PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

A Unified Prompt for Enhancing Heterogeneous Graph Pre-training via Edge-based Message Passing

Fengyu Yan[1], **Xiaobao Wang**[2,*], **Qianhua Tang**[3], **Dongxiao He**[1] and **Di Jin**[1,*]
[1]School of Computer Science and Technology, Tianjin University, Tianjin, China
[2]School of Artificial Intelligence, Tianjin University, Tianjin, China
[3]Tianjin International Engineering Institute, Tianjin University, Tianjin, China
{fengyuyan, wangxiaobao, tangqianhua, hedongxiao, jindi}@tju.edu.cn

Abstract

Inspired by natural language processing prompt learning, recent heterogeneous graph prompt-tuning methods have been developed to better align pre-trained models with downstream tasks. However, existing heterogeneous prompt methods primarily focus on holistic framework design, causing prompts to heavily depend on specific pre-trained models and thus limiting generalization. To address this, we focus on the common encoder module shared across pre-trained models and the multi-relational edge structure unique to heterogeneous graphs, proposing a novel heterogeneous graph prompt tuning method named HGMRP. It simultaneously resolves the issue of inconsistent objectives during prompt optimization when messages propagate across different relational edges. By assigning learnable prompt vectors to different types of edges, HGMRP integrates multi-relational prompts directly into the message-passing process. Furthermore, we extend single-relation prompts into a composite structure that includes both relation-specific and shared components, thereby enhancing expressive capability. Extensive experiments conducted on four widely used heterogeneous graph datasets show that HGMRP can adapt to various types of pre-trained models and significantly outperforms existing heterogeneous prompt methods, validating its effectiveness and superiority.

1 Introduction

Real world complex data, such as recommendation systems, social networks, and academic citation networks [Cao *et al.*, 2025; Hui *et al.*, 2025; Zhao *et al.*, 2020], often contain multi-relational information. These scenarios can be naturally modeled as heterogeneous information networks. For representation learning on heterogeneous information networks, existing semi-supervised methods have been extensively studied [Wang *et al.*, 2019; Hu *et al.*, 2020; Fu *et al.*, 2020a]. However, many real-world scenarios suffer from a lack of high-quality labeled data, which significantly limits the effectiveness of existing semi-supervised models. To address the challenge of limited labeled data, pre-training methods for heterogeneous information networks have emerged. These methods follow a "pre-train, fine-tuning" paradigm and involve carefully designed pre-training tasks [Jing *et al.*, 2021; Wang *et al.*, 2021; Tian *et al.*, 2023], thereby mitigating the drawback of scarce labels.

However, a significant gap often exists between pre-training tasks and downstream tasks in heterogeneous graphs. To mitigate the side effects of such task gaps, existing work has drawn inspiration from the concept of prompts in NLP [Devlin *et al.*, 2019; Radford *et al.*, 2018; Brown *et al.*, 2020], aiming to bridge the connection between upstream and downstream tasks, and has introduced the "prompt-tuning" paradigm during the fine-tuning stage. In homogeneous networks, this paradigm has been reasonably well studied, primarily by adding prompt information to existing node features or representations [Liu *et al.*, 2023; Fang *et al.*, 2023; Huang *et al.*, 2026], or by inserting new nodes to simulate the form of prompts in NLP [Sun *et al.*, 2023]. However, such approaches overlook the crucial relational information in heterogeneous information networks, leading to the loss of effective information in multi-relational scenarios.

To address the introduction of multi-relations, recent years have witnessed a number of prompt-based studies on heterogeneous information networks, which demonstrate the strong expressive power of this paradigm compared with pre-trained models. However, existing heterogeneous prompt methods primarily focus on holistic framework design and exhibit strong dependencies on specific pre-trained models, which limits their generalizability across different pre-training architectures. For example, the prompt design in HGPROMPT [Yu *et al.*, 2024] is tightly coupled with its custom link-prediction-based pre-training scheme; the use of prompts in HGMP [Jiao *et al.*, 2025] is bound to its specific, unified graph-level pre-training task; and the prompt transfer strategy in HetGPT [Ma *et al.*, 2024] is constrained by the type of pre-training framework, which task transfer approach is only reusable within contrastive learning frameworks and difficult to adapt to other pre-training paradigms.

We note that various heterogeneous graph pre-trained models share a common encoder component, and message passing is a core stage in heterogeneous graph representation

[*]Corresponding Authors.

learning. This leads us to consider: *can we design a general prompt mechanism based on the encoder to address the above generalization issue?* While the existing method EdgePrompt [Fu *et al.*, 2025], designed for homogeneous graphs, attempts to incorporate prompts during propagation, it overlooks the critical relational semantic information carried by multi-relation edges in heterogeneous settings. Moreover, employing multi-relational edge-level prompts in message-passing allows prompts to function independently across different semantic relations, thereby alleviating the optimization conflicts inherent in existing methods, such as the heterogeneous graph prompts above, which mainly focus on prompts at the node feature or representation level. For instance, in a recommendation system scenario illustrated in Figure 1, a single prompt attached to a user node propagates to both movie and director nodes during message-passing, yet fails to account for the potential conflicts arising from users' divergent preferences toward different movies and directors when optimizing the prompt.

To achieve the above objectives, we propose a novel method Heterogeneous Graph Multi-Relational Prompt (HGMRP). We leverage edge-based message passing to effectively achieve independent propagation. In heterogeneous graphs, relationships between different types of nodes carry distinct semantic information and play a crucial role during propagation. Therefore, edges of the same type share the same prompt vector, while prompt vectors differ across different edge types. These added multi-relation prompts are incorporated into the message passing process and are trained during downstream fine-tuning tasks. To enhance the multi-relation prompts, we extend the single shared prompt vector per relation into a composite prompt structure, which includes both relation-specific and shared components across different relations. To validate the effectiveness of our method, we conduct extensive experiments on 4 heterogeneous graph datasets. Our approach consistently improves performance across different types of graph pre-trained models and outperforms existing heterogeneous prompt methods. The contributions of our work can be summarized as follows:

- Our proposed method breaks away from the traditional integrated framework of "pre-training, prompt-tuning". It adapts to a wider variety of heterogeneous graph pre-trained models and unifies the approach to prompt-based fine-tuning for heterogeneous graph pre-training.
- We introduce a novel approach to prompt-based fine-tuning for heterogeneous graphs. To the best of our knowledge, it is the first to employ message-passing-based prompts in heterogeneous graph scenarios, highlighting the importance of relational information in such graphs. This opens up new directions for future research on heterogeneous graph prompts.
- We conduct extensive experiments on 4 commonly used heterogeneous graph datasets, comparing our method against 3 categories of heterogeneous graph representation learning approaches. The results demonstrate and validate the effectiveness of our proposed method.

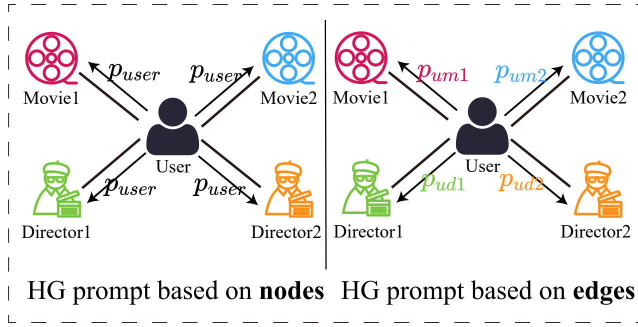

Figure 1: Comparison between HG prompt based on nodes and edges: In a movie recommendation scenario, a single node-based prompt attached to user features propagates to neighboring nodes of different types during message passing, which may lead to inconsistent optimization objectives. In contrast, the edge-based propagation scheme supports independent message flow across distinct relations.

2 Related Work

2.1 Heterogeneous Graph Pre-training

To address the issue of insufficient model expressivity caused by the scarcity of high-quality labels, the pre-training paradigm has been proposed and widely applied to graph data. For handling multi-relational scenarios, numerous studies in recent years have focused on pre-training for heterogeneous graphs. For instance, DMGI [Park *et al.*, 2020] and HDMI [Jing *et al.*, 2021] employ the maximization of mutual information between local node representations and global graph summaries as their pre-training objective. HeCo [Wang *et al.*, 2021] utilizes contrastive learning between positive and negative samples from dual views, while HGMAE [Tian *et al.*, 2023] adopts reconstruction of both structural and attribute information across multiple views. HERO [Mo *et al.*, 2024] aligns homogeneous and heterogeneous perspectives within heterogeneous graphs. All the aforementioned methods solely employ corresponding task heads, such as classifiers, when adapting to specific downstream tasks.

2.2 Prompt-based Heterogeneous Graph Tuning

In natural language processing, prompt tuning helps bridge the gap between pre-training tasks and downstream tasks. Inspired by this, numerous studies in recent years have explored prompt-based methods on graphs, such as ProG [Sun *et al.*, 2023] and GPF [Fang *et al.*, 2023]. These works primarily focus on homogeneous graph scenarios. To address prompt learning in heterogeneous settings, a series of approaches have also been proposed. For instance, HGPROMPT [Yu *et al.*, 2024] introduces a framework from heterogeneous graph pre-training to prompting, HGMP [Jiao *et al.*, 2025] unifies both pre-training and downstream tasks as graph-level tasks, and HetGPT [Ma *et al.*, 2024] designs adaptation methods using positive and negative sample pairs for upstream and downstream tasks. However, existing methods design prompts mainly by modifying features or representations, overlooking the positive and negative effects during the message propagation in the fine-tuning process.

3 Preliminaries

3.1 Heterogeneous Graph

A heterogeneous graph is formally defined as a tuple $G = (V, E, \phi, \psi, X)$. Here, $V = \{v_1, v_2, \ldots, v_N\}$ denotes the set of nodes, and $E = \{e_1, e_2, \ldots, e_M\} \subseteq V \times V$ denotes the set of edges connecting them. The graph's heterogeneity is captured by two mapping functions: the node type mapping $\phi : V \to \mathcal{T}_V$ assigns each node to a type from a predefined set $\mathcal{T}_V$, and the relation type mapping $\psi : E \to \mathcal{R}_E$ assigns each edge to a type from a predefined set $\mathcal{R}_E$, where the condition $|\mathcal{T}_V| + |\mathcal{R}_E| > 2$ must hold, and $X_{\mathcal{T}_i} \in \mathbb{R}^{N \times d_{\mathcal{T}_i}}$ represents the node feature matrix, $\mathcal{T}_i \in \mathcal{T}_V$, with $\mathbf{x}_i \in \mathbb{R}^d$ being the feature vector for node v_i.

3.2 Heterogeneous Graph Pre-training

The pre-training framework for heterogeneous graphs typically consists of two sequential stages. In the first stage, given a heterogeneous graph G, a heterogeneous graph encoder f_θ and a pre-training head g_ϕ. The model parameters θ and ϕ are jointly optimized by minimizing a pre-training loss $\mathcal{L}_{\text{pre}}$ constructed from self-supervised signals $\mathcal{V}_{\text{self}}$ derived from the graph itself depends on the pre-training strategy:

$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathcal{L}_{\text{pre}} \left(f_\theta, g_\phi, V, V_{self} \right).$$

For downstream tasks, the pre-trained encoder f_θ and pre-training head g_ϕ are frozen, and only a task-specific classifier cls_ψ is trained on a limited set of labeled data $\mathcal{D}_{\text{down}} = \{(v_i, y_i)\}$. Its parameters ψ are optimized by minimizing a downstream supervised loss, such as cross-entropy, the fine-tuning process can be described as:

$$\psi^* = \arg\min_{\psi} \mathcal{L}_{\text{down}} \left(f_{\theta^*}, cls_\psi, \mathcal{D}_{\text{down}} \right).$$

3.3 Heterogeneous Graph Prompt Tuning

In this stage, the pre-trained graph encoder f_θ and the pre-training head g_{ϕ^*} are kept frozen to preserve the knowledge acquired during self-supervised pre-training. Only the prompt parameters p and a task-specific classifier cls_ψ are trained using a labeled dataset $\mathcal{D}_{\text{down}}$. The heterogeneous graph prompt tuning can be described as:

$$\psi^*, p^* = \arg\min_{\psi,p} \mathcal{L}_{\text{down}} \left(f_{\theta^*}, cls_\psi, p, \mathcal{D}_{\text{down}} \right).$$

4 Method

In this section, we will briefly introduce the heterogeneous graph pre-training module, followed by the design and application of prompts within the pre-trained model.

4.1 Overview Framework

Our model primarily consists of two parts. As illustrated in Figure 2, the first part is the heterogeneous graph pre-training module, where unified features are fed into a designed pre-training task for computation. For the second part, we introduce a guided propagation-based multi-relational prompt, termed HGMRP, into the pre-trained model. It incorporates a single learnable parameter in each layers of a heterogeneous graph neural network. To further enhance the expressive capacity for multi-relational scenarios, we propose an advanced form, HGMRP+, which designs multiple sets of learnable parameters for different relations across various network layers. Finally, trained with limited labeled data, we obtain controllable propagating fine-tuning prompts.

4.2 Feature Transformation

In heterogeneous scenarios, different types of nodes possess features with varying semantics and dimensions. Existing encoders are primarily divided into meta-path-based methods and attention-based methods. The former constructs homogeneous graphs based on predefined meta-paths, thus requiring no feature transformation, while the latter computes attention scores among different node types and relies on features with consistent semantics and dimensions. Our prompt-based fine-tuning method is applicable to both approaches. The process of unifying features $h_{\mathcal{T}_V}$ of type $\mathcal{T}_\mathcal{V}$ is described as follows:

$$h_{\mathcal{T}_V} = \sigma(w_{\mathcal{T}_V} \times x_{\mathcal{T}_V} + b_{\mathcal{T}_V}), \tag{1}$$

where $\mathcal{T}_\mathcal{V}$ represents the type of the nodes, $w_{\mathcal{T}_\mathcal{V}}$ is the type-specific transformation matrixs, σ is the activation function.

4.3 Pretrained Model

To better illustrate how our prompts are integrated, we briefly outline the two major types of heterogeneous graph pre-training methods that our framework is designed to accommodate. As shown in the overall framework (Figure 2), the transformed node features are fed into a heterogeneous graph encoder to obtain node representations:

$$H_{T_V} = \text{HetGNN}\big(H^0_{\mathcal{T}_V}; \mathcal{E}\big), \tag{2}$$

which are then passed to task-specific pre-training heads for optimization. Specifically, for contrastive pre-training, the objective typically involves learning representations by distinguishing between positive and negative sample pairs. For reconstructive pre-training, a graph decoder is used to recover the original graph data by minimizing the loss between the original and reconstructed features and structures.

All these pre-training tasks employ an encoder-based architecture. Given that our approach incorporates guided propagation-based prompts into the heterogeneous graph encoder, it can be seamlessly adapted to various types of pre-trained models.

4.4 Multiple Relational Prompt Design

To enable the independent propagation of prompts in heterogeneous graphs, inspired by EdgePrompt for homogeneous graphs and taking into account the unique structural specificity of heterogeneous graphs, we first design the HGMRP, which innovatively introduces prompts into the propagation mechanism. To enhance the expressiveness of HGMRP, an advanced version termed HGMRP+ is further proposed. Both are detailed in the following sections.

HGMRP. In the heterogeneous graph encoding stage, to avoid interfering with feature propagation while incorporating prompts, we design a novel edge fusion propagation mechanism. We assign a shared learnable prompt vector for

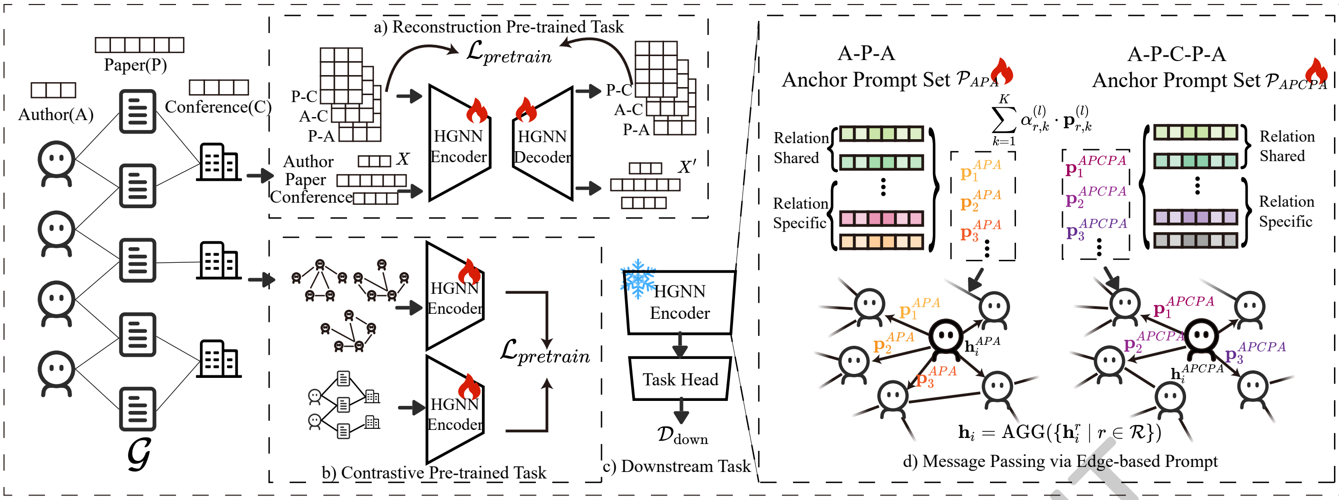

Figure 2: Overview of the HGMRP+: Our method operates within the HGNN encoder of the pre-trained model. The pre-training phase corresponding to a) and b) . After pre-training, the model proceeds to stage c) the downstream task, where parameters in the HGNN encoder are frozen. Prompts are incorporated into the encoder's message-passing mechanism, following the scheme illustrated in d).

each relation type at each layer. In two widely-used heterogeneous graph encoders both propagate node representations during message passing but do not explicitly consider how multi-relation edges can be fused in this process. Therefore, we integrate relation-aware prompts by combining edge messages during the aggregation phase. Specifically, in each metapath view, the propagation with multi-relation prompts is formulated as follows:

$$\mathbf{h}_j^{(l)} = \text{AGG}\left(\left\{\mathbf{h}_j^{(l),r} \mid v_j \in \mathcal{N}_r(v_i)\right\}\right), \tag{3}$$

$$\mathbf{h}_i^{(l+1),r} = \text{COMB}\left(\left\{\mathbf{h}_i^{(l),r}, \mathbf{h}_j^{(l),r}, \mathbf{p}_r^{(l)}\right\}\right), \tag{4}$$

where $\mathbf{p}_r^{(l)}$ denotes the prompt vector for relation r at layer l, and $\mathcal{N}_r(v_i)$ represents the set of neighbors connected to v_i via relation r. After obtaining the prompt-enhanced representations for each view, we aggregate them across different relations to derive the final node representation:

$$\mathbf{h}_i^{(l+1)} = \text{AGG}\left(\left\{\mathbf{h}_i^{(l+1),r} \mid r \in \mathcal{R}\right\}\right), \tag{5}$$

where $\mathcal{R}$ denotes the set of relation types, and AGG denotes an aggregation function such as attention. This entire propagation framework is referred to as HGMRP. In attention-based methods, we likewise design a prompt for each relation type at every layer. During the heterogeneous message passing steps, when computing attention, the prompt vector interacts with the query and key to adjust the attention weights.

HGMRP+. While the aforementioned HGMRP achieves independent aggregation of different semantics during message passing in the fine-tuning stage through the design of relation-specific prompts, it faces limitations in complex pre-trained models. A single shared prompt vector per relation type per layer is often insufficient to fully represent the intricate interactions between relation edges during fine-tuning. To endow the prompt for each relation edge with stronger expressiveness, we propose an enhanced version HGMRP+.

If a unique prompt is assigned to each individual relation edge rather than sharing one across edges of the same relation type within a layer, issues can arise in settings with limited downstream task labels. In such cases, many randomly initialized prompts may remain underfitted due to insufficient training signals, while overfitting could simultaneously occur on edges with more frequent labels. Moreover, such a design is only applicable in semi-supervised settings and fails when new nodes or new relation edges are introduced. Considering these challenges, inspired by GPF [Fang *et al.*, 2023], we introduce an anchor prompt set for each relation type r:

$$\mathcal{P}_r^{(l)} = \{\mathbf{p}_{r,1}^{(l)}, \mathbf{p}_{r,2}^{(l)}, \ldots, \mathbf{p}_{r,K}^{(l)}\}, \tag{6}$$

where l denotes the layer index of the pre-trained model decoder, and K is the hyperparameter for the number of anchor prompts. The prompt for a relation edge in layer l is composed from its corresponding relational anchor prompt set, which is shared among all edges of the same relation type within that layer. Within the anchor prompt set, we design two categories of prompts: one is relation-specific prompts, exclusive to a single relation type within the layer; the other is relation-shared prompts, shared across multiple relation types within the same layer, aiming to capture common information between relations. The prompt for a relation edge in layer l is dynamically composed via an attention mechanism:

$$\mathbf{p}_r^{(l)} = \sum_{k=1}^{K} \alpha_{r,k}^{(l)} \cdot \mathbf{p}_{r,k}^{(l)}, \tag{7}$$

where $\alpha_{r,k}^{(l)}$ denotes the contribution weight of the k-th anchor prompt to the final prompt for relation r at layer l.

In order to compute $\alpha_{r,k}^{(l)}$, the information of an edge is derived from its two connected nodes. Inspired by the attention mechanism in GAT [Velickovic *et al.*, 2018], we calculate it through a combination of the two nodes' representations.

Given an edge connecting node v_i and v_j, we design a function of their representations:

$$\mathbf{q}^{(l),r}(v_i, v_j) = \sigma(\mathbf{W}_r^{(l)} \left[\mathbf{h}_i^{(l-1),r} \| \mathbf{h}_j^{(l-1),r}\right]), \qquad (8)$$

where $\mathbf{W}_r^{(l)} \in \mathbb{R}^{K \times 2d_T}$ is a learnable projection matrix for relation r at layer l, and K denotes the number of anchor prompts. Using the computed q, we obtain attention weights through softmax normalization, which are then used to aggregate the anchor prompt set and derive the actual prompt for the relation edge, which can be expressed as:

$$\alpha_{r,k}^{(l)} = \frac{\exp\left(\mathbf{q}^{(l),r}(v_i, v_j)[k]\right)}{\sum_{k=1}^{K} \exp\left(\mathbf{q}^{(l),r}(v_i, v_j)[k]\right)}. \qquad (9)$$

Thus, we obtain the prompt composition for each relation edge in the graph, and the message passing step in the decoder remains identical to that in HGMRP.

The above describes the construction and fusion process of prompts. The overall heterogeneous prompt-based fine-tuning stage can be summarized as:

$$\psi^*, \mathcal{P}^*, \mathcal{W}^* = \arg\min_{\psi, \mathcal{P}, \mathcal{W}} \mathcal{L}_{\text{down}}\left(f_{\theta^*}, \text{cls}_{\psi}, \mathcal{P}, \mathcal{W}, \mathcal{D}_{\text{down}}\right), \qquad (10)$$

where $\mathcal{P}$ and $\mathcal{W}$ correspond to the set of prompt anchors and the parameters in the attention function for prompt composition, respectively.

5 Experiments

To validate the practical effectiveness of our prompt-tuning approach, we conducted comprehensive experiments. In the following sections, we first introduce the experimental setup, including the datasets, various hyperparameters in our method, as well as the pre-trained models and existing fine-tuning methods used for comparison. We then present the experimental results, which address the following research questions (RQs):

- RQ1: Does our designed model enhance pre-trained models more effectively than existing heterogeneous prompt methods?
- RQ2: How does the number of anchor prompts, as a key hyperparameter in HGMRP+, influence model performance?
- RQ3: What is the impact of limited labeled data on the performance of our method?
- RQ4: Can our relation-edge-based method be successfully combined with existing heterogeneous prompt techniques to yield further performance improvements?

5.1 Experimental Setup

Datasets. We evaluate our approach on four heterogeneous graph benchmarks. The ACM dataset is constructed from ACM [Zhao *et al.*, 2020] conference publications, comprising four node types: paper, author, term, and subject. Relations such as authorship and topic association connect these nodes, and the task is to predict the research area of each paper. DBLP [Fu *et al.*, 2020b] is a scholarly collaboration network consisting of authors, papers, and venues, with relations capturing publication and affiliation patterns; here we perform author classification based on their publication profiles. Freebase [Lv *et al.*, 2021] is a large-scale knowledge graph containing real-world entities with semantic relations, and we focus on classifying book entities. AMiner [Hu *et al.*, 2019] is an academic network involving papers, authors, and venues, where links reflect writing and collaboration relations; the objective is to assign papers to corresponding research topics. The statistical information of the datasets is presented in Table 1.

Dataset	#Nodes	#Types	#Edges	#Rel. Types	Target Node
ACM	8,994	3	25,922	4	paper
DBLP	26,128	4	296,563	3	author
Freebase	180,098	8	1,057,688	36	book
AMiner	55,783	4	153,676	4	paper

Table 1: Statistics of the heterogeneous graph datasets

Baselines. To demonstrate our model's fine-tuning capability in few-shot scenarios, we compare it against the following three types of approaches: 1) traditional semi-supervised heterogeneous representation learning methods, such as HAN [Wang *et al.*, 2019] and HGT [Hu *et al.*, 2020], to illustrate the effectiveness of heterogeneous graph pre-training; 2) widely-used heterogeneous graph pre-training methods, such as HeCo [Wang *et al.*, 2021], HGMAE [Tian *et al.*, 2023], and HDMI [Jing *et al.*, 2021], to highlight the advantage of our fine-tuning strategy; and 3) existing heterogeneous graph prompt-tuning methods, such as HGPROMPT [Yu *et al.*, 2024] and HetGPT [Ma *et al.*, 2024], to verify the effectiveness of our relation-edge-based fine-tuning approach.

Parameter Settings. In this section, we provide a comprehensive description of the parameter settings used in the experiments. To investigate the effectiveness under low-label scenarios, the main experiments adopt 10-shot and 20-shot settings, which refer to randomly selecting 10 or 20 labeled samples per class, respectively. For baseline models, we strictly follow the experimental setups reported in their original papers without modifications, and pre-trained models are also used as originally released without further adjustment. Three evaluation metrics are employed, including, Micro-F1 (measures global classification accuracy by counting total true predictions across all classes), Macro-F1 (calculates the unweighted average of per-class F1 scores to treat each class equally), and AUC (evaluates the model's ranking capability, typically used in binary or multi-class classification tasks to measure separability between classes). Our prompt-tuning method runs five times with random node selections, and the mean and variance are reported.

Our main hyperparameters include the number of anchor prompts per relation in each layer. In the primary experiments, we set the total number of anchor prompts to 10, consisting of 8 relation-specific anchor prompts exclusive to each relation type and 2 relation-shared anchor prompts that are common across different relation types within the same layer. This key parameter will be further analyzed in subsequent

Method	ACM		Freebase		AMiner		DBLP	
	Ma-F1	Mi-F1	Ma-F1	Mi-F1	Ma-F1	Mi-F1	Ma-F1	Mi-F1
10 shot								
HAN	82.37±0.2	82.27±0.4	33.24±1.6	35.36±1.7	60.95±0.8	65.84±1.0	84.93±1.3	84.26±2.9
HGT	70.52±0.3	71.26±0.5	30.93±1.4	34.26±1.9	55.97±2.9	59.26±3.3	75.39±3.2	76.06±1.6
HGPROMPT	78.39±0.5	83.23±0.3	36.48±1.8	45.97±1.8	67.96±2.3	77.05±1.4	89.02±0.2	90.05±0.3
HDMI	87.83±0.4	88.78±0.3	36.78±1.4	45.43±1.7	68.12±1.6	76.93±2.2	89.98±0.4	90.91±1.0
+HetGPT	90.69±0.6	89.15±0.4	37.02±1.6	45.77±2.0	68.37±1.3	77.22±1.9	90.31±0.7	91.27±1.2
+HGMRP+	**90.90±0.8**	**89.58±0.5**	37.91±1.8	46.18±1.7	**68.85±1.1**	77.69±1.7	**90.78±0.5**	**91.86±1.4**
HeCo	87.56±0.1	86.63±0.2	36.43±1.2	45.15±1.6	67.86±1.8	76.55±1.9	89.64±0.6	90.43±0.9
+HetGPT	88.02±0.1	87.15±0.2	36.82±1.4	45.56±1.2	68.32±0.8	77.02±1.9	90.12±0.3	90.93±0.2
+HGMRP+	88.31±0.2	87.47±0.2	37.45±1.3	45.91±1.3	68.67±0.5	77.41±2.2	90.43±0.5	91.27±0.2
HGMAE	88.10±0.1	87.38±0.1	47.33±1.3	52.75±1.8	68.32±0.1	77.01±0.1	88.25±0.1	87.27±0.1
+HetGPT	-	-	-	-	-	-	-	-
+HGMRP+	88.48±0.3	87.91±0.3	**48.86±1.8**	**55.41±1.4**	68.79±0.5	**78.53±0.5**	88.92±0.5	89.86±0.4
20 shot								
HAN	85.41±1.2	85.53±1.3	50.16±0.1	49.95±0.1	66.89±0.5	72.16±0.4	86.49±0.6	84.23±1.1
HGT	87.80±2.8	87.96±5.4	54.87±0.1	51.55±0.3	65.43±0.3	69.48±0.5	87.46±1.1	83.88±0.3
HGPROMPT	87.39±0.3	84.52±0.3	56.43±1.2	61.18±1.8	70.65±0.5	78.65±0.5	92.73±0.2	91.69±0.3
HDMI	90.56±0.8	90.51±0.8	58.64±0.2	61.32±0.4	73.47±0.5	79.55±0.5	91.30±0.2	92.16±0.1
+HetGPT	92.05±0.3	90.85±0.3	58.76±0.3	62.12±0.5	74.17±0.3	79.66±0.1	91.64±0.3	92.43±0.3
+HGMRP+	**92.35±0.4**	**91.15±0.3**	59.63±0.4	62.62±0.6	**74.37±0.6**	79.79±0.2	91.84±0.4	**93.13±0.3**
HeCo	87.93±0.4	86.71±0.4	57.84±0.2	60.12±0.4	71.87±0.5	79.15±0.5	89.02±1.4	91.78±1.8
+HetGPT	88.43±0.6	88.52±0.6	59.53±1.0	62.10±1.6	72.25±0.6	79.70±0.6	91.17±0.4	92.05±0.5
+HGMRP+	88.98±0.3	89.40±0.4	59.98±1.2	63.40±1.4	72.85±0.6	**80.92±0.7**	91.60±0.2	92.40±0.2
HGMAE	90.36±0.4	90.15±0.6	61.42±0.2	64.61±1.1	72.52±0.3	80.07±0.6	91.42±0.6	92.17±0.6
+HetGPT	-	-	-	-	-	-	-	-
+HGMRP+	91.74±0.6	91.92±0.4	**63.41±0.2**	**65.52±0.2**	72.91±0.3	80.87±0.3	**92.88±0.0**	92.67±0.1

Table 2: Node classification results. Best results are in bold, and improvements brought by HGMRP are underlined. The '+' indicates the integration of the prompt method.

ablation studies. During training, we adopt the same learning rate as in the original pre-training stage for fine-tuning the classifier, and increase the number of training epochs to 200 with early stopping enabled. The Adam optimizer (PyTorch implementation) is adopted for parameter optimization. All experiments are conducted on an RTX 5090 GPU.

5.2 Node Classification (RQ1)

To investigate the capability of our prompt-tuning approach in few-shot settings, we conduct 10-shot and 20-shot experiments on four datasets. The results are summarized in Table 2. Based on the results, the following conclusions can be drawn: 1) Under both 10-shot and 20-shot settings, mainstream pre-training methods (HDMI, HeCo, HGMAE) consistently outperform traditional semi-supervised heterogeneous graph methods, demonstrating the effectiveness of pre-training. More importantly, our proposed relation-aware tuning approach achieves noticeable gains over existing prompt-tuning methods. This indicates that controlling message propagation through different relation edges is more effective in this experimental setup than simply adding prompts on nodes for propagation. 2) Our designed prompt mechanism can be seamlessly integrated into various mainstream pre-trained models and consistently yields strong performance. In contrast, HGPROMPT only reports results in a single setting due to its own proprietary framework, and HetGPT is restricted to the HeCo format. This demonstrates that our approach not only preserves performance but also significantly enhances the generalization ability of heterogeneous prompts across different pre-trained architectures.

5.3 Impact of Anchor Prompt Number (RQ2)

In our method, employing a single prompt per relation type per layer results in limited expressive capacity, while assigning a distinct prompt to every edge can lead to underfitting due to scarce labeled data. To balance these trade-offs, we

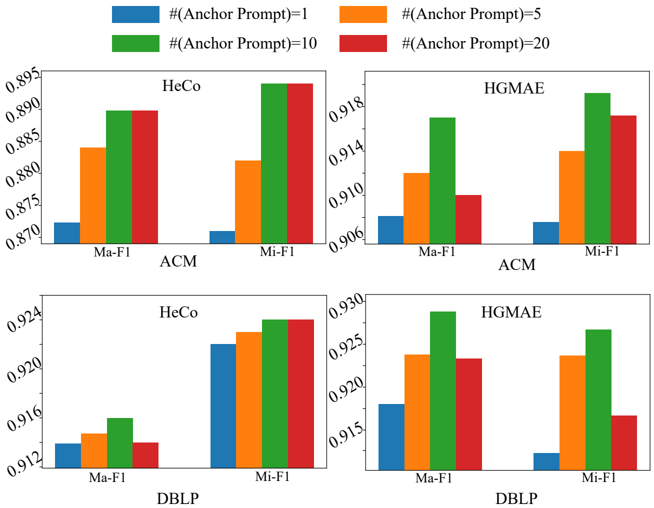

Figure 3: Impact of anchor prompt number

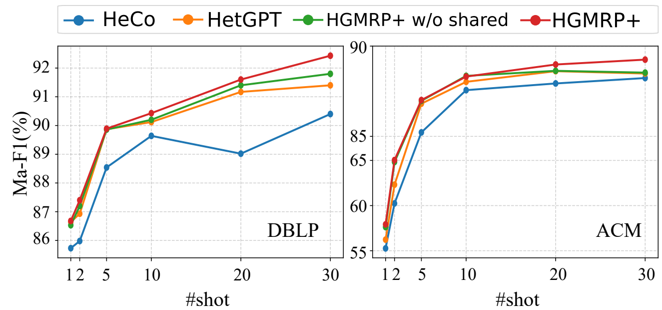

Figure 4: Impact of shot number

introduce HGMRP+, in which the size of the anchor prompt set acts as a key hyperparameter. Setting this size to 1 reverts the model to HGMRP, and letting it approach the total number of edges mimics the case of a dedicated prompt per edge. Therefore, we design the following experiment to examine the impact of the number of anchor prompts. All experiments are conducted under a 20-shot setting. The results are shown in Figure 3. It can be observed that with limited labeled data, a configuration of 10 anchor prompts consistently outperforms the use of 20 prompts across different datasets and pre-trained models. This is because an excessively large set of anchor prompts leads to insufficient training. Thus, an appropriate number of prompts is essential to fully exploit the expressive power of the pre-trained model.

5.4 Performance with Different Shots (RQ3)

The previous experiments fixed the amount of labeled data during fine-tuning. To analyze the impact of varying sample sizes on model performance, we conduct experiments with different numbers of shots. We also examine the scenario in HGMRP where anchor prompts are not shared across different relations within the same layer, denoted in the figure as HGMRP w/o shared. The experimental results are presented in Figure 4. Based on the above results, it is observed that across varying label settings, our method consistently outperforms existing prompt techniques. Furthermore, the designed shared anchor prompts demonstrate superior effectiveness compared to the non-shared version, confirming the value of sharing anchor prompts across different semantic relations. Even in the extreme low-data scenario of 1-shot, where performance is understandably lower due to insufficient prompt training, our approach still achieves better results than the original pre-training method.

5.5 Integration with Other HG Prompt (RQ4)

Our method is grounded in multi-relational propagation on heterogeneous graphs, which does not conflict in application with existing prompts designed for node representation in such graphs. To investigate the practical effect of using both types jointly, we integrate our approach with HetGPT and HGPROMPT, the former employs HeCo as the pre-training task, while the latter uses its own pre-training framework. The results are shown in Table 3, demonstrating that in heterogeneous graphs, our message-passing-based method is compatible with node-level representation and feature-oriented prompts, and further enhances the performance of the original methods.

Dataset	Metric	HGP	+HGMRP	HetGPT	+HGMRP
ACM	Ma-F1	77.29±0.2	80.71±0.3	88.43±0.6	**90.15±0.8**
	Mi-F1	72.54±0.1	75.76±0.6	88.52±0.6	**89.17±0.3**
	AUC	79.86±0.2	80.53±0.2	96.73±0.3	**96.93±0.4**
DBLP	Ma-F1	92.73±0.2	**92.92±0.3**	91.17±0.4	91.24±0.7
	Mi-F1	91.69±0.3	**92.73±0.3**	92.05±0.5	92.33±0.4
	AUC	98.43±0.3	**98.73±0.3**	96.90±0.4	97.24±0.2

Table 3: Performance comparison across different HG prompt methods with best bolded, improved underlined and "+" indicates HGMRP plugged in.

6 Conclusion

In this paper, we propose a relation-edge prompt termed HGMRP. It addresses the lack of a unified design in existing heterogeneous prompt methods, which predominantly operate on node features or representations and may lead to inconsistent optimization objectives when propagating across neighboring relations. An enhanced version, HGMRP+, is further introduced. Our approach is compatible with mainstream pre-trained heterogeneous graph models. Extensive experimental results demonstrate that our method outperforms existing heterogeneous graph prompting techniques, confirming the strong fine-tuning capability of relation-edge-level prompts. While the proposed method achieves strong results in single-domain settings, future work will explore cross-domain scenarios with diverse relational semantics and investigate serializing such relation-edge prompts for integration with large language models.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2023YFC3304503), the National Natural Science Foundation of China (No. 62422210, 92370111, 62302333, 62272340, 62276187).

References

[Brown *et al.*, 2020] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901, 2020.

[Cao *et al.*, 2025] Yuwei Cao, Liangwei Yang, Zhiwei Liu, Yuqing Liu, Chen Wang, Yueqing Liang, Hao Peng, and Philip S. Yu. Graph-sequential alignment and uniformity: Toward enhanced recommendation systems. In *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025*, pages 888–892. ACM, 2025.

[Devlin *et al.*, 2019] 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, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[Fang *et al.*, 2023] Taoran Fang, Yunchao Zhang, Yang Yang, Chunping Wang, and Lei Chen. Universal prompt tuning for graph neural networks. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

[Fu *et al.*, 2020a] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. MAGNN: metapath aggregated graph neural network for heterogeneous graph embedding. In *WWW '20: The Web Conference 2020*, pages 2331–2341, 2020.

[Fu *et al.*, 2020b] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. MAGNN: metapath aggregated graph neural network for heterogeneous graph embedding. In *WWW '20: The Web Conference 2020*, pages 2331–2341. ACM / IW3C2, 2020.

[Fu *et al.*, 2025] Xingbo Fu, Yinhan He, and Jundong Li. Edge prompt tuning for graph neural networks. In *The Thirteenth International Conference on Learning Representations, ICLR 2025*, 2025.

[Hu *et al.*, 2019] Binbin Hu, Yuan Fang, and Chuan Shi. Adversarial learning on heterogeneous information networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019*, pages 120–129, 2019.

[Hu *et al.*, 2020] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *WWW '20: The Web Conference 2020*, pages 2704–2710, 2020.

[Huang *et al.*, 2026] Yongqi Huang, Jitao Zhao, Dongxiao He, Xiaobao Wang, Yawen Li, Yuxiao Huang, Di Jin, and Zhiyong Feng. One prompt fits all: Universal graph adaptation for pretrained models. *Advances in Neural Information Processing Systems*, 38:85497–85530, 2026.

[Hui *et al.*, 2025] Yunming Hui, Inez Maria Zwetsloot, Simon Trimborn, and Stevan Rudinac. Domain-informed negative sampling strategies for dynamic graph embedding in meme stock-related social networks. In *Proceedings of the ACM on Web Conference 2025, WWW 2025*, pages 518–529. ACM, 2025.

[Jiao *et al.*, 2025] Pengfei Jiao, Jialong Ni, Di Jin, Xuan Guo, Huan Liu, Hongjiang Chen, and Yanxian Bi. HGMP: heterogeneous graph multi-task prompt learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2025*, pages 2982–2990, 2025.

[Jing *et al.*, 2021] Baoyu Jing, Chanyoung Park, and Hanghang Tong. HDMI: high-order deep multiplex infomax. In *WWW '21: The Web Conference 2021*, pages 2414–2424, 2021.

[Liu *et al.*, 2023] Zemin Liu, Xingtong Yu, Yuan Fang, and Xinming Zhang. Graphprompt: Unifying pre-training and downstream tasks for graph neural networks. In *Proceedings of the ACM Web Conference 2023*, page 417–428, 2023.

[Lv *et al.*, 2021] Qingsong Lv, Ming Ding, Qiang Liu, Yuxiang Chen, Wenzheng Feng, Siming He, Chang Zhou, Jianguo Jiang, Yuxiao Dong, and Jie Tang. Are we really making much progress?: Revisiting, benchmarking and refining heterogeneous graph neural networks. In *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1150–1160, 2021.

[Ma *et al.*, 2024] Yihong Ma, Ning Yan, Jiayu Li, Masood S. Mortazavi, and Nitesh V. Chawla. Hetgpt: Harnessing the power of prompt tuning in pre-trained heterogeneous graph neural networks. In *Proceedings of the ACM on Web Conference 2024, WWW 2024*, pages 1015–1023. ACM, 2024.

[Mo *et al.*, 2024] Yujie Mo, Feiping Nie, Ping Hu, Heng Tao Shen, Zheng Zhang, Xinchao Wang, and Xiaofeng Zhu. Self-supervised heterogeneous graph learning: a homophily and heterogeneity view. In *The Twelfth International Conference on Learning Representations, ICLR 2024*, 2024.

[Park *et al.*, 2020] Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. Unsupervised attributed multiplex network embedding. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 5371–5378. AAAI Press, 2020.

[Radford *et al.*, 2018] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training, 2018.

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

[Tian *et al.*, 2023] Yijun Tian, Kaiwen Dong, Chunhui Zhang, Chuxu Zhang, and Nitesh V. Chawla. Heterogeneous graph masked autoencoders. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023*, pages 9997–10005, 2023.

[Velickovic *et al.*, 2018] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.

[Wang *et al.*, 2019] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S. Yu. Heterogeneous graph attention network. In *The World Wide Web Conference, WWW 2019*, pages 2022–2032, 2019.

[Wang *et al.*, 2021] Xiao Wang, Nian Liu, Hui Han, and Chuan Shi. Self-supervised heterogeneous graph neural network with co-contrastive learning. In *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1726–1736, 2021.

[Yu *et al.*, 2024] Xingtong Yu, Yuan Fang, Zemin Liu, and Xinming Zhang. Hgprompt: Bridging homogeneous and heterogeneous graphs for few-shot prompt learning. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024*, pages 16578–16586. AAAI Press, 2024.

[Zhao *et al.*, 2020] Jianan Zhao, Xiao Wang, Chuan Shi, Zekuan Liu, and Yanfang Ye. Network schema preserving heterogeneous information network embedding. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 1366–1372. ijcai.org, 2020.