THE WEB CONFERENCE ACM
Research Track | Graph Algorithms and Modeling for the Web

MultiGPrompt for Multi-Task Pre-Training and Prompting on Graphs

Xingtong Yu, Chang Zhou, Yuan Fang, Xinming Zhang

University of Science and Technology of China
SMU Singapore Management University

Motivation

Problem

- To cater to diverse downstream tasks, pre-training should broadly extract knowledge from various aspects.

Challenges

- Different pretext tasks often have different objectives, directly combining them lead to task interference.
- Multiple pretext tasks further complicates the alignment of downstream objectives with the pre-trained model.

C1: How can we leverage diverse pre-text tasks for graph models in a synergistic manner?

C2: How can we transfer both task-specific and global pre-trained knowledge?

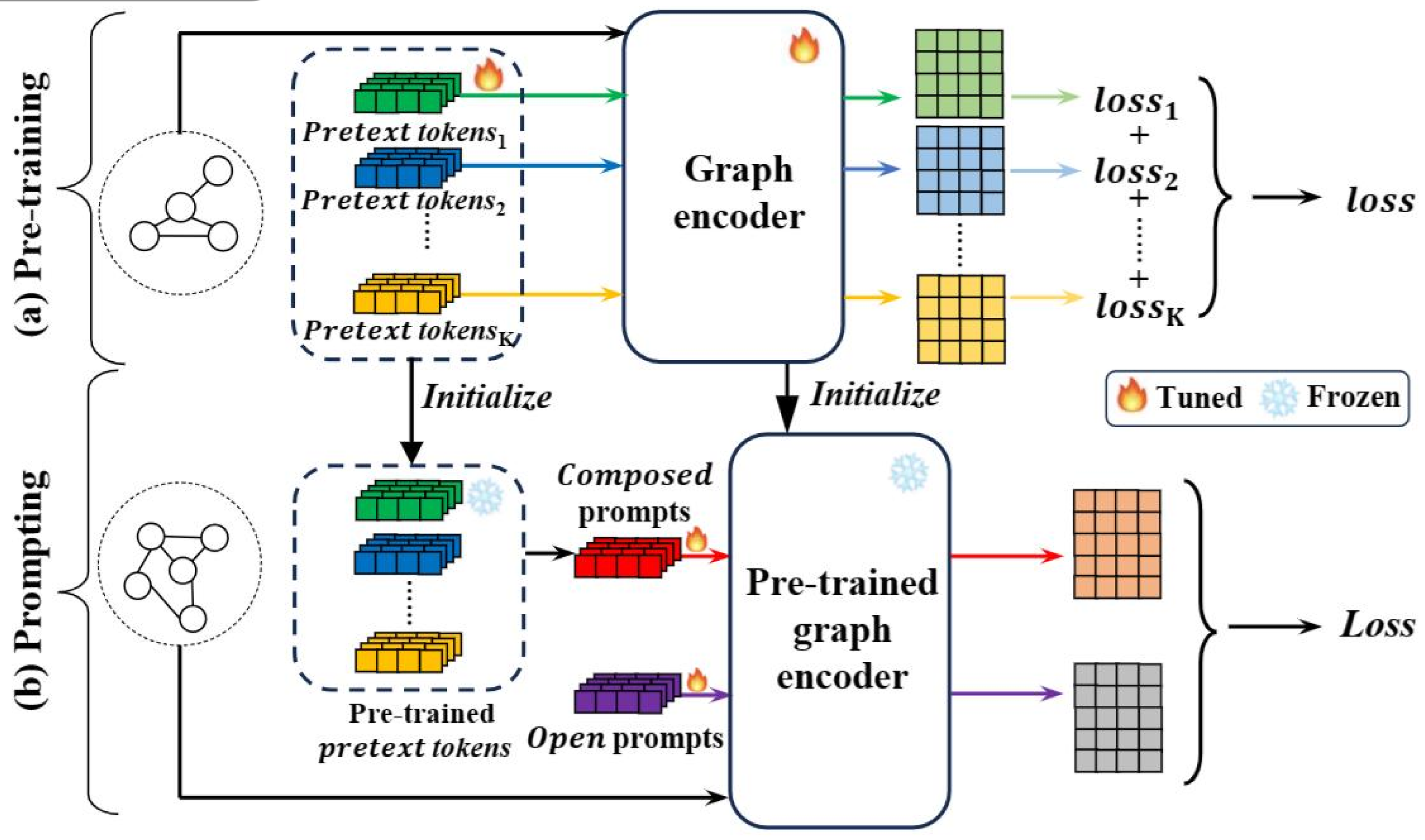

Figure 1: Illustration of MultiGPrompt. (a) Multi-task pre-training on graphs. (b) Prompting on downstream tasks.

Proposed Method: MultiGPrompt

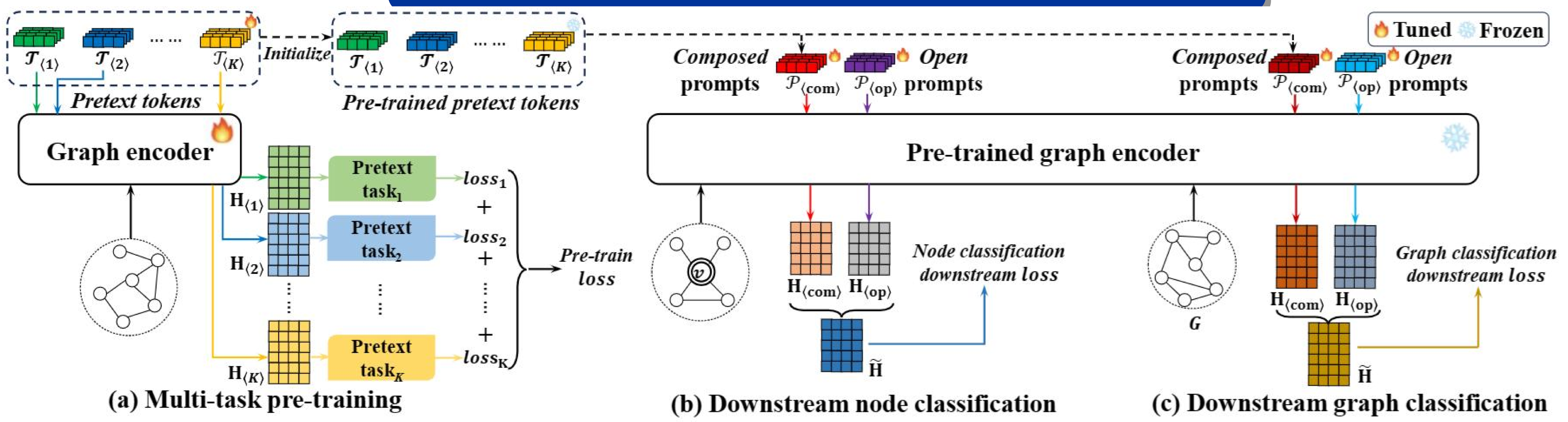

Figure 2: Overall framework of MultiGPrompt, consisting of two main stages: (a) Multi-task pre-training, and (b)/(c) Prompt-based learning for downstream few-shot tasks.

Multi-task pre-training

Pretext tokens

$$\mathcal{T}_{\langle k\rangle}=\{\mathbf{t}_{\langle k\rangle,0},\mathbf{t}_{\langle k\rangle,1},\ldots,\mathbf{t}_{\langle k\rangle,L}\}$$

Add token to each layer of graph encoder

$$\mathbf{H}^{l+1}=\mathrm{MP}(\mathbf{t}_{\langle k\rangle,l}\odot\mathbf{H}^{l},\mathbf{A};\theta^{l})$$

Graph encoder output embedding

$$\mathbf{H}_{\mathbf{t}}=\textsc{GraphEncoder}_{\mathbf{t}}(\mathbf{X},\mathbf{A};\Theta)$$

Overall embedding

$$\mathbf{H}_{\langle k\rangle}=\sum_{l=0}^{L}\alpha_l\mathbf{H}_{\mathbf{t}_{\langle k\rangle,l}}$$

Pre-Training Objective

$$\mathcal{L}_{\text{pre}}(\mathcal{H};\mathcal{T},\Theta)=\sum_{k=1}^{K}\beta_k\mathcal{L}_{\text{pre}_{\langle k\rangle}}(\mathbf{H}_{\langle k\rangle};\mathcal{T}_{\langle k\rangle},\Theta),$$

Prompt tuning

Composed prompt

$$\mathcal{P}_{\langle\text{com}\rangle}=\{\mathbf{p}_{\langle\text{com}\rangle,0},\mathbf{p}_{\langle\text{com}\rangle,1},\cdots,\mathbf{p}_{\langle\text{com}\rangle,L}\}$$

$$\mathbf{p}_{\langle\text{com}\rangle,l}=\textsc{Compose}(\mathbf{t}_{\langle 1\rangle,l},\mathbf{t}_{\langle 2\rangle,l},\ldots,\mathbf{t}_{\langle K\rangle,l};\Gamma)$$

Open prompt

$$\mathcal{P}_{\langle\text{op}\rangle}=\{\mathbf{p}_{\langle\text{op}\rangle,0},\mathbf{p}_{\langle\text{op}\rangle,1},\cdots,\mathbf{p}_{\langle\text{op}\rangle,L}\}$$

Aggregate dual prompt

$$\tilde{\mathbf{H}}=\textsc{Aggr}(\mathbf{H}_{\langle\text{com}\rangle},\mathbf{H}_{\langle\text{op}\rangle};\Delta)$$

Experiment

Table 1: Summary of datasets.

	Graphs	Graph classes	Avg. nodes	Avg. edges	Node features	Node classes	Task* (N/G)
Cora	1	-	2,708	5,429	1,433	7	N
Citeseer	1	-	3,327	4,732	3,703	6	N
PROTEINS	1,113	2	39.06	72.82	1	3	N, G
ENZYMES	600	6	32.63	62.14	18	3	N, G
BZR	405	2	35.75	38.36	3	-	G
COX2	467	2	41.22	43.45	3	-	G

* indicates the type(s) of downstream task associated with each dataset: "N" for node classification and "G" for graph classification.

Table 2: Accuracy evaluation on few-shot node and graph classification.

Methods	Node classification: Cora	Node classification: Citeseer	Node classification: PROTEINS	Node classification: ENZYMES	Graph classification: BZR	Graph classification: COX2	Graph classification: PROTEINS	Graph classification: ENZYMES
GCN	28.57 ± 5.07	31.27 ± 4.53	43.31 ± 9.35	48.08 ± 4.71	<u>56.33</u> ± 10.40	50.95 ± 23.48	50.56 ± 3.01	17.10 ± 3.53
GAT	28.40 ± 6.25	30.76 ± 5.40	31.79 ± 20.11	35.32 ± 18.72	50.69 ± 23.66	50.58 ± 26.16	50.59 ± 12.43	16.80 ± 2.97
DGI/InfoGraph	54.11 ± 9.60	45.00 ± 9.19	45.22 ± 11.09	48.05 ± 14.83	52.57 ± 18.14	54.62 ± 15.36	48.21 ± 12.35	21.69 ± 5.98
GraphCL	51.96 ± 9.43	43.12 ± 9.61	46.15 ± 10.94	48.88 ± 15.98	54.11 ± 16.63	54.29 ± 17.31	53.69 ± 11.92	21.57 ± 5.20
GPPT	15.37 ± 4.51	21.45 ± 3.45	35.15 ± 11.40	35.37 ± 9.37	-	-	-	-
GraphPrompt	<u>54.25</u> ± 9.38	<u>45.34</u> ± 10.53	<u>47.22</u> ± 11.05	<u>53.54</u> ± 15.46	54.60 ± 10.53	<u>54.35</u> ± 14.78	<u>54.73</u> ± 8.87	<u>25.06</u> ± 7.56
MultiGPrompt	**57.72** ± 9.94	**54.74** ± 11.57	**48.09** ± 11.49	**54.47** ± 15.36	**60.07** ± 12.48	**56.17** ± 12.84	**56.02** ± 8.27	**26.63** ± 6.22

Results are reported in percent. The best method is bolded and the runner-up is underlined.

Table 3: Ablation study on prompt design for multi-task pre-training.

Methods	Pretext token	Composed prompt	Open prompt	Node classification: Cora	Node classification: Citeseer	Node classification: PROTEINS	Node classification: ENZYMES	Graph classification: BZR	Graph classification: COX2	Graph classification: PROTEINS	Graph classification: ENZYMES
Variant 1	×	×	×	56.58	50.69	46.48	48.04	49.63	54.35	55.72	21.07
Variant 2	×	×	✓	56.54	53.08	47.79	51.09	47.56	54.89	55.61	24.23
Variant 3	✓	×	×	45.00	52.36	45.11	50.55	57.14	54.43	55.67	21.06
Variant 4	✓	×	✓	56.59	50.63	47.64	50.52	57.52	55.21	55.12	24.30
Variant 5	✓	✓	×	56.83	53.72	47.50	53.11	55.71	53.04	55.15	23.33
MultiGPrompt	✓	✓	✓	**57.72**	**54.74**	**48.09**	**54.47**	**60.07**	**56.17**	**56.02**	**26.63**

Results are evaluated using classification accuracy, reported in percent. The best variant is bolded.

Baselines

- Supervised: GCN, GAT
- Pre-train,Finetune: DGI/InfoGraph, GraphCL
- Prompting: GPPT, GraphPrompt

Observation

- MultiGPrompt consistently outperforms all baselines and variations.

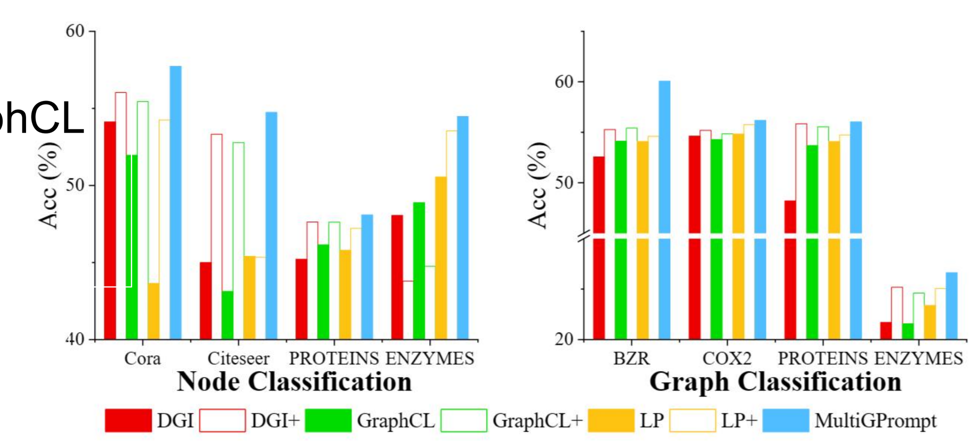

Figure 5: Ablation study on pretext tasks.