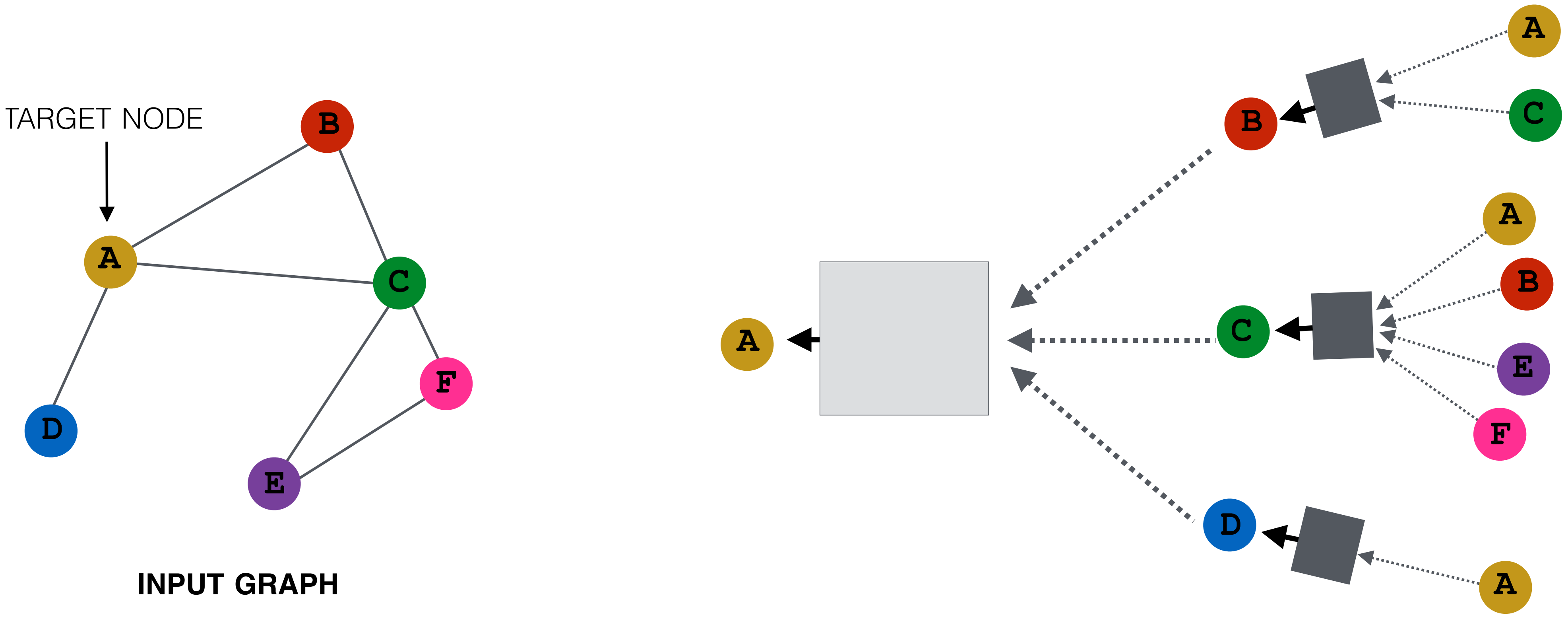
Neural Graph Matching for Pre-training Graph Neural Networks

Yupeng Hou[1], Binbin Hu[2], Wayne Xin Zhao[1,✉], Zhiqiang Zhang[2], Jun Zhou[2], Ji-Rong Wen[1]

1. Gaoling School of Artificial Intelligence, Renmin University of China
2. Ant Group

Background – GNN

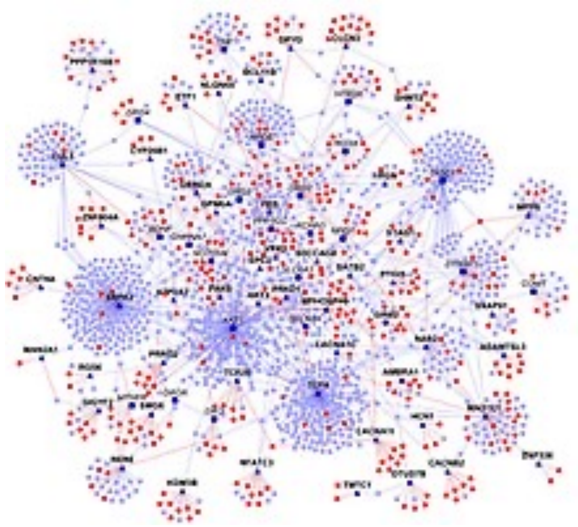
- Graph Neural Networks (GNNs)

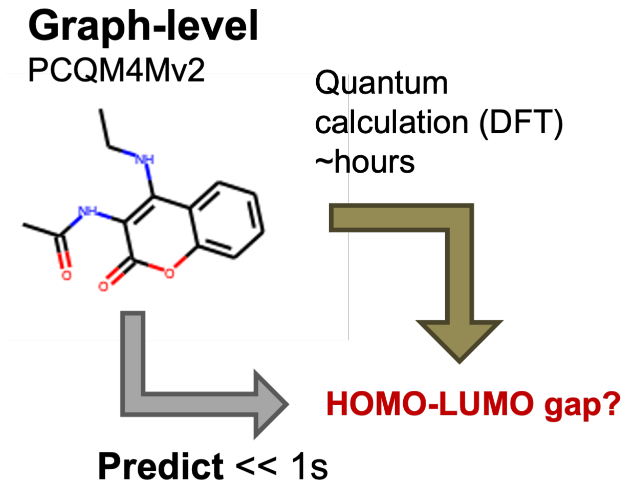


William L. Hamilton et al. Representation Learning on Graphs: Methods and Applications. IEEE Data Engineering Bulletin 2017.

Background – GNN

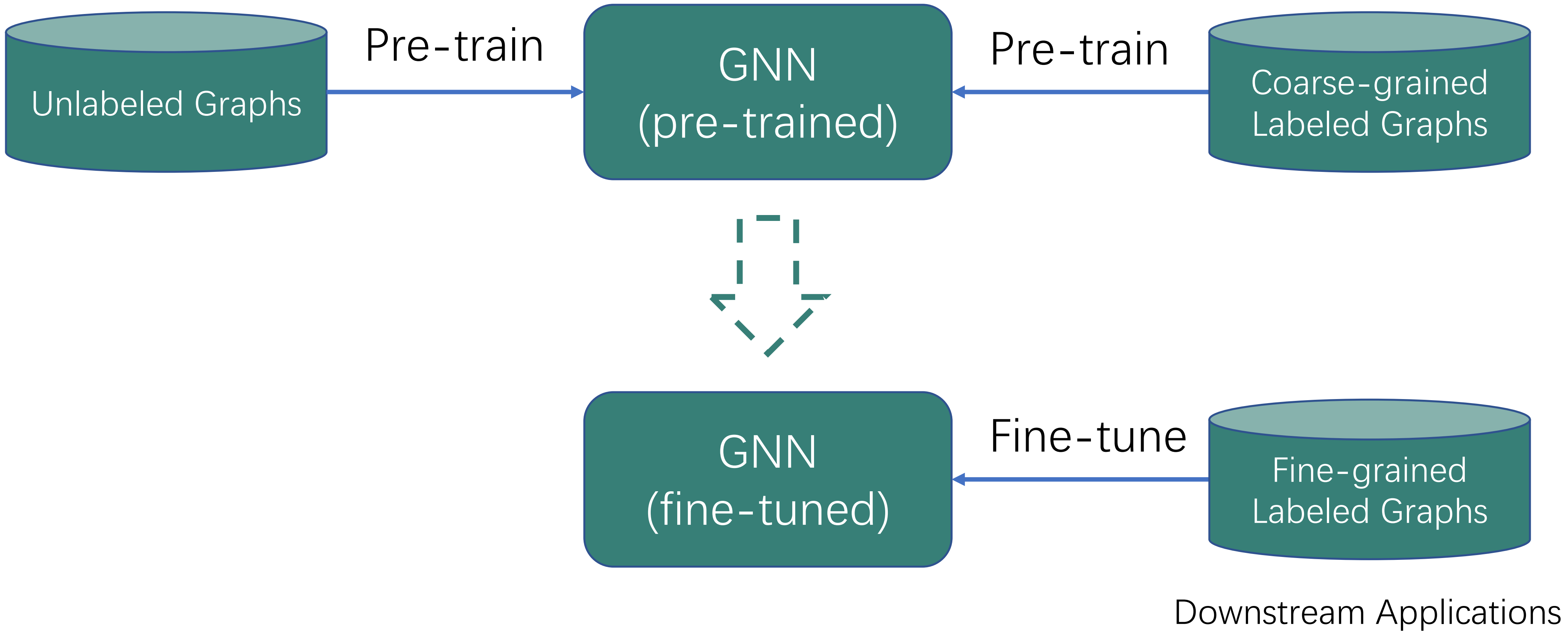
- Data for Graph Neural Networks (GNNs)
  - Requiring abundant task-specific labeled graph data
  - Sparse (especially for scientific domains)





Ganapathiraju MK et al. Schizophrenia interactome with 504 novel protein-protein interactions. NPJ Schizophrenia 2016.
Weihua Hu et al. OGB-LSC: A Large-Scale Challenge for Machine Learning on Graphs. NeurIPS 2021.

GNN Pre-training



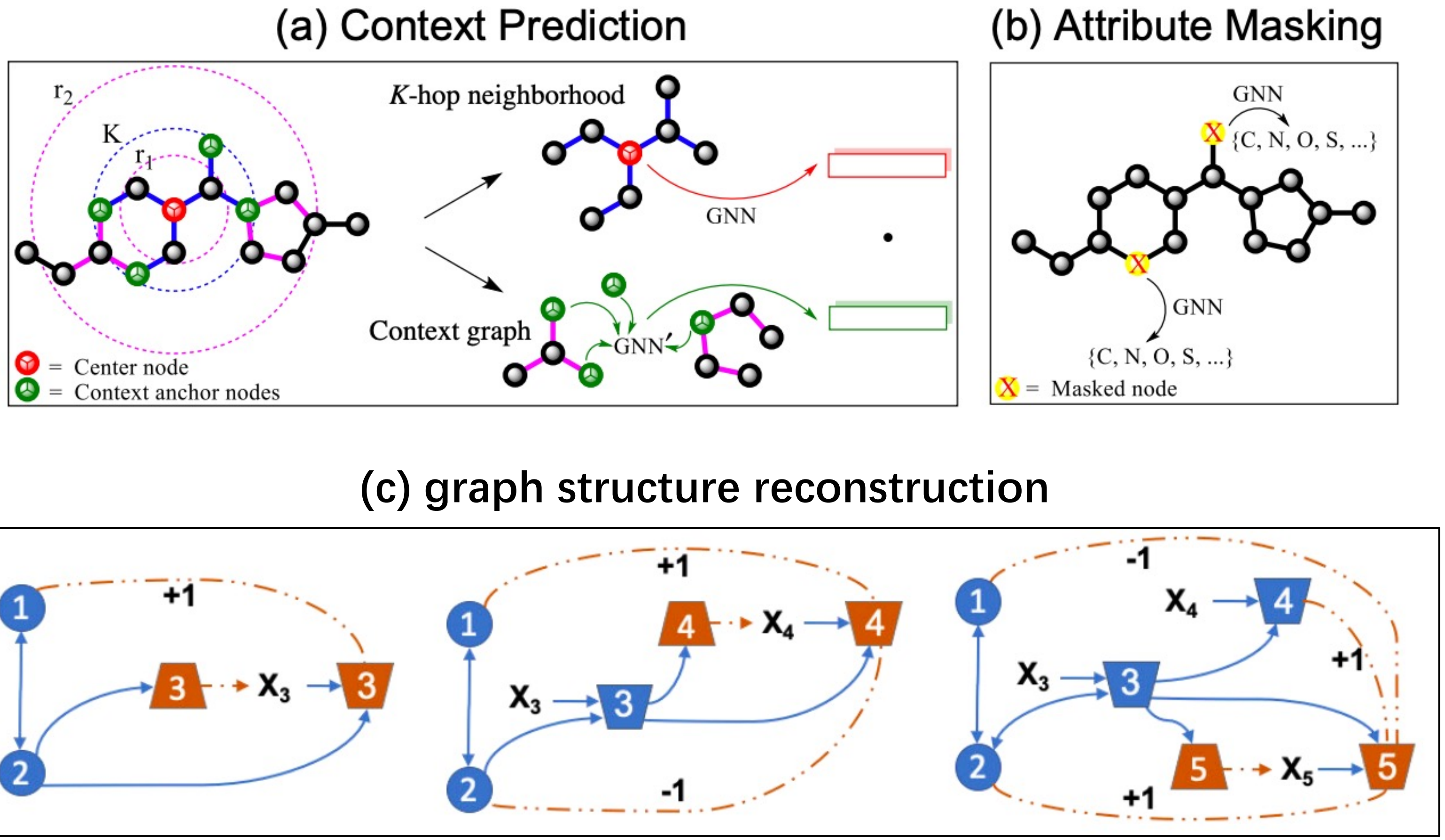
GNN Pre-training

GNN pre-training strategies

(1) node-level tasks
(2) graph-level tasks

Node-level tasks

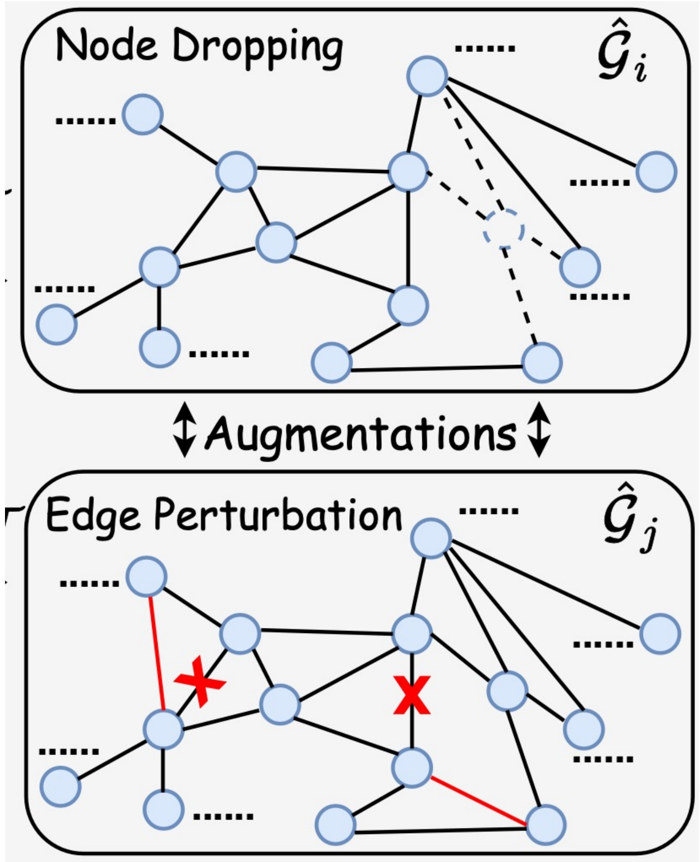
GNN Pre-training

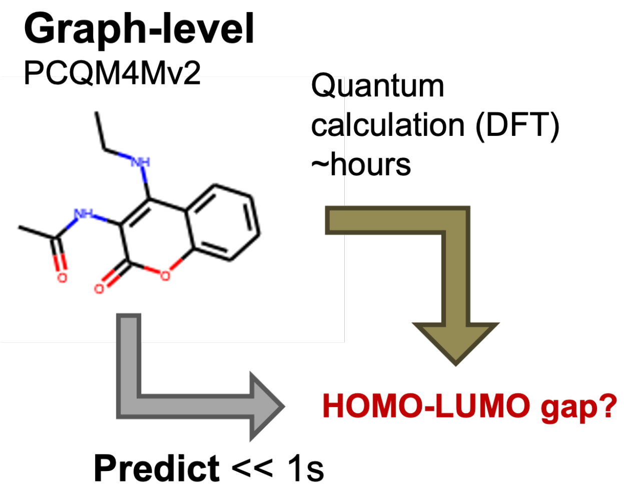


Weihua Hu et al. Strategies for Pre-training Graph Neural Networks. ICLR 2020.
Ziniu Hu et al. GPT-GNN: Generative Pre-Training of Graph Neural Networks. KDD 2020.

Graph-level tasks

GNN Pre-training

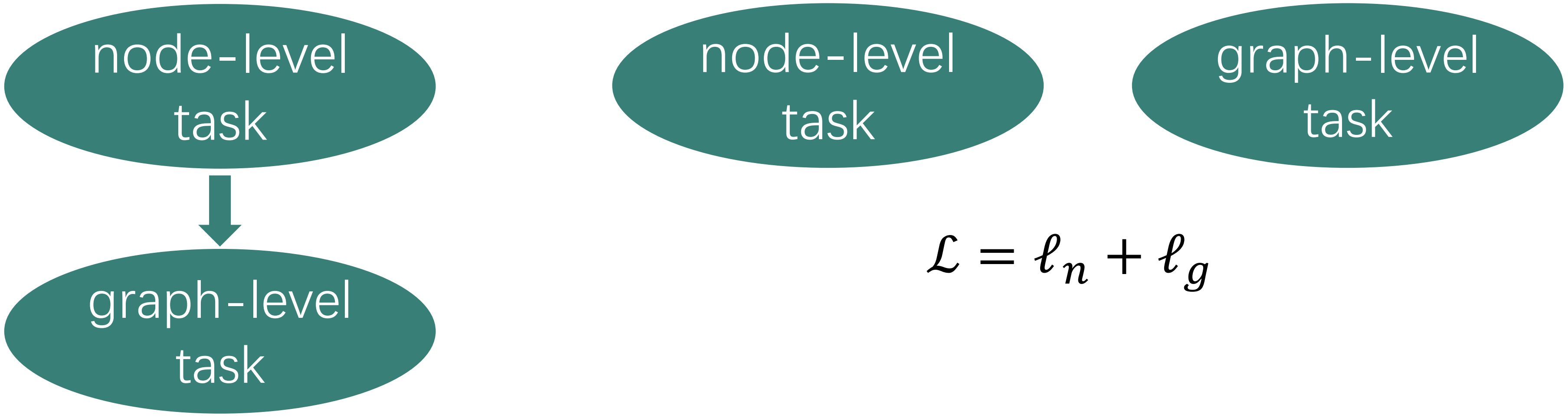




Yuning You et al. Graph Contrastive Learning with Augmentations. NeurIPS 2020.
Weihua Hu et al. OGB-LSC: A Large-Scale Challenge for Machine Learning on Graphs. NeurIPS 2021.

Challenge

- Combining both node- and graph-level optimization goals
  - essential for GNN pre-training



(a) two-stage **(b) multi-task**

- Each individual task is not aware of all the optimization goals at different levels

Challenge

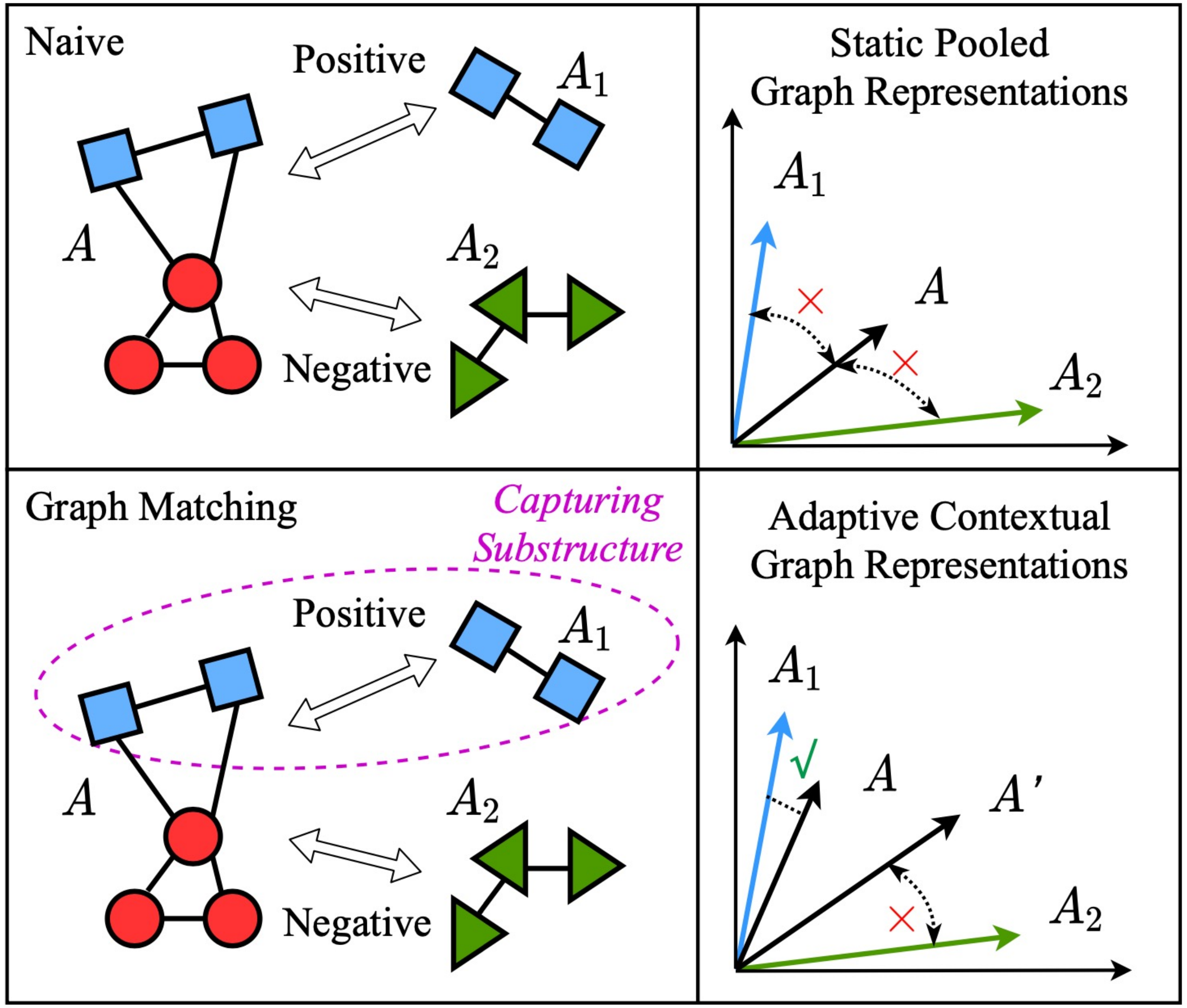
- Combining both node- and graph-level optimization goals
  - essential for GNN pre-training

(c) one single task? 🤔

- Can we have one single GNN pre-training task,
- capturing node- and graph-level characteristics simultaneously

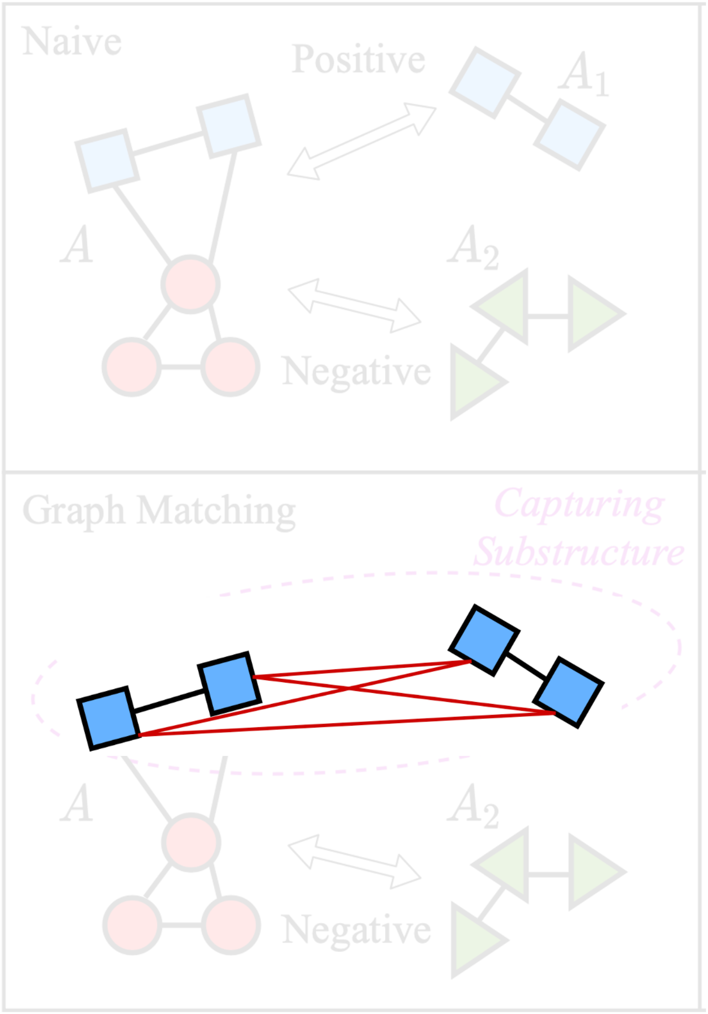
Idea

- Neural Graph Matching Task



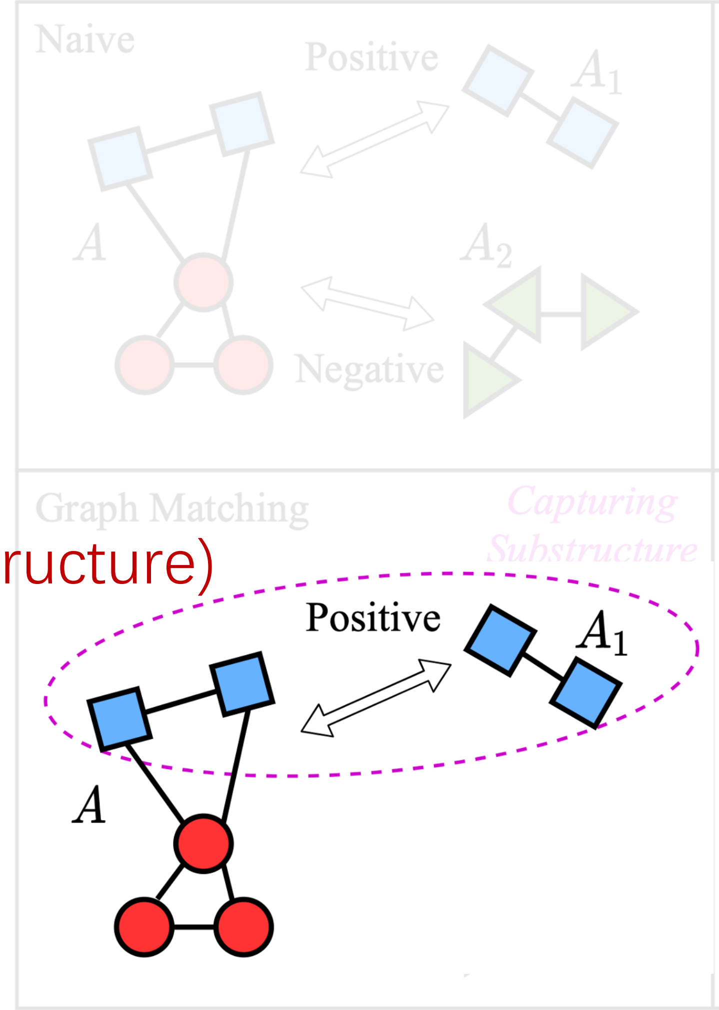
Idea

- Neural Graph Matching Task
  - Node-level correspondence (structural matching)



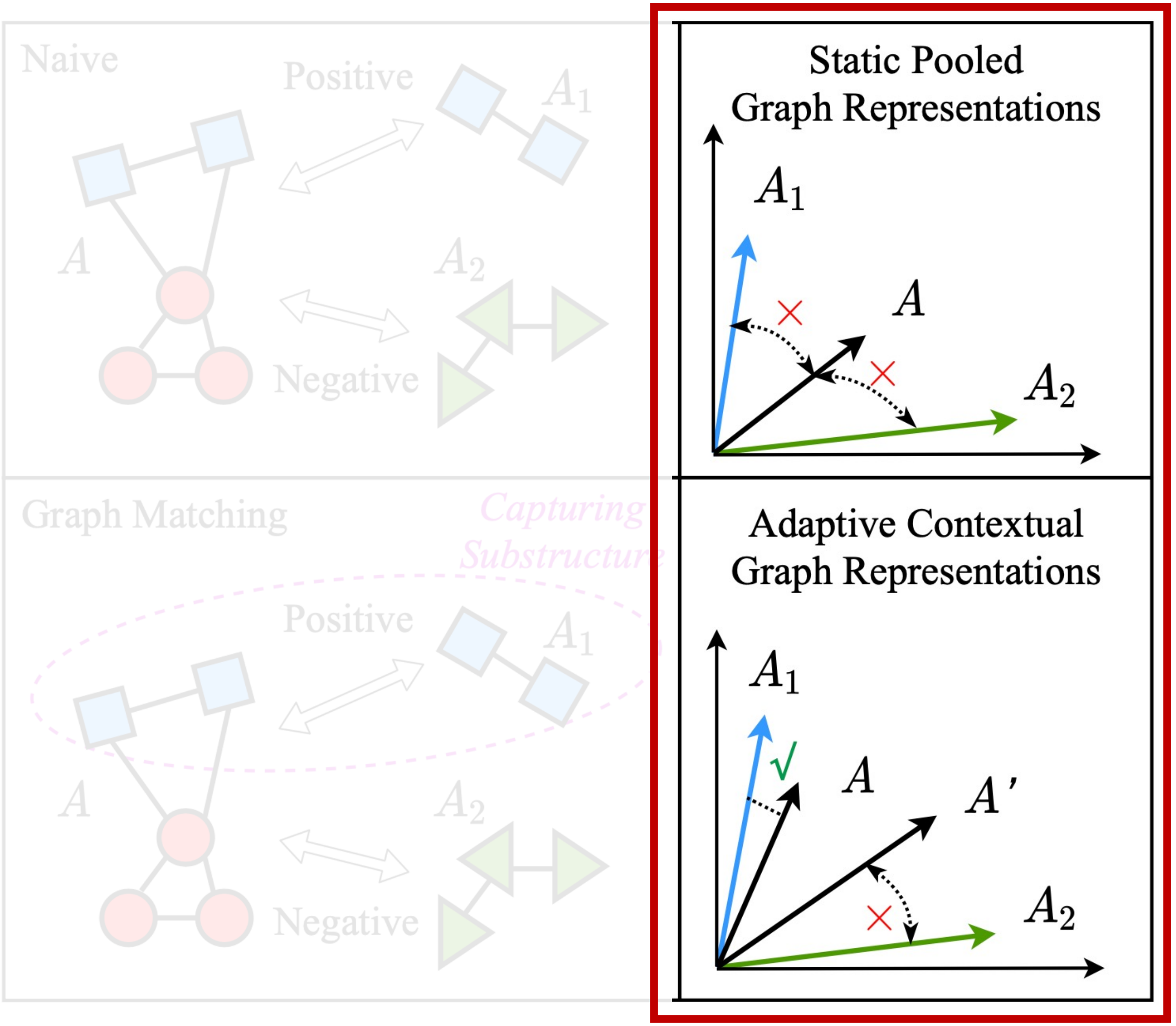
Idea

- Neural Graph Matching Task
  - Node-level correspondence
    (structural matching)
  - Graph-level properties
    (whether containing shared substructure)



Idea

- Neural Graph Matching Task
  - Adaptive graph representations (while pre-training)

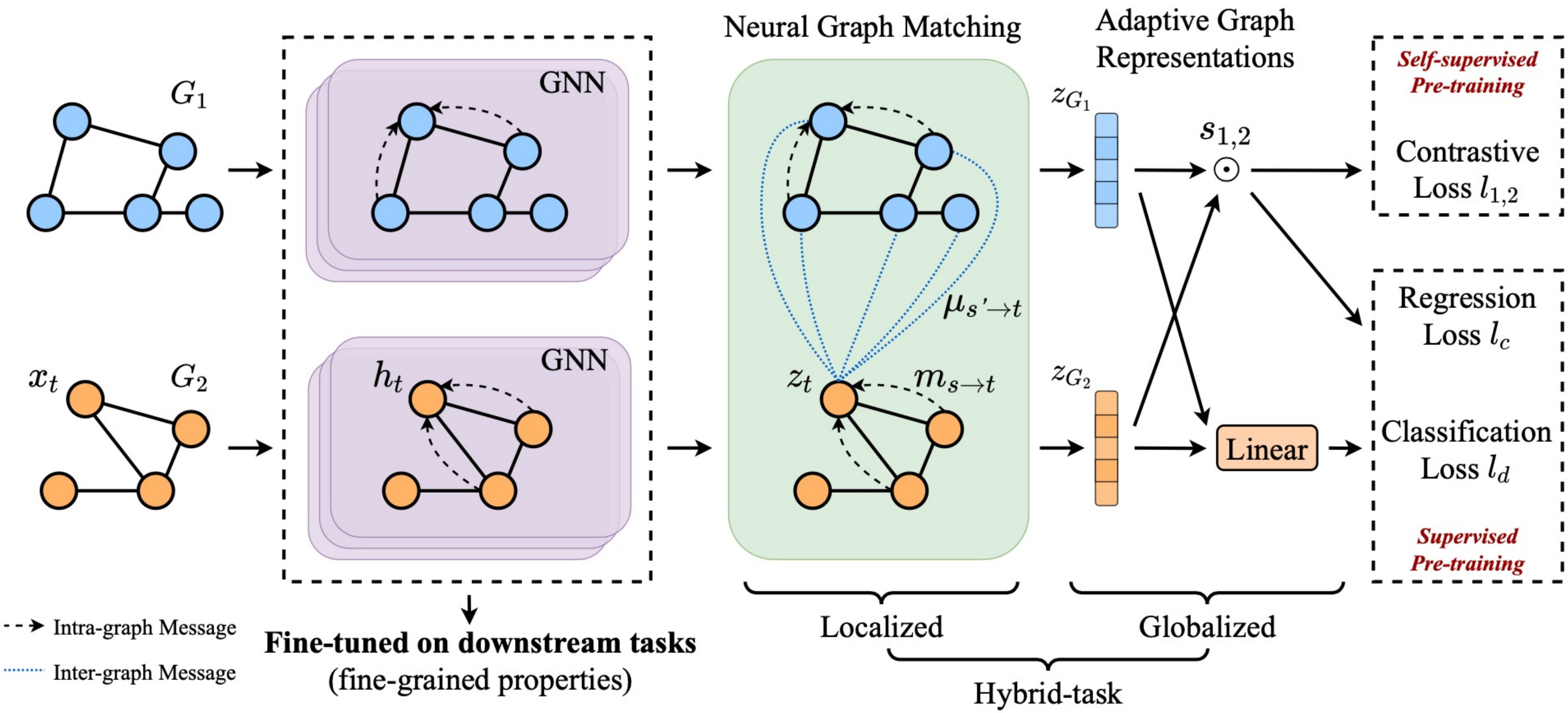


Methodology, GMPT

Graph Matching based GNN Pre-Training, GMPT

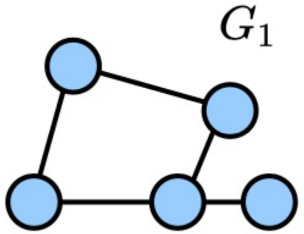
Self-supervised Pre-training

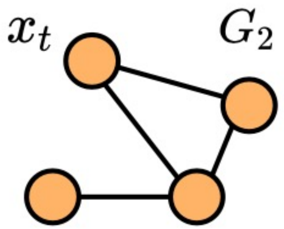
GMPT-CL



Self-supervised Pre-training

GMPT-CL



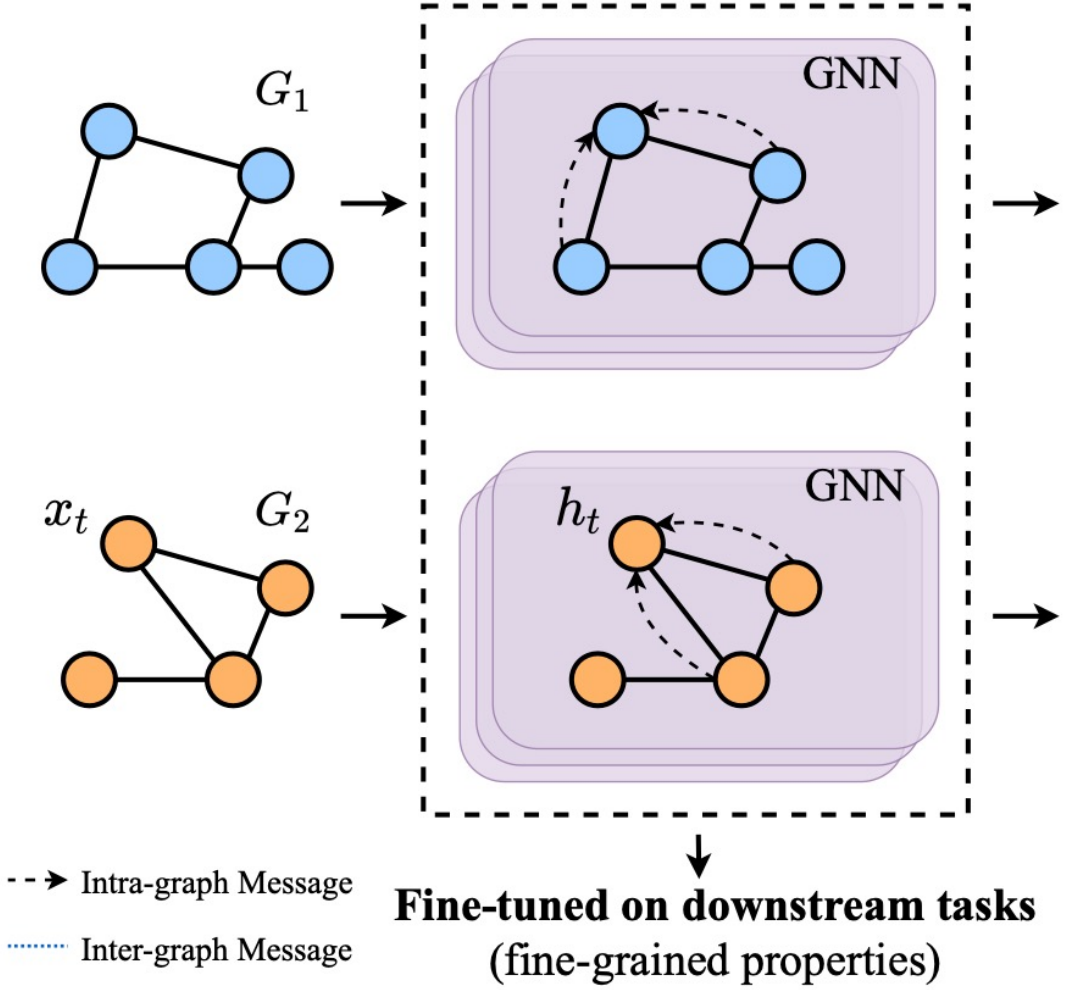


Graph Augmentation

Node/edge perturbation
Subgraph sampling
... ...

Self-supervised Pre-training

GMPT-CL



GNN Encoding

Self-supervised Pre-training

GMPT-CL

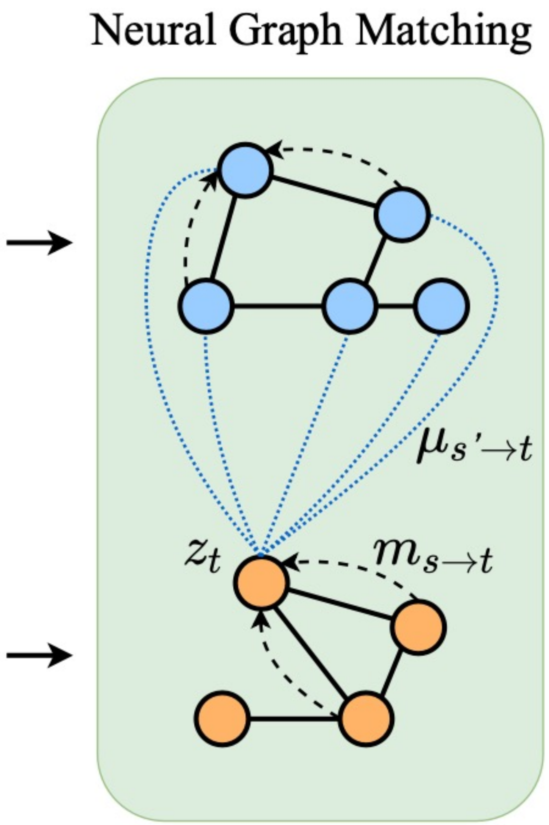
Graph Matching Module

$$\boldsymbol{m}_{s\to t} = \mathrm{MSG}_{\mathrm{intra}}\left(\boldsymbol{h}_s^{(1)}, \boldsymbol{h}_t^{(1)}, \boldsymbol{e}_{st}\right),$$

$$\boldsymbol{\mu}_{s'\to t'} = \mathrm{MSG}_{\mathrm{inter}}\left(\boldsymbol{h}_{s'}^{(1)}, \boldsymbol{h}_{t'}^{(2)}\right),$$

$$\boldsymbol{\mu}_{s'\to t'} = a_{s'\to t'} \cdot \boldsymbol{h}_{s'}^{(1)},$$

$$a_{s'\to t'} = \frac{\exp(\mathrm{sim}(\boldsymbol{h}_{s'}^{(1)}, \boldsymbol{h}_{t'}^{(2)}))}{\sum_{k\in\tilde{G}_2}\exp(\mathrm{sim}(\boldsymbol{h}_{s'}^{(1)}, \boldsymbol{h}_k^{(2)}))}$$



Self-supervised Pre-training

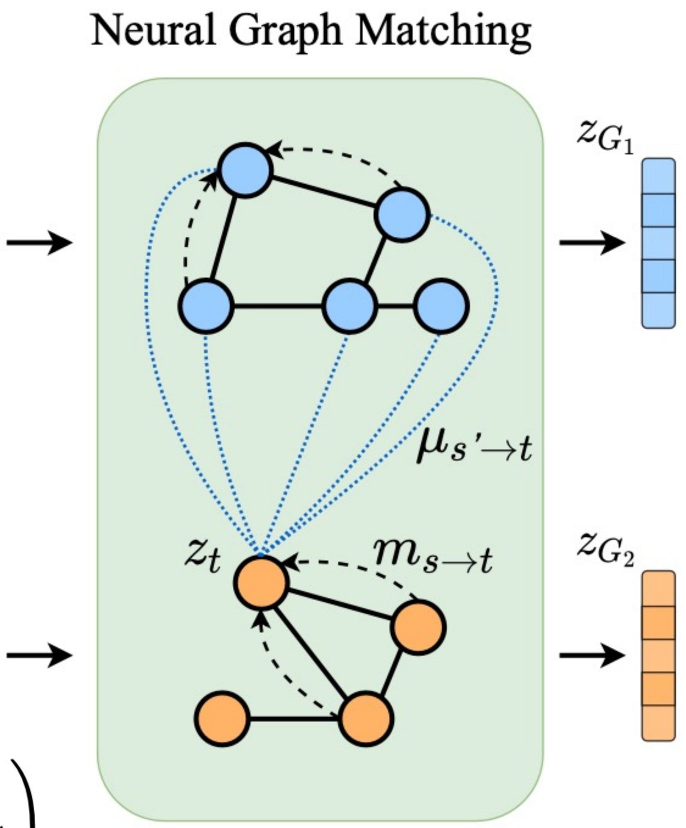
GMPT-CL

Adaptive Graph Representations

$$\boldsymbol{m}_{s \to t} = \mathrm{MSG}_{\mathrm{intra}}\left(\boldsymbol{h}_s^{(1)}, \boldsymbol{h}_t^{(1)}, \boldsymbol{e}_{st}\right),$$

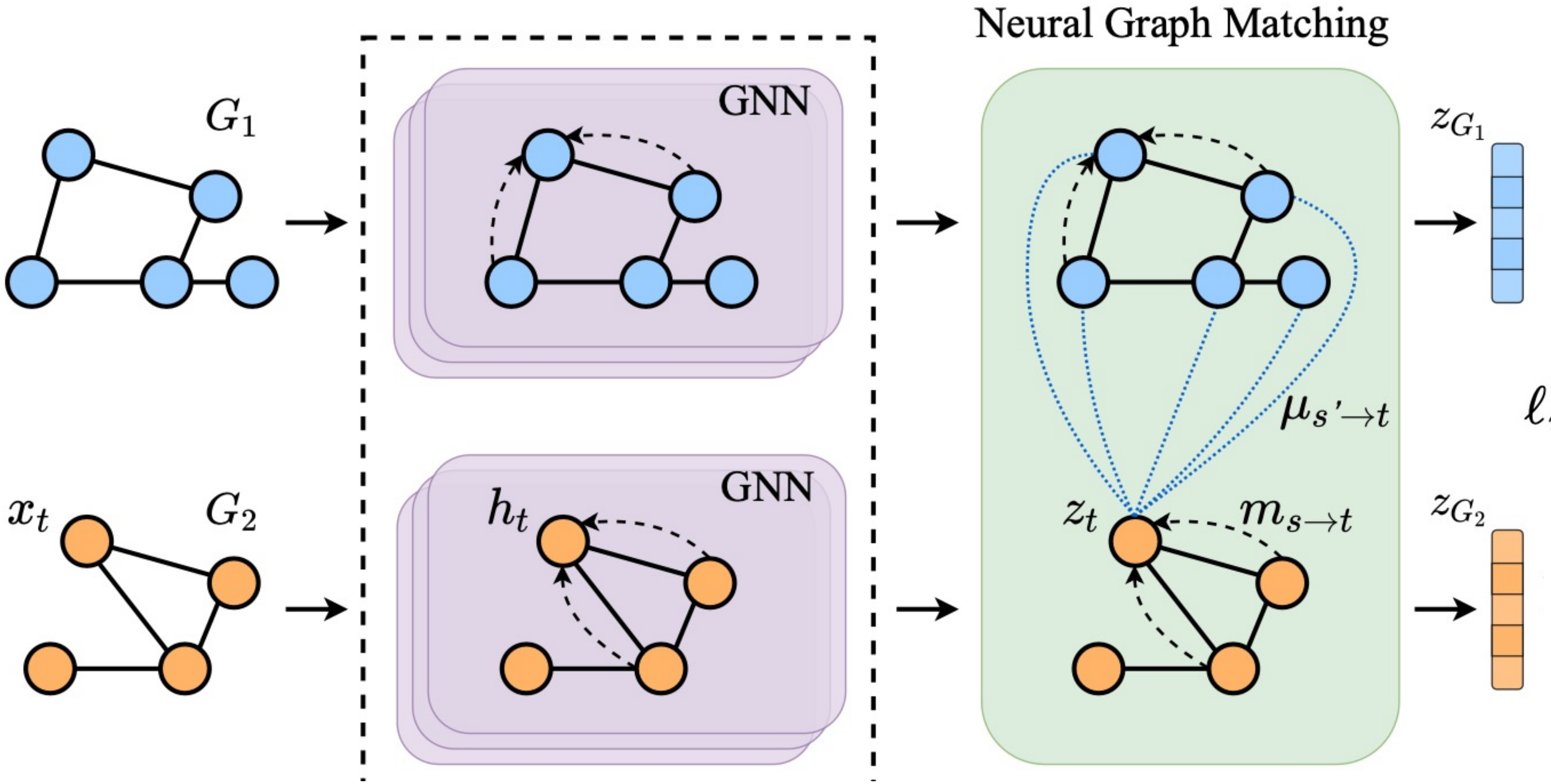
$$\boldsymbol{\mu}_{s' \to t'} = \mathrm{MSG}_{\mathrm{inter}}\left(\boldsymbol{h}_{s'}^{(1)}, \boldsymbol{h}_{t'}^{(2)}\right),$$

$$\boldsymbol{z}_t = \mathrm{Update}\left(\boldsymbol{h}_t, \sum_{s \in \mathcal{N}_{\mathrm{intra}}} \boldsymbol{m}_{s \to t}, \sum_{s' \in \mathcal{N}_{\mathrm{inter}}} \boldsymbol{\mu}_{s' \to t}\right),$$



Self-supervised Pre-training

GMPT-CL



Contrastive Loss

$$\ell_{i,j} = -\log \frac{\exp\left(s_{i,j}/\tau\right)}{\sum_{k\neq i}\exp\left(s_{i,k}/\tau\right)},$$

Approximate Contrastive Training

GMPT-CL

Graph Matching

-> node-node similarity calculation

-> high computation cost

-> additional $O(m^2 \cdot d)$, where $m = \sum_{i=1}^{2n} |\mathcal{V}_i|$

quadratic time and space cost

Approximate Contrastive Training

GMPT-CL

Sampling!

Comparison from $2n \times 2n$ to $\boxed{q} \times 2n$

Additional Cost	Naive	Approximate Contrastive Training
Time	$O(m^2 \cdot d)$	$O(\frac{q}{2n} \cdot m^2 \cdot d)$

Approximate Contrastive Training

GMPT-CL

Gradients Accumulation

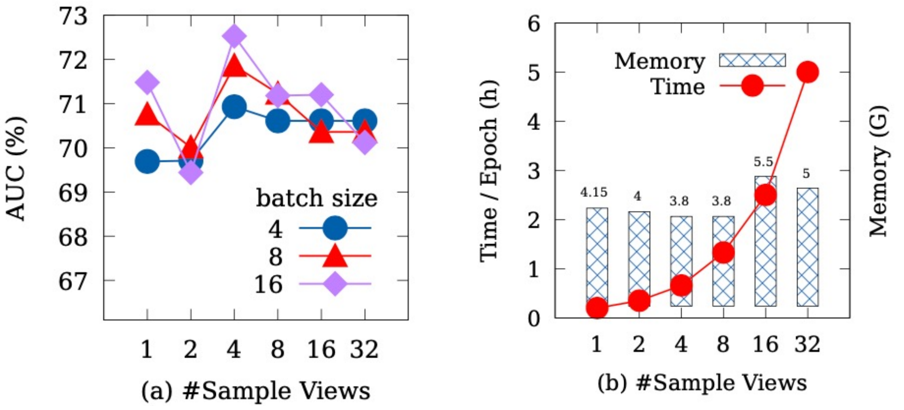
Additional Cost	Naive	Approximate Contrastive Training
Time	$O(m^2 \cdot d)$	$O(\frac{q}{2n} \cdot m^2 \cdot d)$
Space	$O(m^2 \cdot d)$	$O\left(\boxed{\frac{1}{2n}} \cdot m^2 \cdot d\right)$

Approximate Contrastive Training

GMPT-CL

Both theoretically and empirically verified

> **Lemma 3.2.** *Optimizing Eqn. (3.5) with approximate contrastive training algorithm has the same optimization lower bound as originally in expectation.*



Additional Cost	Naive	Approximate Contrastive Training
Time	$O(m^2 \cdot d)$	$O(\frac{q}{2n} \cdot m^2 \cdot d)$
Space	$O(m^2 \cdot d)$	$O\left(\frac{1}{2n} \cdot m^2 \cdot d\right)$

Supervised Pre-training

GMPT-Sup

Supervised pre-training with coarse-grained labeled graphs

Jointly predict graph properties

Supervised Pre-training

GMPT-Sup

Continuous labels as real-value vectors

$$s_p = \mathrm{sim}(\boldsymbol{y}_1, \boldsymbol{y}_2), s_g = \mathrm{sim}(\boldsymbol{z}_{G_1}, \boldsymbol{z}_{G_2}),$$

Similar graphs ~ similar labels

$$\ell_c = \mathrm{MSE}(s_p, s_g)$$

Supervised Pre-training

GMPT-Sup++

Discrete labels

$$\ell_d = \sum_{k=1,2} \mathrm{BCE}(\boldsymbol{y}_k, \boldsymbol{W}_k \cdot \boldsymbol{z}_{G_k} + \boldsymbol{b}_k),$$

Experiments

Benchmark datasets

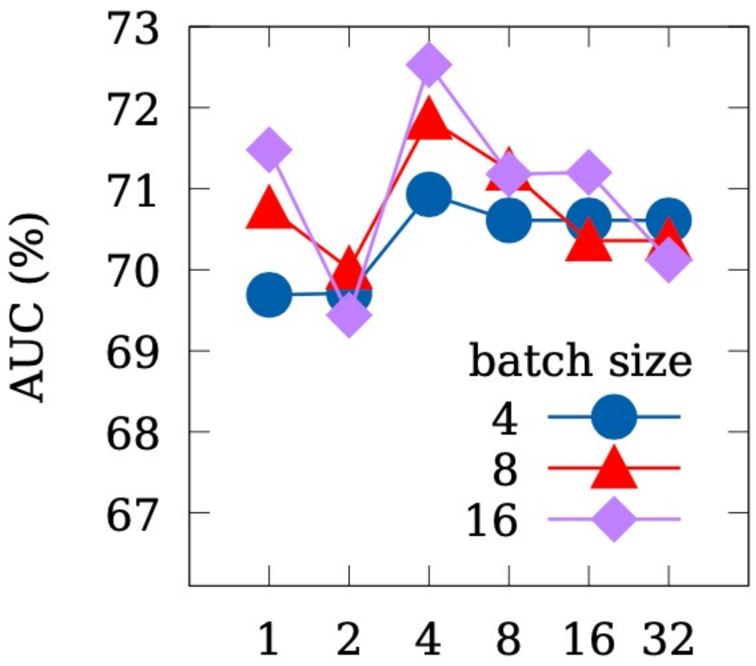
Dataset	Bio	Chem
#(sub)graphs for self-supervised PT	307K	2,000K
#(sub)graphs for supervised PT/FT	88K	456K
#Coarse-grained labels for PT	5,000	1,310
#Downstream FT tasks	40	678*

Experiments

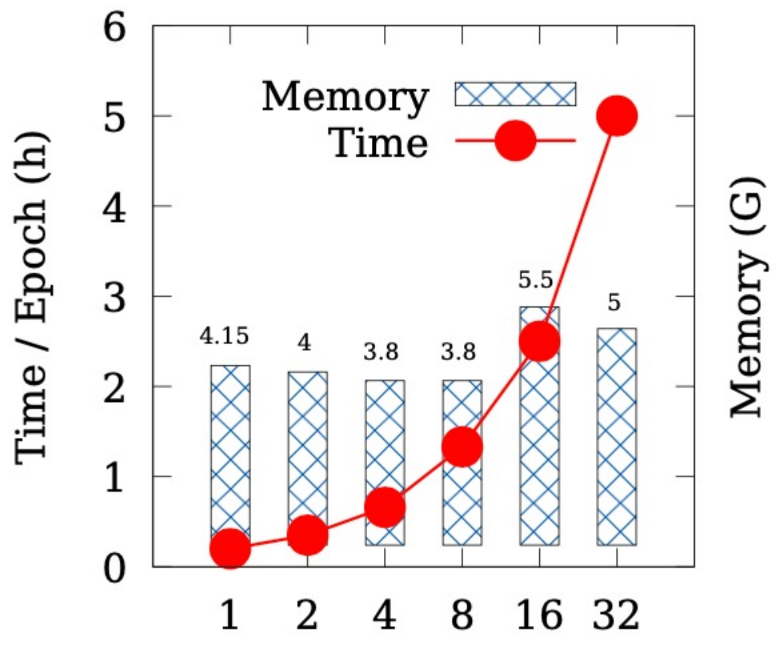
Pre-training methods	Bio					Chem
	GCN	GraphSAGE	GAT	GIN	Average	
w/o pre-training	63.20 ± 1.00	65.70 ± 1.20	68.20 ± 1.10	64.80 ± 1.00	65.48	67.0
Infomax	62.83 ± 1.22	67.21 ± 1.84	66.94 ± 2.61	64.10 ± 1.50	65.27	70.3
EdgePred	63.18 ± 1.12	66.05 ± 0.78	65.72 ± 1.17	65.70 ± 1.30	65.16	70.3
ContextPred	62.81 ± 1.87	66.47 ± 1.27	67.86 ± 1.19	65.20 ± 1.60	65.59	71.1
AttrMasking	62.40 ± 1.35	63.32 ± 1.01	61.72 ± 2.70	64.40 ± 1.30	62.96	70.9
GraphCL	67.05 ± 1.16	**71.53** ± 0.46	65.68 ± 3.98	67.88 ± 0.85	68.04	70.8
L2P-GNN	66.48 ± 1.59	69.89 ± 1.63	69.15 ± 1.86	70.13 ± 0.95	68.91	70.4
GMPT-CL	**70.65** ± 0.53	70.29 ± 0.21	**71.07** ± 0.14	**72.53** ± 0.42	**71.13**	**71.5**

Pre-training methods	Bio	Chem
w/o pre-training	64.8 ± 1.0	67.0
PropPred	69.0 ± 2.4	70.0
GMPT-Sup	**70.84 ± 0.59**	–
GMPT-Sup_{++}	70.73 ± 0.42	**70.4**

Experiments

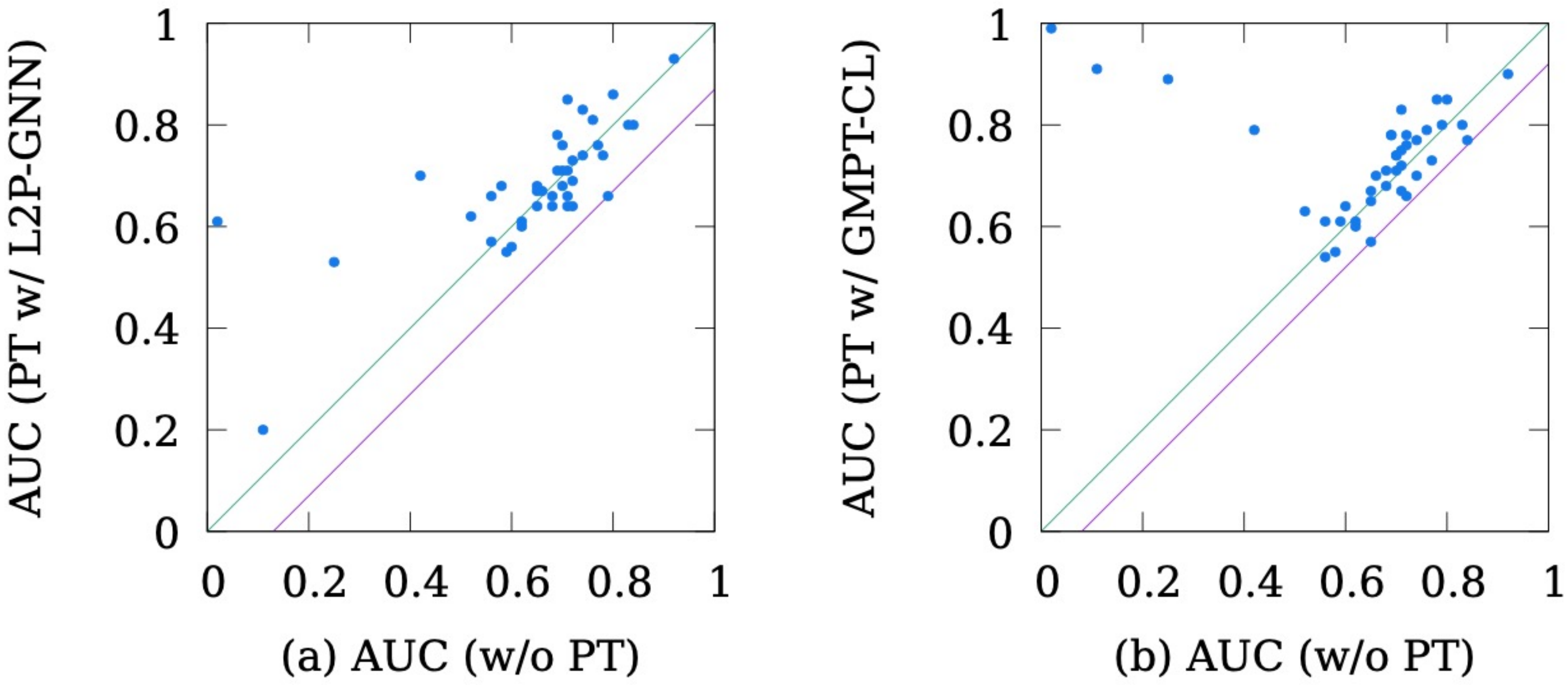


(a) #Sample Views



(b) #Sample Views

Experiments



Conclusion & QA

Yupeng Hou
houyupeng@ruc.edu.cn

- GMPT: Graph Matching tasks for pre-training GNNs
  - Node- and graph-level characteristics in one single task;
  - Adaptive graph representations;
  - 🌟 https://github.com/RUCAIBox/GMPT for code & pre-trained models