Universal Prompt Tuning for Graph Neural Networks

Taoran Fang[1], Yunchao Zhang[1], Yang Yang[1*],
Chunping Wang[2],Lei Chen[2]

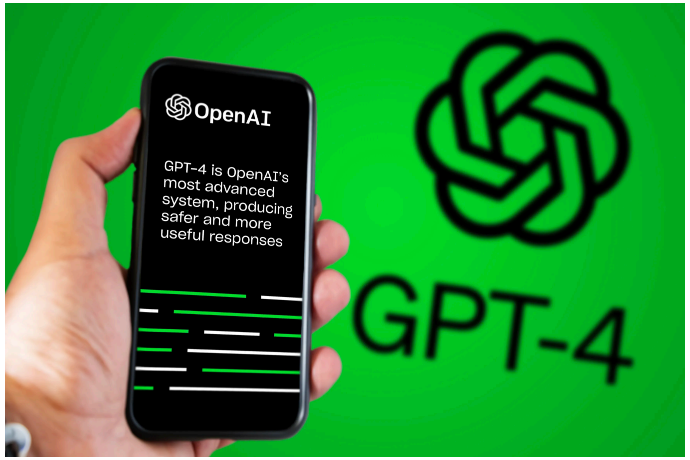
[1]Zhejiang University
[2]Finvolution Group

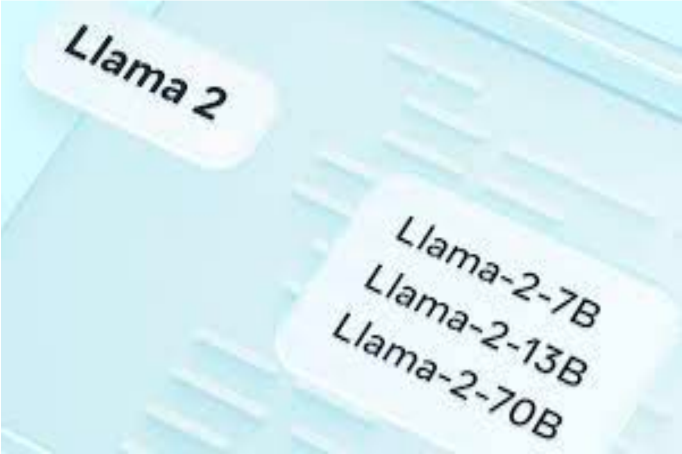
Background

Prompt tuning has achieved a great success in adapting large language model (LLM).

- e.g. GPT-4, Llama 2, ChatGLM …

This technique leads the way for adapting pre-trained models in a new direction.

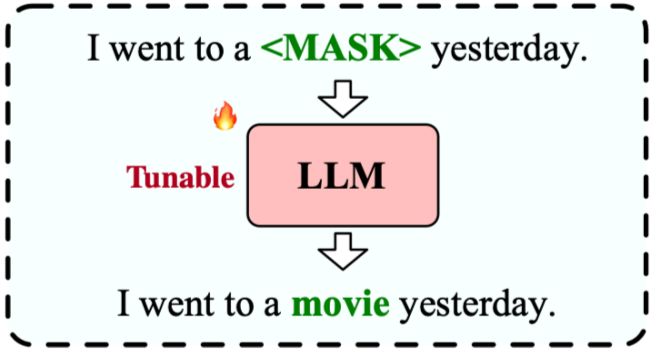




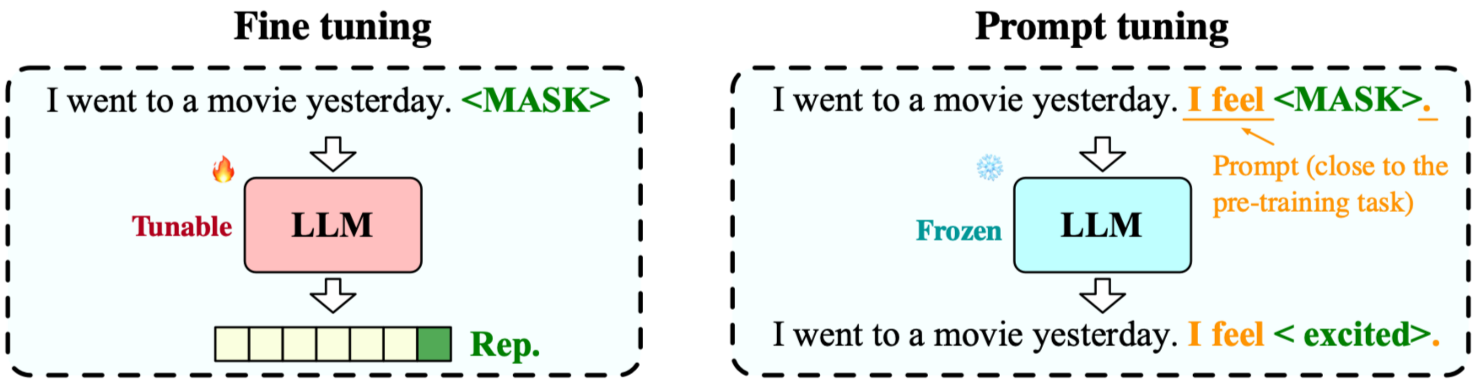
Background

Prompt tuning a pre-trained LLM

- Step 1: Pre-training an LLM using the Masked Language Modeling (MLM).

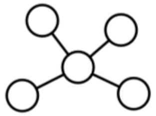


- Step 2: Reformulating the downstream task by a prompt on the input sentence.



Background

Pre-trained LLM vs Pre-trained GNNs

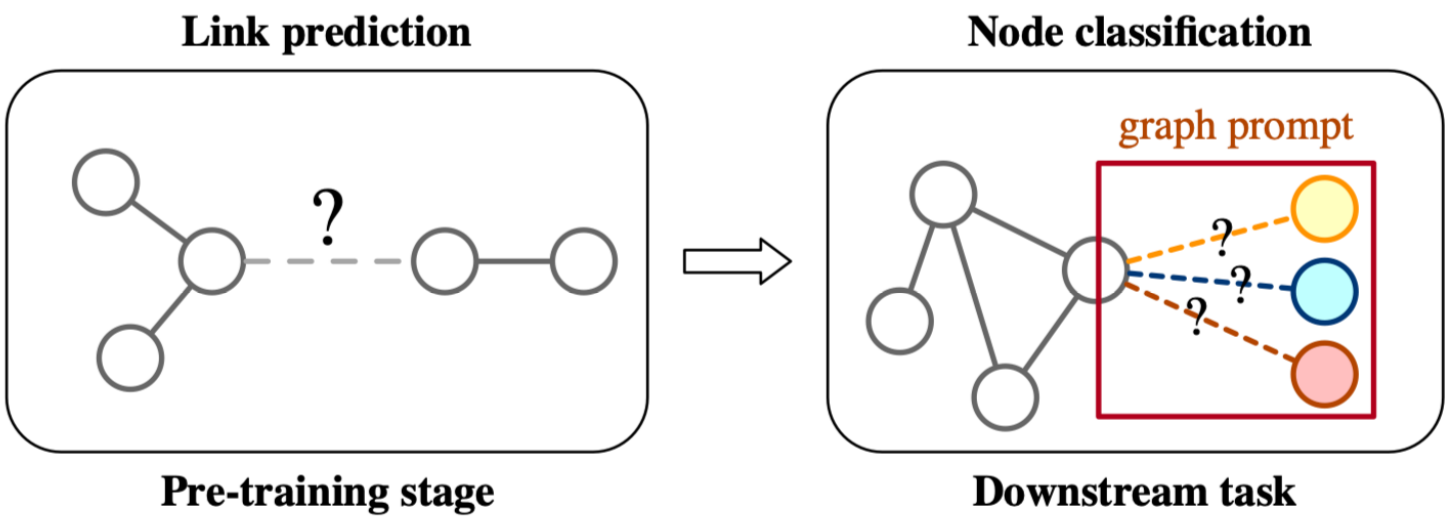
	LLM	**GNNs**
Input	[Sentence] *I went to a movie yesterday.*	[Graph] 
Pre-training Task	Masked language modeling (MLM)	Link prediction, Attribute masking, Contrastive learning, …
Prompt Template	*I went to a movie yesterday.* *I feel <MASK>.*	?

How to apply prompt tuning on pre-trained GNNs?

Background

Existing graph prompt tuning methods for GNNs.

- Some pioneering works GPPT[1] and GraphPrompt[2] utilize graph prompt tuning by modifying the downstream task to the **link prediction**, which is consistent with the pre-training strategy they use.



[1] Mingchen Sun et al. “GPPT: Graph Pre-training and Prompt Tuning to Generalize Graph Neural Networks.”
[2] Zemin Liu et al. “GraphPrompt: Unifying Pre-Training and Downstream Tasks for Graph Neural Networks.”

Background

Limitations

❑In practice

- There is no unified pre-training task for GNNs, making it challenging to design general prompting functions.
- Existing methods have limited applicability and are only compatible with models pre-trained by the link prediction.

❑In theory

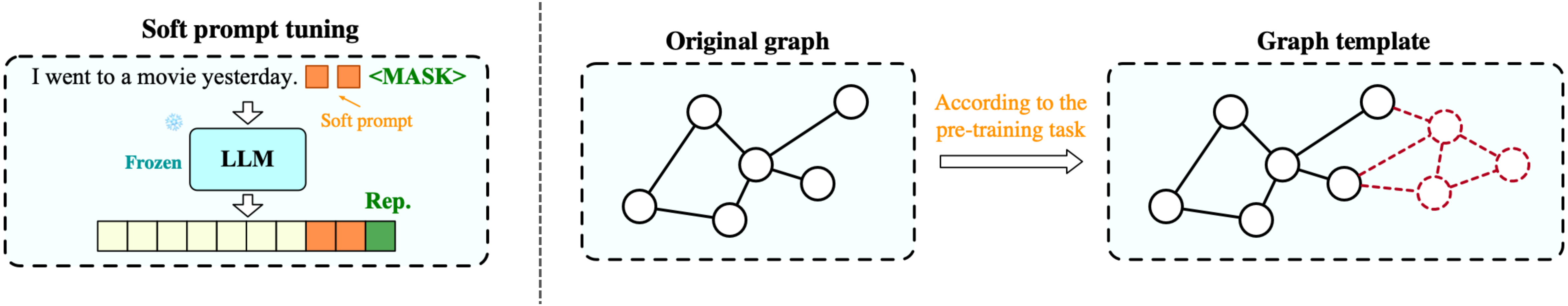
- Existing prompt-based tuning methods for GNN models are designed based on intuition, lacking theoretical guarantees for their effectiveness.

Methodology

Graph prompt tuning

- Step 1: ***Template design***. We generate the graph template, which includes learnable components in its adjacency matrix and feature matrix.

$$\mathcal{G}^*: (\mathbf{A}^*, \mathbf{X}^*) = \psi_t(\mathcal{G})$$



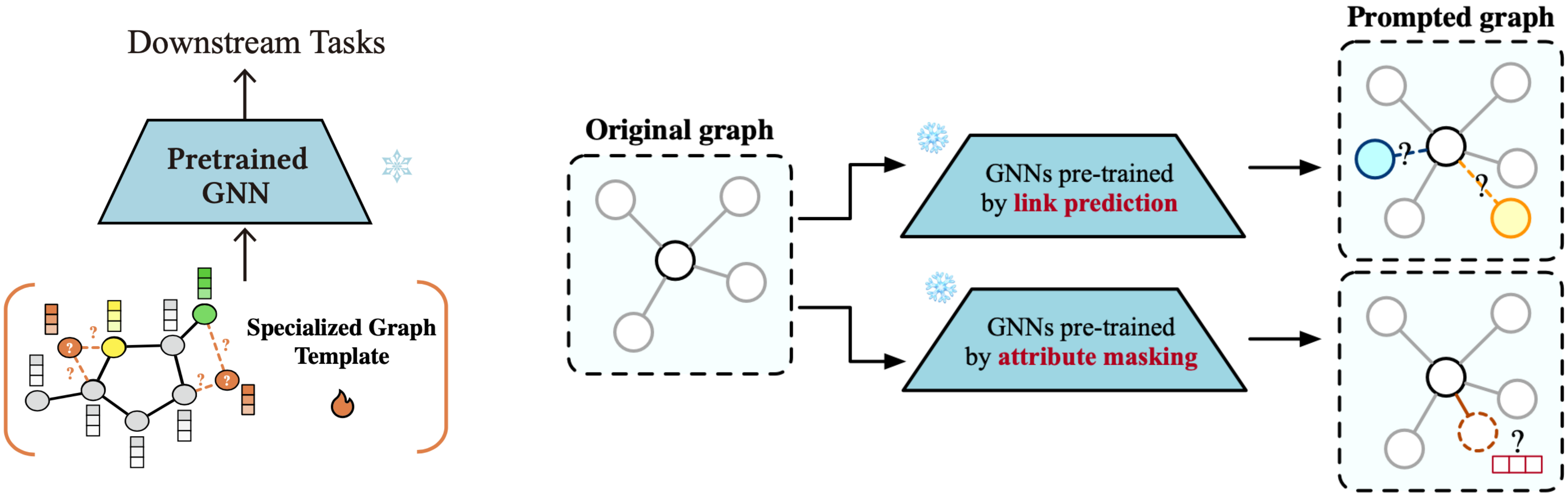
- Step 2: ***Prompt optimization***. We search for the optimal prompt parameters according to the downstream task.

$$\max_{\hat{\mathbf{A}} \in \mathbb{A}, \hat{\mathbf{X}} \in \mathbb{X}, \theta} P_{f,\theta}(y|\mathcal{G}^*)$$

Methodology

Specialized graph prompt tuning

- According to the motivation of prompt tuning, the graph prompt design is close related to the pre-training task involved.



However, there are so many pre-training strategies in the graph field. Can we design a universal graph prompt tuning method for all these strategies?

Methodology

Universal graph prompt tuning

- **Graph Prompt Feature (GPF)**

GPF focuses on incorporating additional learnable parameters into the feature space of the input graph.

$$p \in \mathbb{R}^F$$

The learnable vector p is added to the graph features $\mathbf{X}$ to generate the prompted features $\mathbf{X}^*$.

$$\mathbf{X} = \{x_1, x_2, \ldots, x_N\} \quad \mathbf{X}^* = \{x_1 + p, x_2 + p, \ldots, x_N + p\}$$

- **Graph Prompt Feature-Plus (GPF-plus)**

GPF-plus sets a different feature vector for each node in the graph.

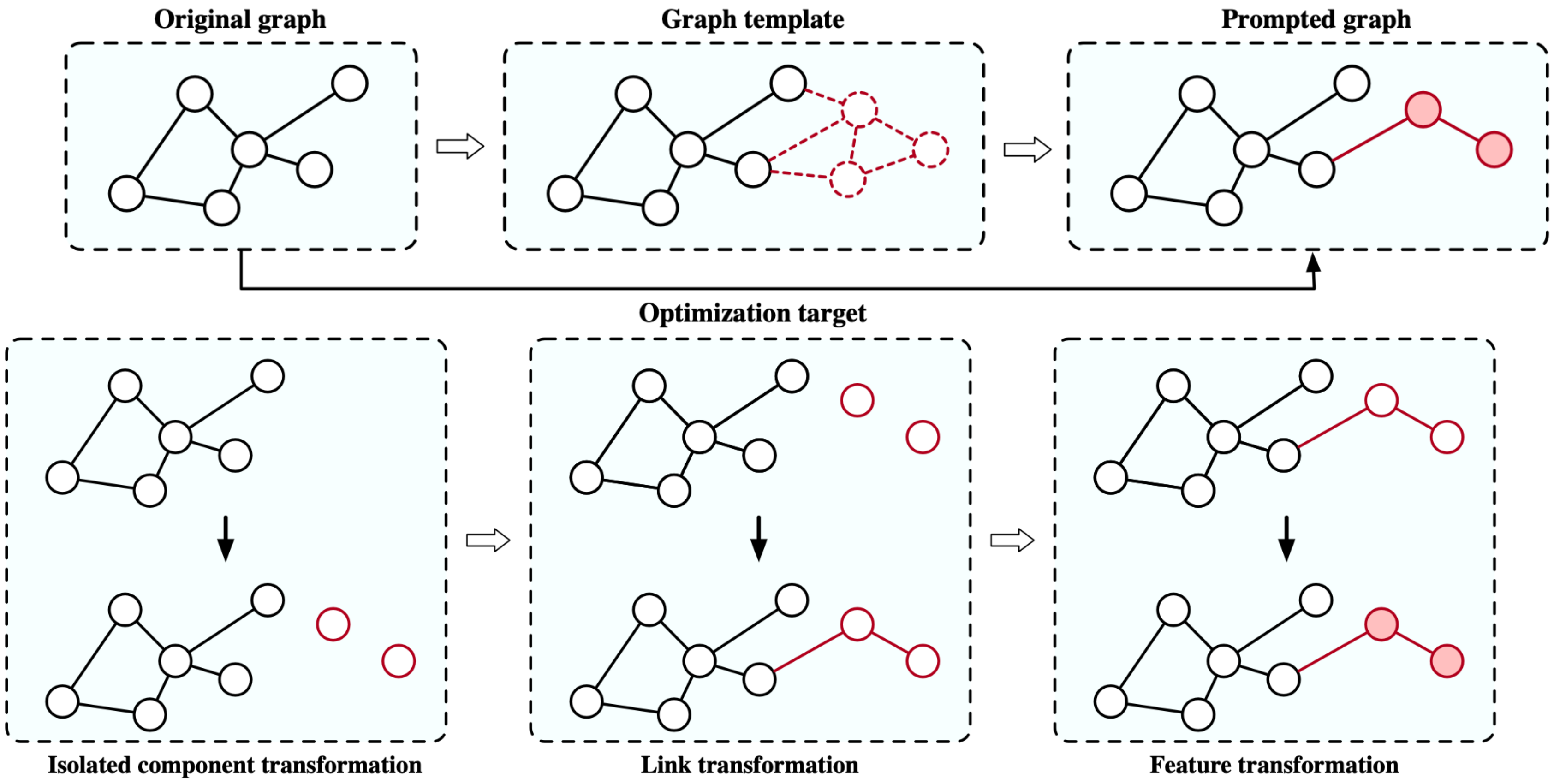
$$p_1, p_2, \ldots p_N \in \mathbb{R}^F$$

$$\mathbf{X} = \{x_1, x_2, \ldots, x_N\} \quad \mathbf{X}^* = \{x_1 + p_1, x_2 + p_2, \ldots, x_N + p_N\}$$

Theoretical Analysis

Rethinking the process of graph prompt tuning

- Complex *template design* and *prompt optimization* can be divided into several simple steps.



Theoretical Analysis

Rethinking the process of graph prompt tuning

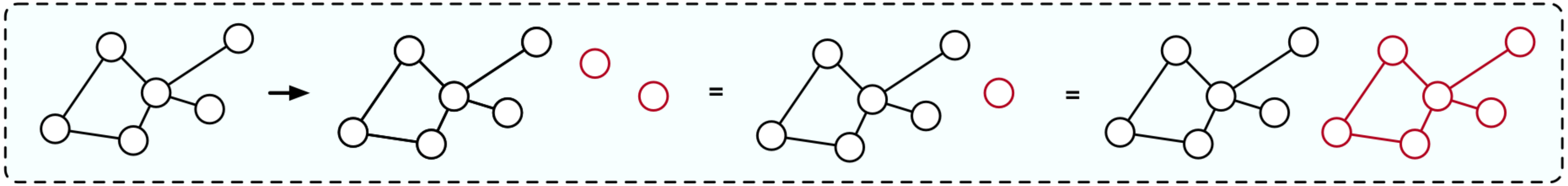
We assume the pre-trained GNN model is a single layer *GIN* with *sum* pooling.

$$\mathbf{H} = (\mathbf{A} + (1+\epsilon) \cdot \mathbf{I}) \cdot \mathbf{X} \cdot \mathbf{W}$$

$$h_{\mathcal{G}} = \sum_{v_i \in \mathcal{V}} h_i$$

- **Isolated component transformation**

We can find a GPF that achieves equivalent effect

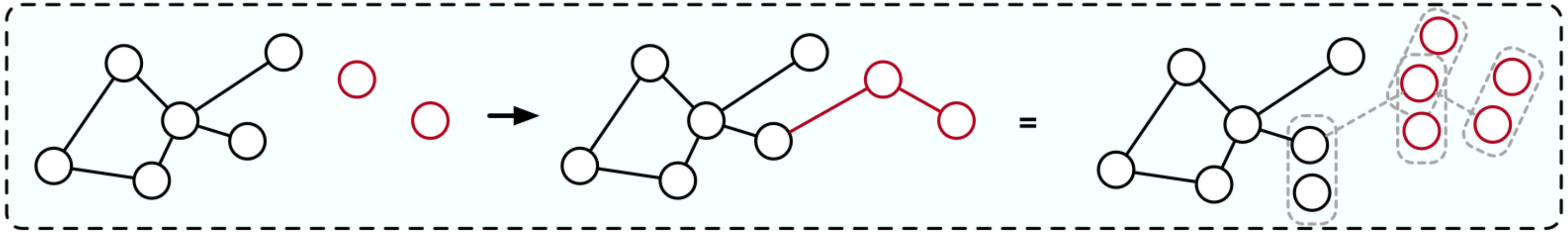


p_1

Theoretical Analysis

- **Link transformation**

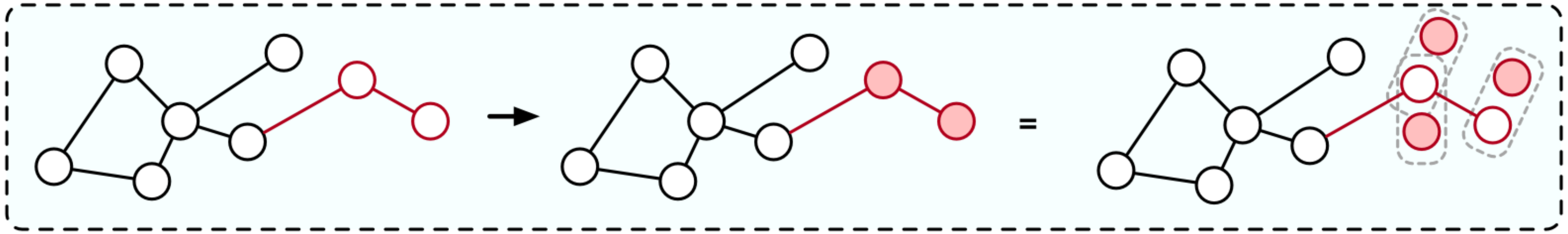
We can find a GPF that achieves equivalent effect



p_2

- **Feature transformation**

We can find a GPF that achieves equivalent effect



p_3

Therefore, we can find a GPF $p = p_1 + p_2 + p_3$ that achieves the equivalent effect.

Theoretical Analysis

The universal capability of GPF

Theorem 1. Given a pre-trained GNN model f, an input graph $\mathcal{G}: (\mathbf{A}, \mathbf{X})$, an arbitrary prompting function $\psi_t(\cdot)$, for any prompted graph $\hat{\mathcal{G}}: (\widehat{\mathbf{A}}, \widehat{\mathbf{X}})$ in the candidate space of the graph template $\mathcal{G}^* = \psi_t(\mathcal{G})$, there exists a GPF extra feature vector $\hat{p}$ that satisfies:

$$f(\mathbf{A}, \mathbf{X} + \hat{p}) = f(\hat{\mathbf{A}}, \hat{\mathbf{X}})$$

GPF can achieve equivalent performance to any specialized graph prompting method. This conclusion inspires many future works such as [1].

[1] Xiangguo Sun et al. "All in One: Multi-Task Prompting for Graph Neural Networks."

Theoretical Analysis

The effectiveness guarantee of GPF

Theorem 2. For a pre-trained GNN model f, graphs $D = \{(\mathcal{G}_1: (\mathbf{A}_1, \mathbf{X}_1)), \ldots, (\mathcal{G}_m: (\mathbf{A}_m, \mathbf{X}_m))\}$ under the non-degeneracy condition, and a linear projection head θ, there exists $\mathcal{Y} = \{y'_1, \ldots, y'_m\}$ for $y_1 = y'_1, \ldots, y_m = y'_m$ that satisfies:

$$l_{\mathrm{GPF}} = \min_{p,\theta} \sum_{i}^{m} (f(\mathbf{A}_i, \mathbf{X}_i + p) \cdot \theta - y_i)^2 < l_{\mathrm{FT}} = \min_{f,\theta} \sum_{i}^{m} (f(\mathbf{A}_i, \mathbf{X}_i) \cdot \theta - y_i)^2$$

GPF is not weaker than fine-tuning.

$$\mathrm{H} = \boxed{\mathrm{A} \cdot \mathrm{X}} \cdot \boxed{\mathrm{W}}$$

Input tuning: prompt tuning.

Model tuning: fine-tuning

We are the first to compare the effectiveness of prompt tuning to fine-tuning.

Empirical Analysis

Pre-training Strategy	Tuning Strategy	BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE	PPI	**Avg.**
Infomax	FT	**67.55** ±2.06	78.57 ±0.51	65.16 ±0.53	63.34 ±0.45	70.06 ±1.45	**81.42** ±2.65	77.71 ±0.45	81.32 ±1.25	71.29 ±1.79	72.93
	GPF	66.83 ±0.86	79.09 ±0.25	66.10 ±0.53	**66.17** ±0.81	73.56 ±3.94	80.43 ±0.53	76.49 ±0.18	83.60 ±1.00	77.02 ±0.42	74.36
	GPF-plus	67.17 ±0.36	**79.13** ±0.70	**66.35** ±0.37	65.62 ±0.74	**75.12** ±2.45	81.33 ±1.52	**77.73** ±1.14	**83.67** ±1.08	**77.03** ±0.32	**74.79**
AttrMasking	FT	66.33 ±0.55	78.28 ±0.05	65.34 ±0.30	66.77 ±0.13	74.46 ±2.82					
	GPF	**68.09** ±0.38	**79.04** ±0.90	66.32 ±0.42	**69.13** ±1.16	75.06 ±1.02					
	GPF-plus	67.71 ±0.64	78.87 ±0.31	**66.58** ±0.13	68.65 ±0.72	**76.17** ±2.98					
ContextPred	FT	**69.65** ±0.87	78.29 ±0.44	66.39 ±0.57	64.45 ±0.6	73.71 ±1.57					
	GPF	68.48 ±0.88	79.99 ±0.24	**67.92** ±0.35	66.18 ±0.46	74.51 ±2.72					
	GPF-plus	69.15 ±0.82	**80.05** ±0.46	67.58 ±0.54	**66.94** ±0.95	**75.25** ±1.88	**84.48** ±0.78	78.40 ±0.16	**85.81** ±0.43	**77.71** ±0.21	**76.15**
GCL	FT	69.49 ±0.35	73.35 ±0.70	62.54 ±0.26	60.63 ±1.26	**75.17** ±2.14	69.78 ±1.44	**78.26** ±0.73	75.51 ±2.01	67.76 ±0.78	70.27
	GPF	71.11 ±1.20	**73.64** ±0.25	62.70 ±0.46	61.26 ±0.53	72.06 ±2.98	70.09 ±0.67	75.52 ±1.09	78.55 ±0.56	67.60 ±0.57	70.28
	GPF-plus	**72.18** ±0.93	73.35 ±0.43	**62.76** ±0.75	**62.37** ±0.38	73.90 ±2.47	**72.94** ±1.87	77.51 ±0.82	**79.61** ±2.06	**67.89** ±0.69	**71.39**

Dataset	Tuning Strategy	Tunable Parameters	Relative Ratio (%)
Chemistry	FT	$\sim$ 1.8M	100
	GPF	$\sim$ 0.3K	0.02
	GPF-plus	$\sim$ 3-12K	0.17-0.68
Biology	FT	$\sim$ 2.7M	100
	GPF	$\sim$ 0.3K	0.01
	GPF-plus	$\sim$ 3-12K	0.11-0.44

GPF and GPF-plus achieved better results than fine-tuning in 80% of the experiments.

Empirical Analysis

Comparison with existing graph prompt-based methods

Pre-training Strategy	Tuning Strategy	BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE	PPI	**Avg.**
EdgePred	FT	66.56 ±3.56	78.67 ±0.35	**66.29** ±0.45	64.35 ±0.78	69.07 ±4.61	79.67 ±1.70	77.44 ±0.58	80.90 ±0.92	71.54 ±0.85	72.72
	GPPT	64.13 ±0.14	66.41 ±0.04	60.34 ±0.14	54.86 ±0.25	59.81 ±0.46	63.05 ±0.34	60.54 ±0.54	70.85 ±1.42	56.23 ±0.27	61.80
	GPPT (w/o ol)	69.43 ±0.18	78.91 ±0.15	64.86 ±0.11	60.94 ±0.18	62.15 ±0.69	82.06 ±0.53	73.19 ±0.19	70.31 ±0.99	76.85 ±0.26	70.97
	GraphPrompt	69.29 ±0.19	68.09 ±0.19	60.54 ±0.21	58.71 ±0.13	55.37 ±0.57	62.35 ±0.44	59.31 ±0.93	67.70 ±1.26	49.48 ±0.96	61.20
	GPF	**69.57** ±0.21	79.74 ±0.03	65.65 ±0.30	67.20 ±0.99	**69.49** ±5.17	82.86 ±0.23	77.60 ±1.45	81.57 ±1.08	76.98 ±0.20	74.51
	GPF-plus	69.06 ±0.68	**80.04** ±0.06	65.94 ±0.31	**67.51** ±0.59	68.80 ±2.58	**83.13** ±0.42	**77.65** ±1.90	**81.75** ±2.09	**77.00** ±0.12	**74.54**

GPF and GPF-plus achieved better results than specialized graph prompt methods by a large margin.

Empirical Analysis

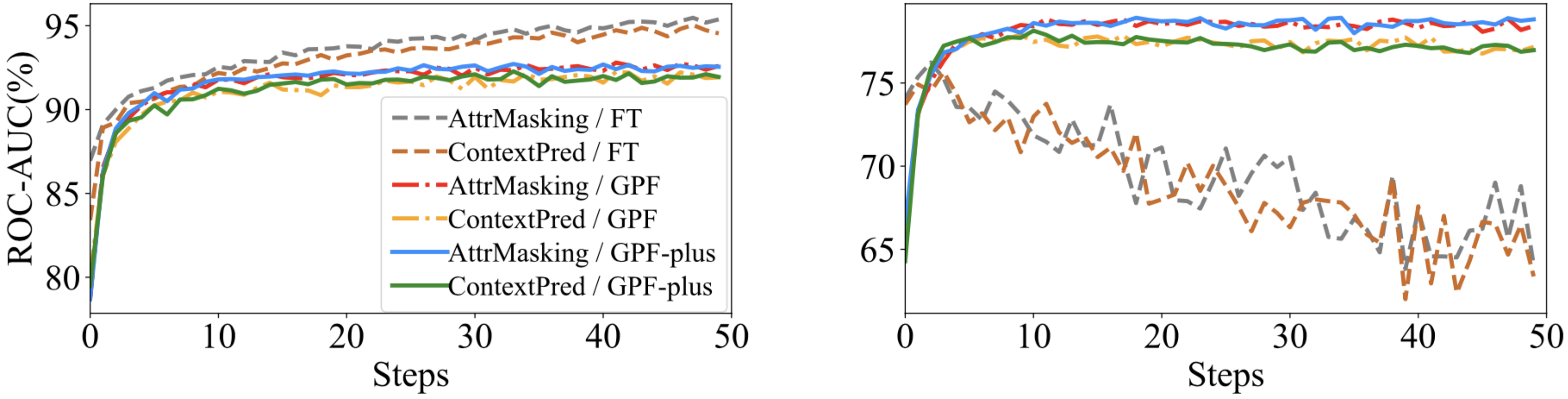
Full-shot (50-shot) experiments

Pre-training Strategy	Tuning Strategy	BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE	PPI	**Avg.**
Infomax	FT	53.81 ±3.35	61.42 ±1.19	53.93 ±0.59	50.77 ±2.27	58.6 ±3.48	66.12 ±0.63	65.09 ±1.17	52.64 ±2.64	48.79 ±1.32	56.79
	GPF	55.52 ±1.84	65.56 ±0.64	56.76 ±0.54	50.29 ±1.61	62.44 ±4.11	**68.00** ±0.61	**67.68** ±1.09	54.49 ±2.54	**54.03** ±0.34	59.41
	GPF-plus	**58.09** ±2.12	**65.71** ±0.37	**57.13** ±0.48	**51.33** ±1.14	**62.96** ±3.27	67.88 ±0.42	66.80 ±1.43	**56.56** ±6.81	53.78 ±0.45	**60.02**
EdgePred	FT	48.88 ±0.68	60.95 ±1.46	55.73 ±0.43	51.30 ±2.21	57.78 ±4.03	66.88 ±0.53	64.22 ±1.57	61.27 ±6.10	47.62 ±1.50	57.18
	GPF	50.53 ±1.35	64.46 ±0.93	57.33 ±0.65	**51.35** ±0.76	**68.74** ±6.03	68.08 ±0.39	**66.22** ±1.90	**62.85** ±5.91	52.81 ±0.38	**60.26**
	GPF-plus	**54.49** ±4.60	**64.99** ±0.53	**57.69** ±0.61	51.30 ±1.18	66.64 ±2.40	**68.16** ±0.48	62.05 ±3.39	62.60 ±2.48	**53.30** ±0.34	60.13
AttrMasking	FT	51.26 ±2.33	60.28 ±1.73	53.47 ±0.46	50.11 ±1.63	61.51 ±1.45	59.35 ±1.31	67.18 ±1.59	55.62 ±5.04	48.17 ±2.45	56.32
	GPF	54.24 ±0.74	64.24 ±0.40	**56.84** ±0.28	**50.62** ±0.88	**65.34** ±1.93	61.34 ±0.60	67.94 ±0.48	**57.31** ±6.71	51.26 ±0.32	58.79
	GPF-plus	**58.10** ±1.92	**64.39** ±0.30	56.78 ±0.25	50.30 ±0.78	63.34 ±0.85	**63.84** ±1.13	**68.05** ±0.97	57.29 ±4.46	**51.35** ±0.32	**59.27**

GPF and GPF-plus have a greater advantage over fine-tuning in few-shot scenarios.

Empirical Analysis

Training process analysis



(a) Training Curve (b) Test Curve

Fully fine-tuning a pre-trained GNN model may lose the model's generalization ability. GPF and GPF-plus can significantly alleviate this issue and maintain superior performance on the test set.

Universal Prompt Tuning for Graph Neural Networks

- ✓ We propose a universal prompt tuning method for graph neural networks, which can be applied to the models pre-trained by any strategy.
- ✓ We provide theoretical guarantees and design principles for graph prompt tuning, offering valuable insights for future investigations in this field.

THANKS | Q&A

More relevant research of our group: ***http://yangy.org***

Contact: ***fangtr@zju.edu.cn***, ***yangya@zju.edu.cn***

Github: ***https://github.com/zjunet/GPF***

Group Homepage

Code Repository