GraphBridge: Towards Arbitrary Transfer Learning in GNNs

Li Ju, Xingyi Yang, Qi Li, Xinchao Wang

{jujulili,xyang,liqi}@u.nus.edu, xinchao@nus.edu.sg

International Conference On Learning Representations

National University of Singapore

Motivation

GNNs are conventionally trained on a per-domain, per-task basis, which creates a significant barrier in GNN model reuse and arbitrary graph knowledge transfer. There are 2 main challenges towards arbitrary graph transfer learning:

Task heterogeneity (Multi Input & Output)

- Hard to fit various input dimensions and output forms of downstream tasks, considering that we only have a frozen pre-trained backbone without any additional dimension transformation.

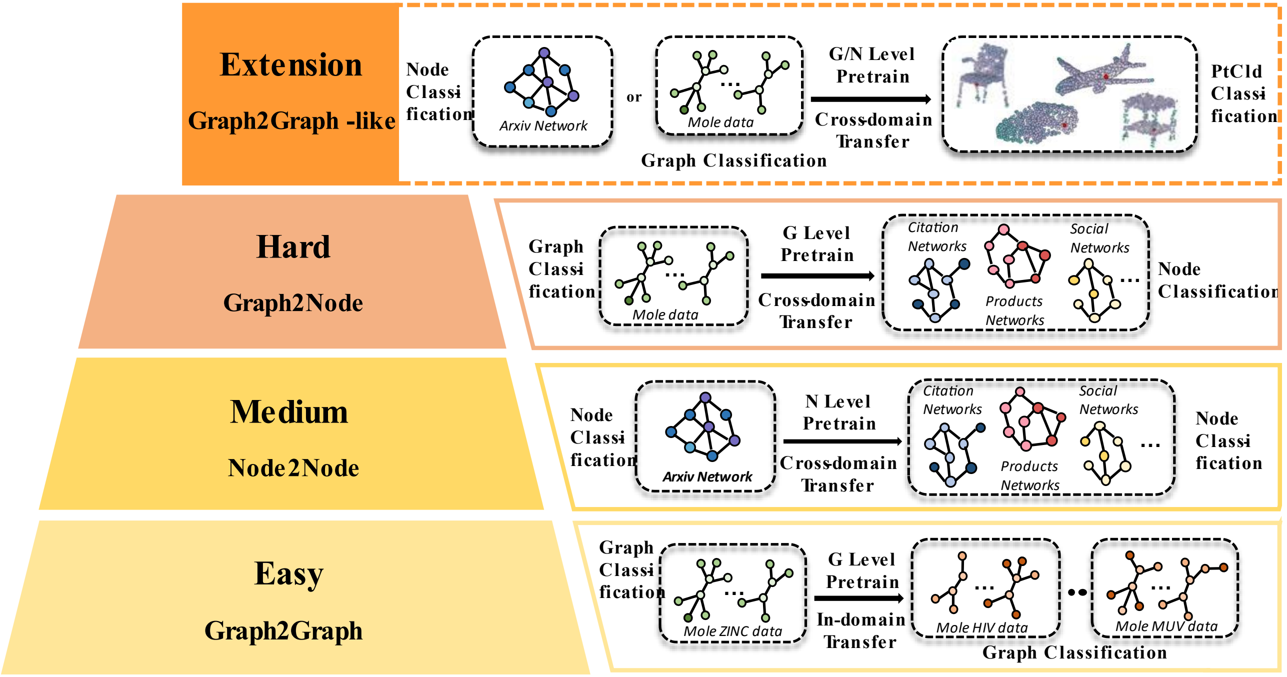
Domain heterogeneity

- Insufficient tuning methods' capacity to make good use of the knowledge from pre-trained models for adaptation in the target domain, since the substantial domain gap between upstream and downstream tasks exists.

Major Work

To enable knowledge transfer across disparate tasks and domains in GNNs, we:

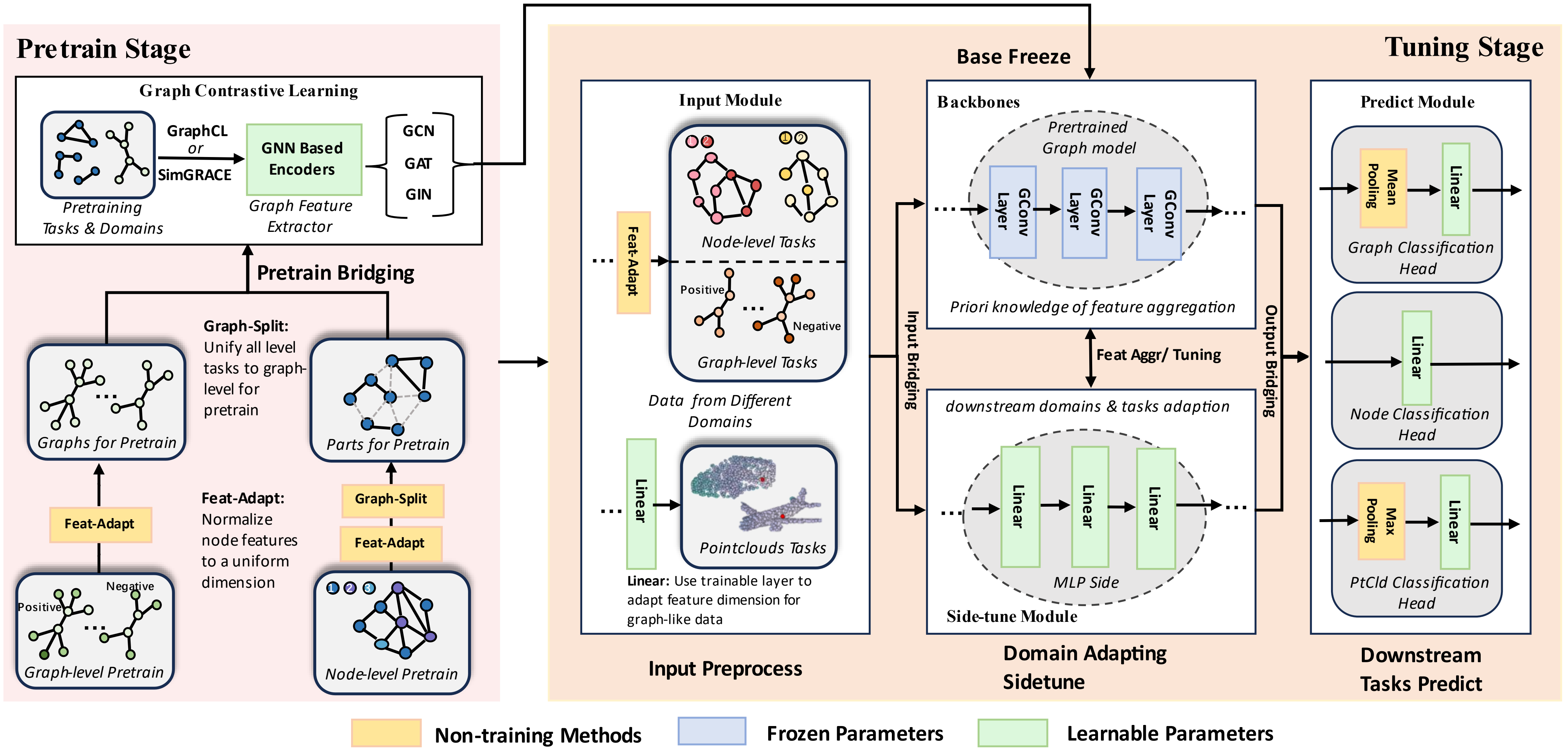
- Propose **GraphBridge** to enable knowledge transfer across disparate tasks and domains in GNNs without modifying task configurations or graph topologies;
- Design the **resource-efficient graph side-tuning method** to save temporary and memory resources in arbitrary graph transfer learning;



- Present the **Task Pyramid** for evaluating transfer tasks across different graph domains of varying difficulty.
- Validate that GraphBridge attains SOTA performance on different datasets across different task scenarios. with only **5% ~ 20%** of the tunable parameter.

Methods - Framework

GraphBridge Framework comprises a **Pre-training Stage**, aimed at extracting generalized graph knowledge, and a **Tuning Stage** dedicated to downstream tasks adaptation.

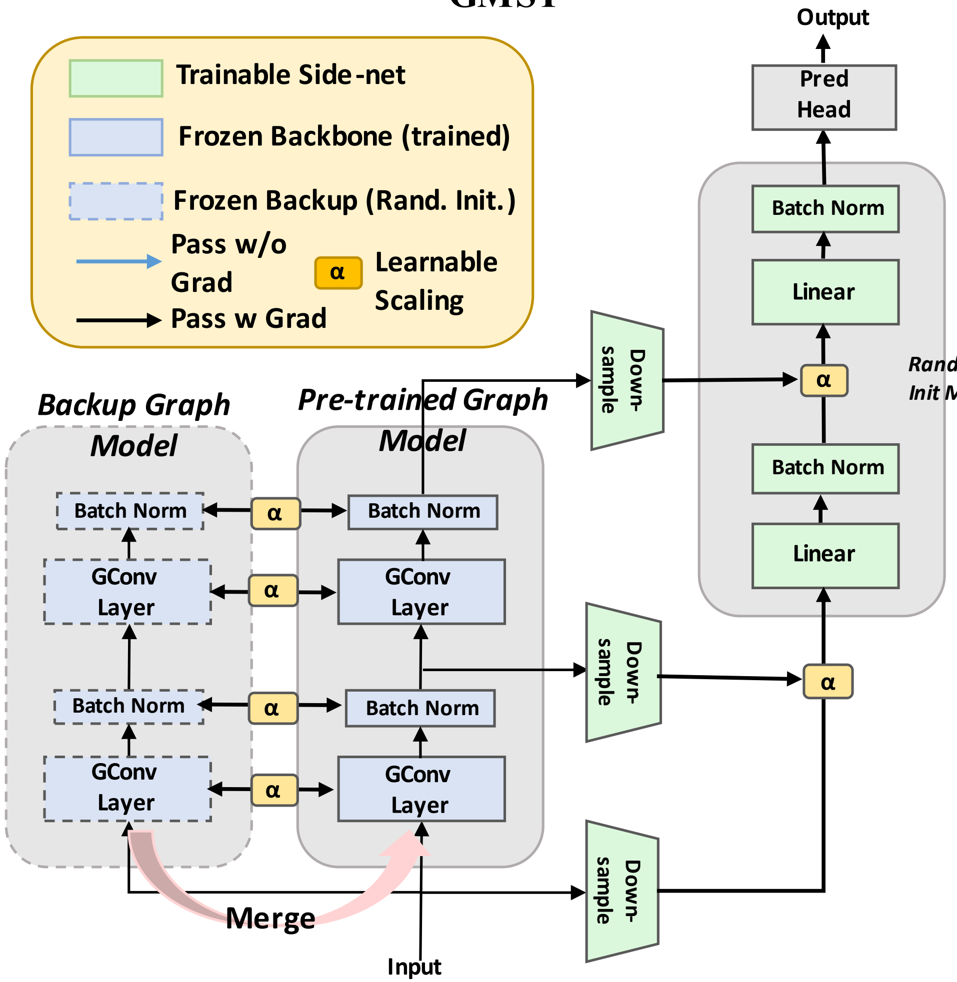


Methods - Side tuning

In the tuning stage, we introduce novel graph side tuning technique, enabling effective transfer learning of different graph tasks.

- On the one hand, **Side tuning** showcases resource efficiency by maintaining performance with fewer parameter manipulations.
- On the other hand, the flexible architecture of **Side tuning** facilitates the design of solutions to address negative transfers occurring during large gap domain transfer.

GMST



Experiments - Setups

Dataset

Names	Task Descriptions	Feature Dimensions	Nodes	Edges	# Graphs
1. Flickr [12]	Online Images Classification	500	89,250	899,756	1
2. Cora [6]	Machine-Learning Paper Classification	1,433	2,708	5,429	1
3. Citeseer [6]	Computer-Science Paper Classification	3,703	3,327	4,732	1
4. Pubmed [6]	Diabete-related Publication Classification	500	19,717	44,338	1
5. ogbn-arxiv [2, 8]	Subject Area Prediction of arXiv Papers	128	169,343	1,166,243	1
6. Amazon Computers [5]	Computer-Product Category Prediction	767	13,752	574,418	1
7. BACE [10]	Molecule Property Classification	2	$\sim$ 34.1	$\sim$ 73.7	1,513
8. BBBP [10]	Molecule Property Classification	2	$\sim$ 23.9	$\sim$ 51.6	2,039
9. ClinTox [10]	Molecule Property Classification	2	$\sim$ 26.1	$\sim$ 55.5	1,484
10. HIV [10]	Molecule Property Classification	2	$\sim$ 25.5	$\sim$ 54.9	41,127
11. SIDER [10]	Molecule Property Classification	2	$\sim$ 33.6	$\sim$ 70.7	1,427
12. Tox21 [10]	Molecule Property Classification	2	$\sim$ 18.6	$\sim$ 38.6	7,831
13. MUV [10]	Molecule Property Classification	2	$\sim$ 24.2	$\sim$ 52.6	93,087
14. ToxCast [10]	Molecule Property Classification	2	$\sim$ 18.7	$\sim$ 38.4	8,597
15. ZINC-full [10]	Molecule Property Classification	2	$\sim$ 23.2	$\sim$ 49.8	249,456
16. ModelNet10 [9]	3D Object Recognition	3	$\sim$ 9,508.2	$\sim$ 37,450.5	4,899

Pre-training Methods

- SOTA graph-level pre-training methods: GraphCL & SimGRACE.

Foundational Backbone of the Base Model

- Classic GNN Architectures: 2-layers GCN, GAT and GIN.

Experiments

Easy

Results of Graph2Graph Transfer. Test ROC-AUC (%) performances on molecular prediction benchmarks with different pre-train-tuning workflows. **Imp.** refers to the improvement of parameter-efficient tuning methods in comparison to the fine-tuning.

Pre-train Methods	Tuning Methods	BACE	BBBP	ClinTox	HIV	SIDER	Tox21	MUV	ToxCast	Avg.	Imp.
GraphCL[14]	FT[1]	74.6±2.2	68.6±2.3	69.8±2.2	78.5±1.2	59.6±0.7	74.4±0.5	73.7±2.7	62.9±0.4	70.3	–
	Adapter[4]	76.1±2.2	67.8±1.4	72.0±3.8	77.8±1.3	59.6±1.3	74.9±0.9	75.0±2.11	63.1±0.4	70.7	0.4%
	GBST	73.2±0.7	65.1±0.4	64.7±0.1	70.0±0.5	60.6±0.2	70.7±0.1	74.8±0.3	61.4±0.1	67.6	-2.7%
	GSST	79.3±0.2	69.5±1.0	71.1±0.4	72.8±0.9	60.6±0.1	72.1±0.1	78.0±0.7	62.9±0.1	**70.9**	0.6%
SimGRACE[11]	FT[1]	74.7±1.0	65.5±1.0	53.8±2.3	74.6±1.2	58.1±0.6	71.9±0.4	71.0±1.9	61.3±0.4	66.3	–
	Adapter[4]	74.9±1.7	64.6±1.3	53.9±7.0	72.3±1.2	57.2±0.9	71.4±0.6	71.8±1.4	61.3±0.6	65.9	-0.4%
	GBST	65.6±2.1	64.1±1.3	53.1±0.7	69.2±0.4	57.7±0.5	70.8±0.1	72.1±2.4	60.6±0.2	64.3	-2.0%
	GSST	73.0±0.6	65.4±0.2	57.2±0.3	69.1±0.1	57.9±0.2	72.3±0.3	74.4±0.5	61.6±0.1	**66.4**	0.1%

Experiments

Medium

Results of Node2Node Transfer. Test Acc. (%) on diverse node-classification benchmarks with different tuning methods and backbones under different data pre-training methods.

Pre-train Methods	Tuning Methods	Citeseer			PubMed			Cora			Amazon			Flickr		
		GCN	GAT	GIN	GCN	GAT	GIN	GCN	GAT	GIN	GCN	GAT	GIN	GCN	GAT	GIN
—	Scratch Train[3, 7]	64.30	69.21	55.10	75.70	75.10	65.80	76.90	77.00	72.10	92.37	92.33	91.89	53.07	52.97	53.15
GraphCL[14]	FT[1]	56.62	56.87	52.83	69.96	70.21	67.32	**74.43**	**73.32**	62.40	**92.22**	**92.00**	**91.02**	**53.32**	**52.85**	**53.90**
	GBST	56.10	53.52	54.37	70.94	68.83	68.60	60.41	59.57	57.08	88.84	87.79	85.28	49.68	45.69	49.69
	GSST	54.00	55.82	56.44	69.88	71.83	69.07	63.30	64.00	59.13	88.95	84.77	85.13	49.72	44.32	49.54
	GAST	59.70	61.30	54.41	71.10	71.13	69.79	69.60	70.30	57.30	89.24	88.80	86.59	49.71	4716	49.53
	GMST	**59.30**	**63.40**	**58.80**	**72.10**	**75.00**	**72.60**	73.10	72.32	**65.40**	89.42	90.19	86.15	51.92	47.70	49.94
SimGRACE[11]	FT[1]	58.90	57.60	45.50	71.30	71.70	64.10	72.9	71.20	64.40	**92.37**	**92.29**	**91.28**	**53.60**	**50.81**	**53.77**
	GBST	53.00	48.80	47.20	69.70	69.40	64.80	62.30	59.60	51.90	88.91	88.62	85.71	49.11	45.82	49.67
	GSST	52.00	52.10	49.50	68.00	70.00	67.30	64.60	59.30	53.90	88.84	87.86	80.26	48.93	45.20	49.71
	GAST	54.30	51.30	47.80	69.90	71.80	63.80	63.60	63.60	51.20	88.95	89.42	85.53	49.16	46.71	49.67
	GMST	**61.60**	**63.40**	**58.90**	**73.20**	**75.80**	**72.70**	**75.10**	**72.20**	**66.70**	90.88	90.53	84.19	50.56	47.71	51.16

Experiments

Hard

Results of Graph2Node Transfer. Test Acc. (%) on diverse node-classification benchmarks with different tuning methods and backbones under different data pre-training methods.

Pre-train Methods	Tuning Methods	Citeseer GCN	Citeseer GAT	Citeseer GIN	PubMed GCN	PubMed GAT	PubMed GIN	Cora GCN	Cora GAT	Cora GIN	Amazon GCN	Amazon GAT	Amazon GIN	Flickr GCN	Flickr GAT	Flickr GIN
—	Scratch Train[3, 7]	64.30	69.21	55.10	75.70	75.10	65.80	76.90	77.00	72.10	92.37	92.33	91.89	53.07	52.97	53.15
GraphCL[14]	FT[1]	52.60	49.90	46.90	68.60	68.00	63.50	69.20	60.90	63.10	**91.82**	89.09	**90.44**	**52.3**	**49.87**	**52.38**
	GBST	49.00	48.40	50.90	64.40	62.40	67.10	51.60	50.00	55.00	87.02	85.61	84.81	48.18	45.67	47.98
	GSST	48.70	49.90	50.20	64.60	64.40	65.80	52.20	51.30	56.00	83.68	80.77	80.23	48.22	44.49	47.55
	GAST	48.40	50.50	50.10	68.80	64.90	68.80	58.90	52.10	55.50	88.44	87.50	84.70	50.31	46.12	47.37
	GMST	**61.90**	**62.40**	**57.90**	**73.10**	**73.70**	**73.90**	**74.80**	**72.30**	**66.50**	88.62	**89.79**	85.50	51.30	47.34	49.54
SimGRACE[11]	FT[1]	53.70	53.00	43.30	59.30	68.10	61.80	65.00	64.30	57.70	**92.26**	89.46	**90.88**	**52.42**	**48.05**	**53.50**
	GBST	48.50	46.00	50.20	60.70	61.30	66.40	53.30	49.90	49.60	85.13	86.22	83.93	47.97	45.95	47.18
	GSST	49.90	45.40	51.90	58.20	61.60	65.30	54.50	49.20	48.30	80.95	81.68	77.39	47.80	44.86	48.93
	GAST	48.20	45.10	48.00	61.60	65.80	66.20	56.30	52.70	48.70	89.31	89.06	85.42	48.72	46.66	48.99
	GMST	**63.00**	**62.20**	**58.50**	**72.70**	**74.80**	**73.30**	**74.30**	**71.10**	**65.20**	89.67	**89.46**	85.26	50.24	47.29	51.57

Experiments

Extension

Results of Graph2PtCld Transfer. I configure pre-trained models on graph-level and node-level data across distinct types of graph models as the initialization for backbone.

Pre-train Methods	Tuning Methods	Graph-level: **ogbg-molhiv**			Node-level: **ogbn-arxiv**		
Backbones		GCN	GAT	GIN	GCN	GAT	GIN
—	Scratch Train[3, 7]	77.4	81.6	87.9	77.4	81.6	87.9
GraphCL[14]	FT[1]	78.1	74.9	83.9	73.4	**75.4**	**87.3**
	GBST	76.6	**79.2**	83	73.8	74.6	86.0
	GAST	79.3	78.2	82.9	**76.4**	72.1	83.8
	GSST	**81.7**	78.6	84.7	70.4	74.0	80.4
	GMST	77.8	74.8	82.5	75.6	74.5	80.8
SimGRACE[11]	FT[1]	78.6	70.9	77.3	72.8	78.0	**79.7**
	GBST	75.7	77.43	75.4	72.1	76.8	79.3
	GAST	76.9	74.3	74.1	73.3	77.5	79.3
	GSST	**79.3**	**81.8**	77.3	71.8	69.2	63.9
	GMST	65.2	65.1	76.4	**78.1**	**78.1**	78.7

Experiments

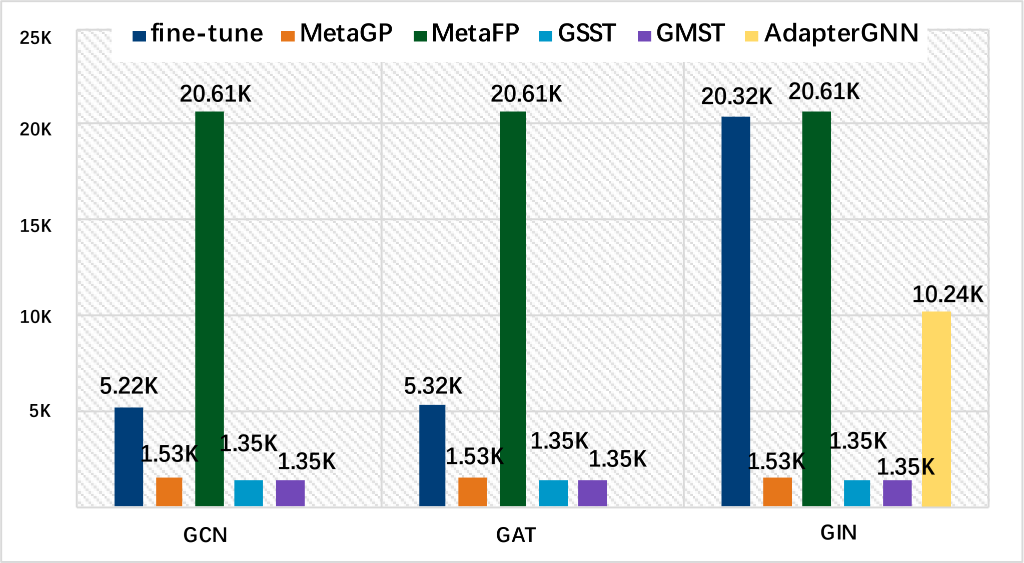
Efficacy & Efficiency

GSST	GMST	Dataset: **Cora**					
		Node-level Pretrain			Graph-level Pretrain		
		GCN	GAT	GIN	GCN	GAT	GIN
✗	✗	13.80	11.60	17.30	13.50	31.90	17.00
✓	✗	63.30	64.00	59.13	52.20	51.30	56.00
✗	✓	**73.10**	**72.32**	**65.40**	**74.80**	**72.30**	**66.50**

Ablation studies on the effects of GSST and GMST modules designed for arbitrary graph transfer learning. we conduct tests under three training scenarios: direct inference without tuning, tuning with GSST, and tuning with GMST.

Speed-up. ↑ **(%)**	GCN	GAT	GIN
FT	3.3	7.2	1.7
MetaFP	77.3	68.6	74.3
Adapter	-	-	20.2
GSST	39.4	57.6	26.3
GSMT	31.6	52.3	20.8

Training speed-up of different tuning methods compared to Scratch Training. We selected the transfer experiment on Cora dataset in the challenging scenario as a representative.



Adjustable parameter sizes in different tuning methods across distinct backbones. We conduct statistics on 5-layer backbones.

Conclusion

NUS National University of Singapore

- We introduce a novel GraphBridge framework for **resource-efficient** graph transfer learning toward **arbitrary** downstream tasks and domains.
- We will strive to tackle transfer tasks across more **benchmarks**, **backbones** and **pre-training methods** in the future.

Thank you!

Code: https://github.com/jujulili888/GraphBridge

International Conference On Learning Representations

National University of Singapore