2022), GPF-Plus (Fang et al. 2023) and ProG (Sun et al. 2023a).

Effectiveness Analysis (RQ1)

In this section, we evaluate the performance of different models under varying privacy budgets ϵ. We conducted experiments with privacy budgets $\epsilon \in \{4.0, \ldots, 10.0\}$ on four benchmark datasets: Cora, CiteSeer, PubMed and ENZYMES. Figure 3 illustrates the accuracy of various methods within the set privacy strength range. Our proposed scheme demonstrates a clear upward trend in performance as ϵ increases. Importantly, even at relatively low privacy budgets, our scheme maintains a stable and competitive performance level. Compared to baseline methods, our scheme exhibits significantly stronger robustness under low to moderate privacy constraints.

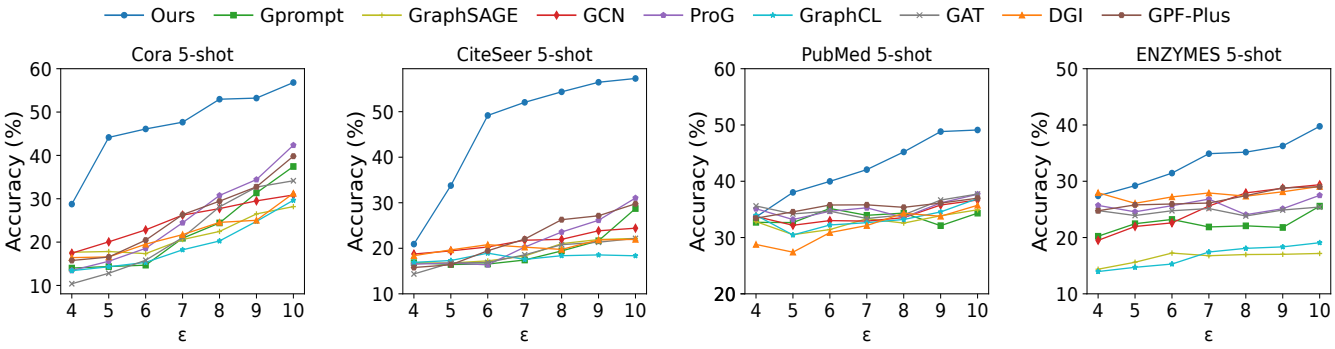

Figure 3: Performance (%) comparison of our proposed scheme and baselines under different privacy budgets.

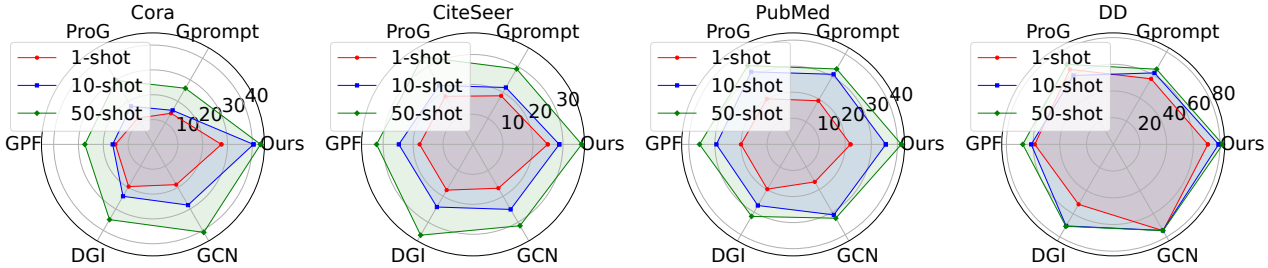

Figure 4: Performance (%) under different shot settings.

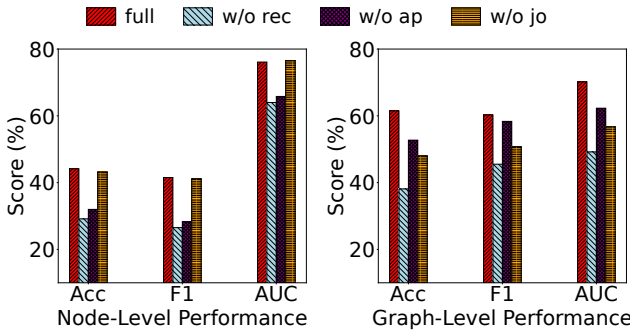

Figure 5: Effectiveness of main components under a privacy budget of $\epsilon = 5.0$.

Performance over Multiple Scales (RQ2)

In this section, we investigate the impact of the number of shots on privacy-disturbed condition across both node-level and graph-level tasks. We set the total privacy budget $\epsilon = 6.0$ and evaluate performance under varying shot counts, including 1, 10, and 50-shot settings. Empirical results support this analysis in Figure 4. We observe a clear positive correlation between the number of shots and downstream task performance under the same privacy constraint. Accuracy consistently improves as the number of shots increases from 1, 10, and 50-shot settings. However, the marginal gain diminishes, and the relative improvement is more dramatic between 1-shot and 10-shot than between 10 and 50, which suggests a possible saturation effect. This finding highlights a crucial trade-off in privacy-aware few-shot scenarios: while increasing the number of shots improves learning, it also increases the cumulative privacy cost if each sample access is considered under differential privacy accounting.

Transferability across Graph Encoders (RQ3)

To assess the transferability of our scheme across different GNN backbones, we conduct experiments under a fixed privacy budget $\epsilon = 10.0$ and a 5-shot learning setting. As shown in Table 1, our scheme consistently achieves competitive and superior performance across a diverse set of graph encoders, including GCN, GAT, and GraphSAGE. Despite architectural differences in message-passing mechanisms and inductive biases, the proposed PAGPL demonstrates strong adaptability. This indicates that the subgraph-based privacy-aware prompt design generalizes well across heterogeneous encoders, which effectively decouples prompt learning from specific GNN architectures.

Ablation Study (RQ4)

In this section, we compare our complete framework with three variants: "w/o rec" refers to the prompt scheme without the reconstructed graph by adaptive structure-wise Bayesian estimation; "w/o ap" is the prompt construction scheme without the privacy-aware module; "w/o jo" refers to the scheme that directly performs prompt tuning without joint optimization. Figure 5 illustrates the performance of our scheme and its various variants on both node-level and graph-level tasks. It can be observed that the "w/o rec" variant suffers a substantial performance drop, indicating that

GPL	GNNs	Cora			CiteSeer			PubMed			ENZYMES			DD		
		Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC
GPF-Plus	GAT	35.25	34.13	62.22	30.25	31.14	65.48	34.29	32.82	55.54	29.62	31.86	32.23	62.26	59.27	68.51
	GraphSAGE	40.27	38.32	61.21	29.21	25.03	68.91	36.71	37.53	56.02	27.31	25.06	34.11	65.62	63.13	69.02
	GCN	38.72	35.22	67.23	32.22	29.11	67.52	35.27	33.19	55.13	29.25	28.02	35.23	67.72	65.25	70.21
GPrompt	GAT	35.15	34.32	67.21	29.51	29.13	54.42	34.27	32.18	55.52	31.16	31.82	32.23	62.61	50.27	68.52
	GraphSAGE	40.27	35.31	69.11	26.13	25.01	58.29	32.71	34.25	56.01	33.32	32.01	34.12	65.61	53.23	71.07
	GCN	37.44	34.60	67.23	28.73	26.37	59.96	34.33	33.38	51.32	33.15	31.02	35.32	67.72	55.15	70.22
ProG	GAT	45.25	40.39	67.12	30.51	29.14	64.47	34.22	32.81	55.52	29.65	27.81	32.32	64.68	60.17	68.25
	GraphSAGE	39.75	38.43	71.11	32.18	25.01	68.92	36.79	36.55	56.01	31.37	30.02	34.11	65.61	61.32	68.04
	GCN	42.35	37.31	74.79	31.04	24.43	62.25	37.69	35.71	56.56	30.45	28.01	37.34	67.17	59.15	71.24
Ours	GAT	55.51	51.37	67.28	53.25	51.18	64.43	45.27	**44.83**	65.54	**39.87**	**37.81**	42.32	77.61	64.78	**78.25**
	GraphSAGE	52.74	48.32	71.11	56.18	**57.54**	**77.91**	42.47	39.51	**66.03**	37.33	35.02	44.71	75.61	63.43	76.08
	GCN	**56.81**	**52.37**	**75.78**	**57.29**	57.40	76.17	**49.11**	43.53	61.18	39.76	35.53	**45.31**	**78.06**	**66.13**	78.21

Table 1: Performance (%) of different GNNs under the 5-shot setting with a privacy budget of $\epsilon = 10.0$.

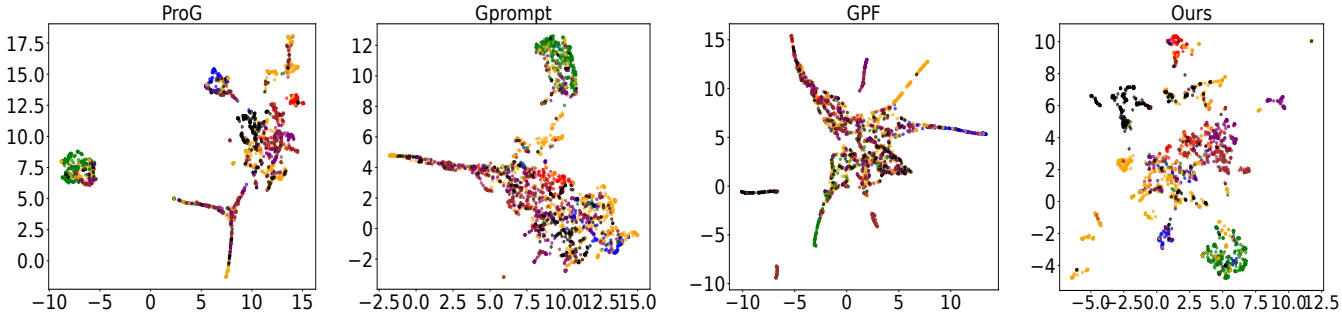

Figure 6: Visualization of node representations under perturbation conditions.

the reconstructed graph plays a vital role in enabling effective prompt construction. This is because the learnable reconstruction compensates for missing or obfuscated structural information caused by privacy mechanisms. Overall, these results verify that each proposed module contributes significantly to the overall effectiveness of our framework.

Prompt Validity Analysis (RQ5)

To qualitatively assess the effectiveness of different schemes under privacy-preserving perturbations, we visualize the prompt representations via UMAP in Figure 6. Our scheme clearly forms well-separated clusters with compact intra-class distributions and minimal inter-class overlap. This observation indicates that the proposed privacy-aware GPL effectively captures task-relevant semantics, even in the presence of strong noise and structural perturbations. The consistent geometry across clusters further suggests improved stability and robustness, as the learned representations remain resilient to noise-induced variations while preserving class-discriminative information.

Conclusion

In this paper, we propose a novel privacy-aware graph prompt learning, which overcomes the limitations of topological prompt generation when dealing with privacy-perturbed graphs. The main conclusions can be drawn from this research work as follows: 1) the adaptive structure-wise Bayesian estimation effectively reconstructs the privacy-noise graph structures; 2) the prompt construction for suppressing residual noise and the progressive privacy consistency optimization achieve resilient alignment with downstream tasks; 3) this work provides a novel viewpoint for GPL, specifically how to effective handle downstream tasks in few-shot settings under privacy noise; and 4) comprehensive experiments show that our proposed scheme demonstrates robust adaptability to different privacy budgets and shot numbers. These results show promising performance in privacy-disturbed scenarios, effectively mitigating the impact of privacy constraints on GPL. In the near future, we plan to explore the application of GPL in multi-agent systems.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 62402106, U22B2025, 62172093), the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20241272), the Fundamental Research Funds for the Central Universities (Grant No. 2242025K30025), and the Start-Up Research Fund of Southeast University (Grant No. RF1028623129).

References

Chen, Q.; Wang, L.; Zheng, B.; and Song, G. 2025. Dagprompt: Pushing the limits of graph prompting with a distribution-aware graph prompt tuning approach. In *Proceedings of the ACM on Web Conference 2025*, 4346–4358.

Dobson, P. D.; and Doig, A. J. 2003. Distinguishing enzyme structures from non-enzymes without alignments. *Journal of molecular biology*, 330(4): 771–783.

Dwork, C.; McSherry, F.; Nissim, K.; and Smith, A. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Proceedings of the 3rd Theory of Cryptography Conference (TCC)*, volume 3876 of *Lecture Notes in Computer Science*, 265–284. Springer.

Fang, T.; Zhang, Y. M.; Yang, Y.; and Wang, C. 2022. Prompt tuning for graph neural networks.

Fang, T.; Zhang, Y. M.; Yang, Y.; Wang, C.; and Chen, L. 2023. Universal Prompt Tuning for Graph Neural Networks. In *Proceedings of the Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS)*.

Fang, X.; Easwaran, A.; and Genest, B. 2025. Adaptive Multi-prompt Contrastive Network for Few-shot Out-of-distribution Detection. In *International Conference on Machine Learning*.

Fang, X.; Easwaran, A.; Genest, B.; and Suganthan, P. N. 2025a. Your data is not perfect: Towards cross-domain out-of-distribution detection in class-imbalanced data. *Expert Systems with Applications*.

Fang, X.; Fang, W.; Wang, C.; Liu, D.; Tang, K.; Dong, J.; Zhou, P.; and Li, B. 2025b. Multi-pair temporal sentence grounding via multi-thread knowledge transfer network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 2915–2923.

Fu, D.; Bao, W.; Maciejewski, R.; et al. 2023. Privacy-preserving graph machine learning from data to computation: A survey. *ACM SIGKDD Explorations Newsletter*, 25(1): 54–72.

Ge, Q.; Zhao, Z.; Liu, Y.; Cheng, A.; Li, X.; Wang, S.; and Yin, D. 2023. Enhancing Graph Neural Networks with Structure-Based Prompt. *arXiv preprint arXiv:2310.17394*.

Grover, A.; and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 1025–1035.

Jia, J.; Gao, P.; Luo, M.; Wu, C.; and Guo, J. 2026. CTEA: Camouflaged topological element attack via causal influence discovery. *Expert Systems with Applications*, 297: 129160.

Jia, J.; Li, R.; Wu, C.; Feng, Y.; Ma, S.; Wang, L.; and Deng, R. H. 2025a. Environment-Adaptive Representation Interaction for Privacy-Perceptual GNNs against Deceptive OOD Attacks. *IEEE Transactions on Information Forensics and Security*.

Jia, J.; Li, R.; Wu, C.; Ma, S.; Wang, L.; and Deng, R. H. 2025b. SIGFinger: A Subtle and Interactive GNN Fingerprinting Scheme via Spatial Structure Inference Perturbation. *IEEE Transactions on Dependable and Secure Computing*.

Jia, J.; Ma, S.; Liu, Y.; Wang, L.; and Deng, R. H. 2023. A Causality-Aligned Structure Rationalization Scheme Against Adversarial Biased Perturbations for Graph Neural Networks. *IEEE Transactions on Information Forensics and Security*, 19: 59–73.

Jia, J.; Yu, J.; Wu, D.; Wu, C.; Zhu, H.; and Wang, L. 2025c. Prompt as a Double-Edged Sword: A Dynamic Equilibrium Gradient-Assigned Attack against Graph Prompt Learning. In *The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

Jin, W.; Derr, T.; Liu, H.; Wang, Y.; Wang, S.; Liu, Z.; and Tang, J. 2020. Self-Supervised Learning on Graphs: Deep Insights and New Direction. *arXiv preprint arXiv:2006.10141*.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.

Li, J.; Chiu, B.; Feng, S.; and Wang, H. 2020. Few-shot named entity recognition via meta-learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(9): 4245–4256.

Li, Y.; Wang, P.; Li, Z.; Yu, J. X.; and Li, J. 2024. Zerog: Investigating cross-dataset zero-shot transferability in graphs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1725–1735.

Liu, K.; and Terzi, E. 2008. Towards Identity Anonymization on Graphs. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 93–106. ACM.

Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; and Neubig, G. 2023a. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM computing surveys*, 55(9): 1–35.

Liu, Z.; Yu, X.; Fang, Y.; and Zhang, X. 2023b. GraphPrompt: Unifying Pre-Training and Downstream Tasks for Graph Neural Networks. In *Proceedings of the ACM Web Conference 2023 (WWW)*, 3584–3595. Austin, TX, USA: ACM.

Lu, Y.; Jiang, X.; Fang, Y.; and Shi, C. 2021. Learning to pre-train graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 4276–4284.

Nguyen, D. C.; Ding, M.; Pathirana, P. N.; Seneviratne, A.; Li, J.; and Poor, H. V. 2021. Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 23(3): 1622–1658.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710.

Satorras, V. G.; and Estrach, J. B. 2018. Few-shot learning with graph neural networks. In *International conference on learning representations*.

Seliem, M. M.; Kamel, A. R.; Taha, I. M.; and El-Nasr, M. M. A. 2025. A Note on Prior Selection in Bayesian Estimation. *Statistics, Optimization & Information Computing*, 13(2): 795–806.

Sen, P.; Namata, G.; Bilgic, M.; Getoor, L.; Gallagher, B.; and Eliassi-Rad, T. 2008. Collective classification in network data. *AI magazine*, 29(3): 93–93.

Shen, H.; Li, G.; Zhong, J.; and Zhou, Y. 2025. LDP: Generalizing to multilingual visual information extraction by language decoupled pretraining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 6805–6813.

Sun, M.; Zhou, K.; He, X.; Wang, Y.; and Wang, X. 2022. Gppt: Graph pre-training and prompt tuning to generalize graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1717–1727.

Sun, X.; Cheng, H.; Li, J.; Liu, B.; and Guan, J. 2023a. All in One: Multi-Task Prompting for Graph Neural Networks. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2120–2131. Long Beach, CA, USA: ACM.

Sun, X.; Zhang, J.; Wu, X.; Cheng, H.; Xiong, Y.; and Li, J. 2023b. Graph Prompt Learning: A Comprehensive Survey and Beyond. *arXiv preprint arXiv:2311.16534*.

Sun, Y.; Zhu, Q.; Yang, Y.; Wang, C.; Fan, T.; Zhu, J.; and Chen, L. 2024. Fine-tuning graph neural networks by preserving graph generative patterns. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 9053–9061.

Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; and Mei, Q. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, 1067–1077.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *International Conference on Learning Representations (ICLR)*.

Veličković, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep Graph Infomax. In *International Conference on Learning Representations (ICLR)*.

Wang, N.; Luo, M.; Ding, K.; Zhang, L.; Li, J.; and Zheng, Q. 2020. Graph few-shot learning with attribute matching. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1545–1554.

Wang, R.; Mou, S.; Wang, X.; Xiao, W.; Ju, Q.; Shi, C.; and Xie, X. 2021. Graph structure estimation neural networks. In *Proceedings of the web conference 2021*, 342–353.

Yazdanjue, N.; Yazdanjouei, H.; Gharoun, H.; Khorshidi, M. S.; Rakhshaninejad, M.; Amiri, B.; and Gandomi, A. H. 2025. A comprehensive bibliometric analysis on social network anonymization: current approaches and future directions. *Knowledge and Information Systems*, 1–80.

Ye, M.; Shen, W.; Du, B.; Snezhko, E.; Kovalev, V.; and Yuen, P. C. 2025. Vertical federated learning for effectiveness, security, applicability: A survey. *ACM Computing Surveys*, 57(9): 1–32.

You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; and Shen, Y. 2020. Graph Contrastive Learning with Augmentations. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 5812–5823.

Yu, Q.; Zou, L.; Luo, X.; Zhao, X.; and Li, C. 2025. Uniform Graph Pre-training and Prompting for Transferable Recommendation. *ACM Transactions on Information Systems*.

Yu, X.; Fang, Y.; Liu, Z.; and Zhang, X. 2024. Hgprompt: Bridging homogeneous and heterogeneous graphs for few-shot prompt learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 16578–16586.

Zhang, C.; Ding, K.; and Li, J. 2025. Few-shot learning on graphs. In *Handbook on Neurosymbolic AI and Knowledge Graphs*, 96–117. IOS Press.

Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li, W.; and Gao, Y. 2021. A survey on federated learning. *Knowledge-Based Systems*, 216: 106775.

Zhang, W.; Jia, J.; Jia, X.; Huang, Y.; Li, X.; Wu, C.; and Wang, L. 2025. PATFinger: Prompt-Adapted Transferable Fingerprinting against Unauthorized Multimodal Dataset Usage. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 403–413.

Zhu, X.; Tan, V. Y.; and Xiao, X. 2023. Blink: link local differential privacy in graph neural networks via Bayesian estimation. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2651–2664.

Zyphur, M. J.; and Oswald, F. L. 2015. Bayesian estimation and inference: A user's guide. *Journal of Management*, 41(2): 390–420.