Would prompt work for graph learning?
An exploration of few-shot learning on graphs

Yuan FANG

School of Computing and Information Systems
Singapore Management University

VALSE Webinar

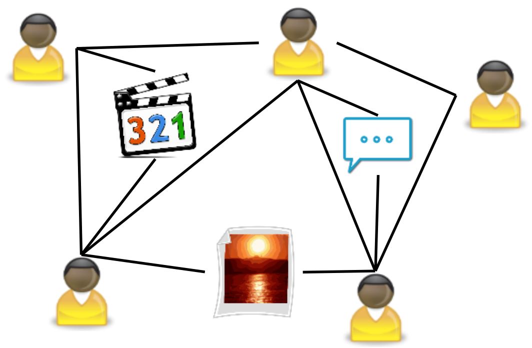
10 Jan 2024

Outline

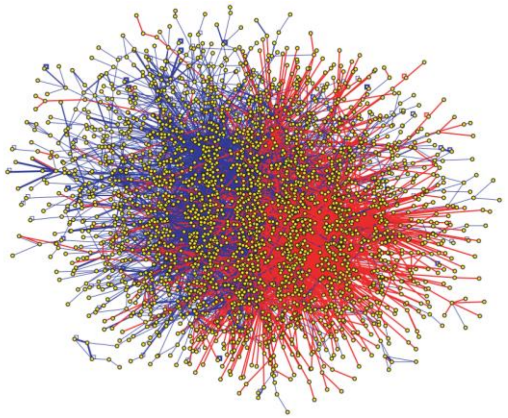
- **Introduction: Data and problems**
- Overview of few-shot methodologies
- Can prompt work on graph + text?
- Can prompt work on graph alone?
- Conclusion

Complex big data as graphs

Social networks

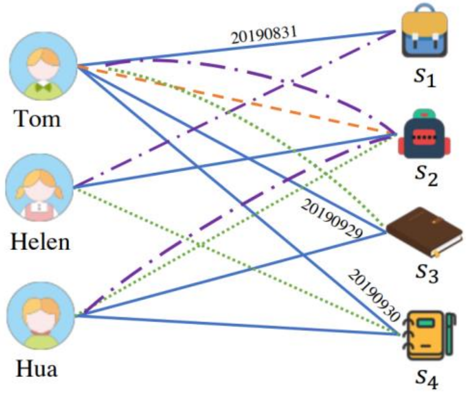


Biology



[Image from RVH05]

E-commerce

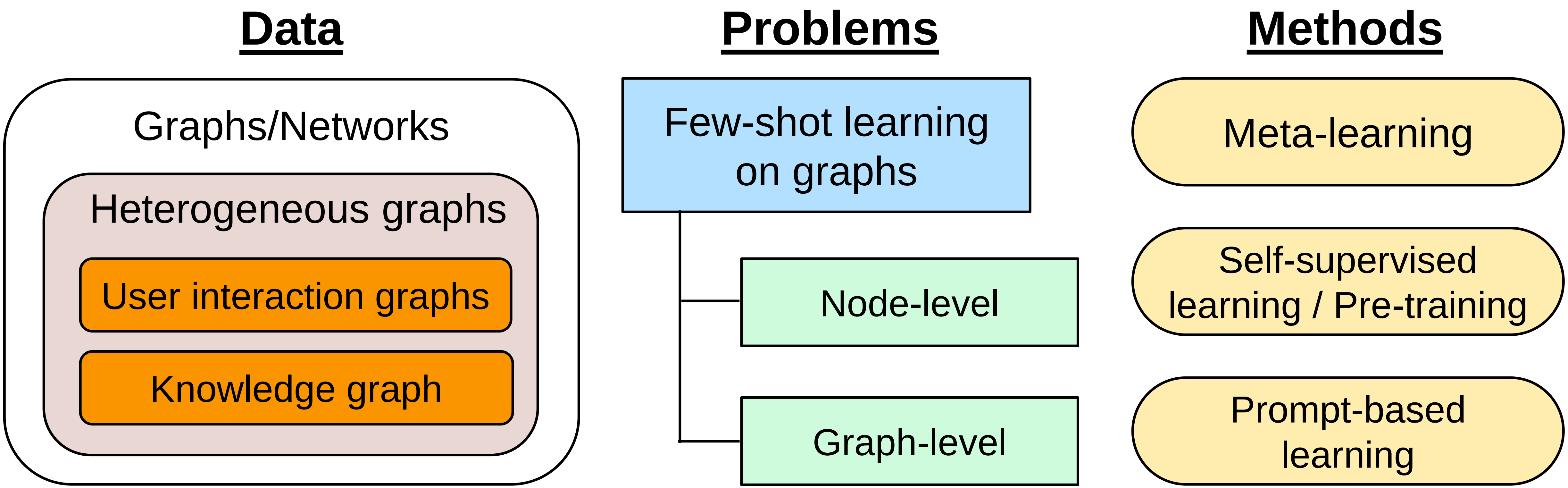


Knowledge graph

[Image from Microsoft]

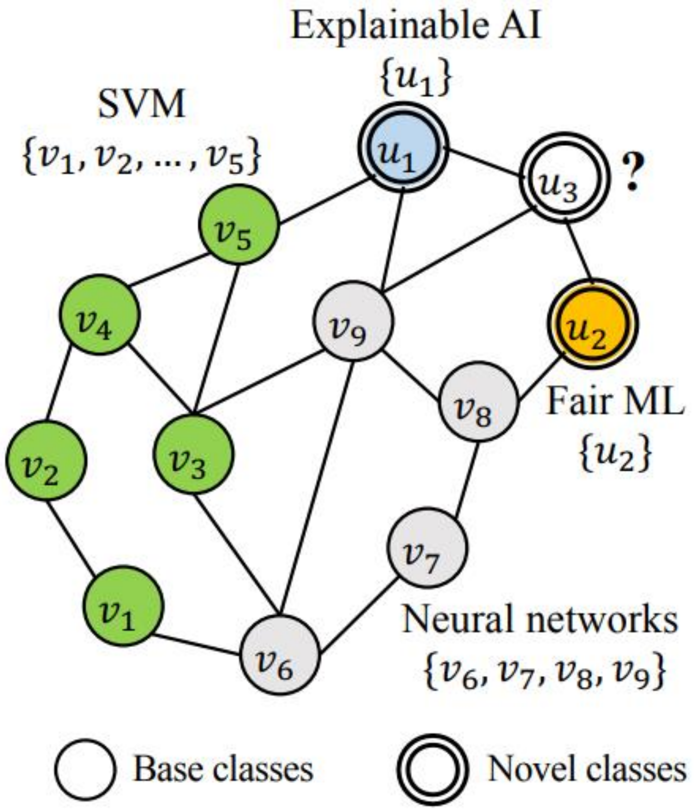
[RVH05] Towards a proteome-scale map of the human protein–protein interaction network. J. Rual, *et al*. Nature: 437(7062), 2005.

Data, Problems and Methods

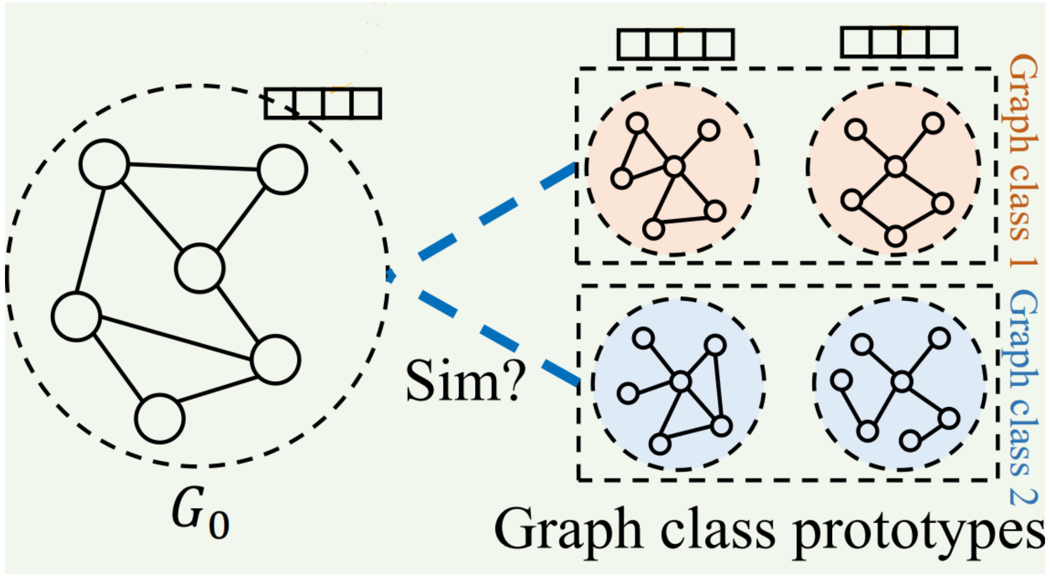


Few-shot problems on graphs

Node classification



Graph classification



[AAAI21] Z. Liu, Y. Fang, C. Liu and S. C. H. Hoi. *Relative and Absolute Location Embedding for Few-Shot Node Classification on Graph.*

[WWW23] Z. Liu, X. Yu, Y. Fang and X. Zhang. *GraphPrompt: Unifying Pre-Training and Downstream Tasks for Graph Neural Networks.*

Outline

- Introduction: Data and problems
- **Overview of few-shot methodologies**
- Can prompt work on graph + text?
- Can prompt work on graph alone?
- Conclusion

Why supervised learning does not work?

Supervised learning

Train

Cat

Dog

Banana

Apple

Test

?

?

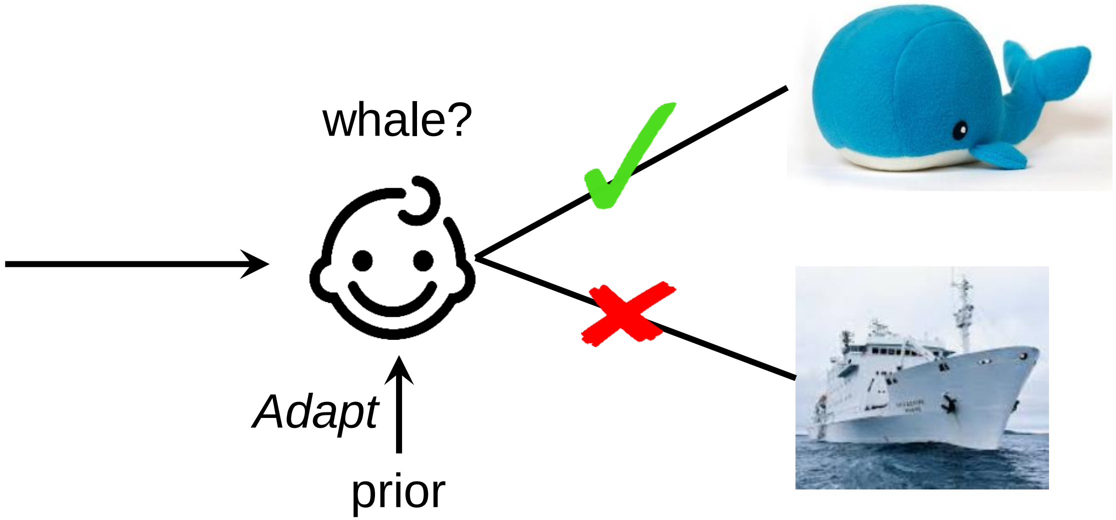
Learn a classifier

$f_\theta($  $) \rightarrow \mathrm{dog}$

Need many, many labelled data!
Hard to deal with novel classes.

How humans learn?

One example of toy whale



Even toddlers can learn novel classes very quickly with one/few examples…by generalizing from prior knowledge.

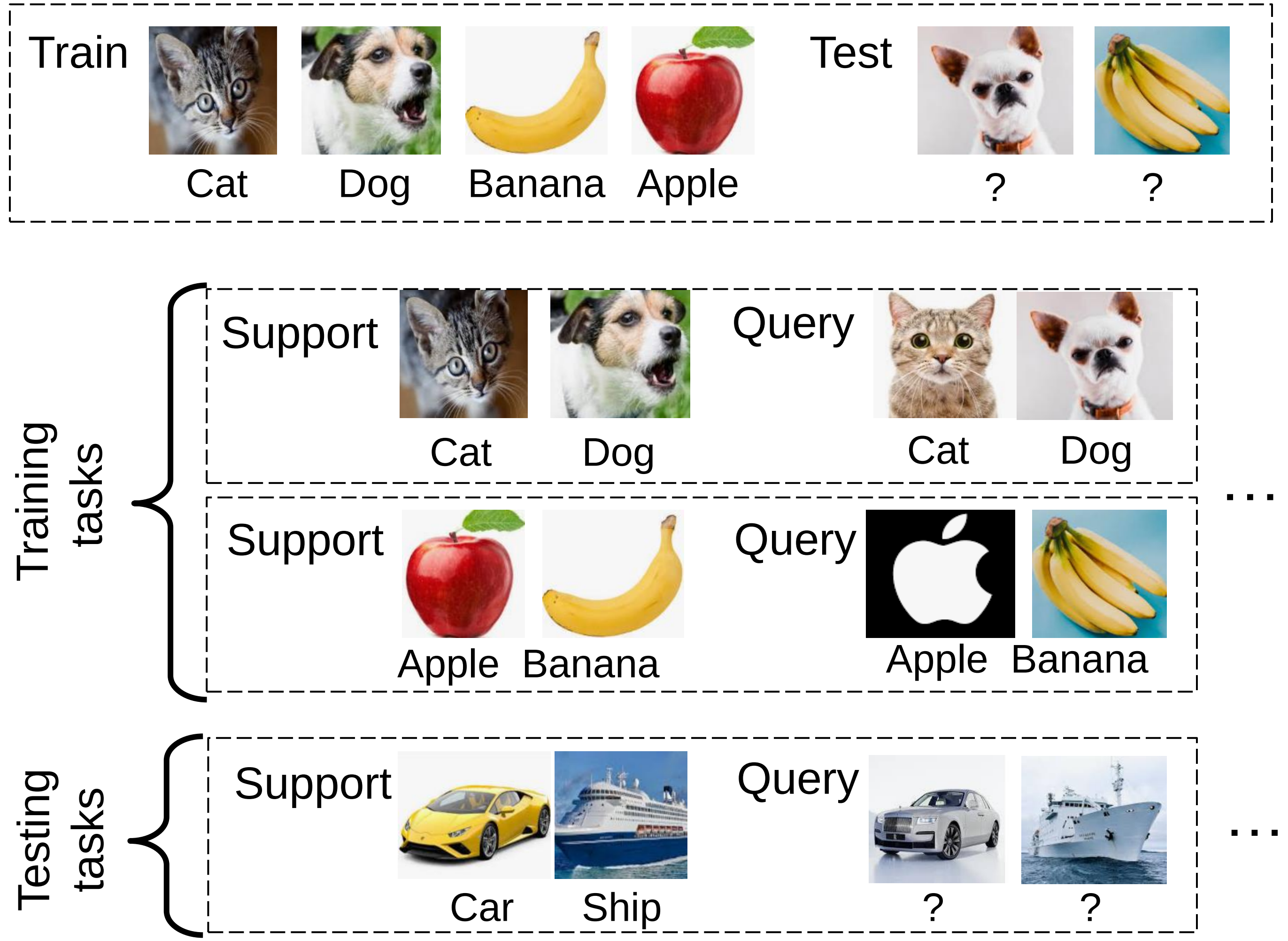
[Images from the Web]

From supervised learning to meta-learning

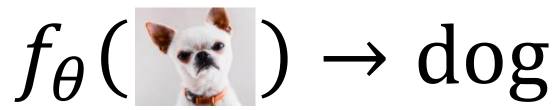
Supervised learning

Meta-learning

(MAML [FAL17])



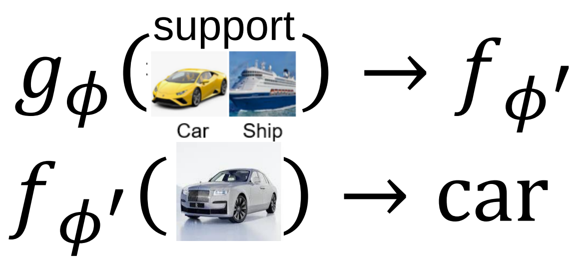
Learn a classifier



Need many, many labelled data!
Hard to deal with novel classes.

Learn a prior ϕ from the training tasks

↓ *Adapt*

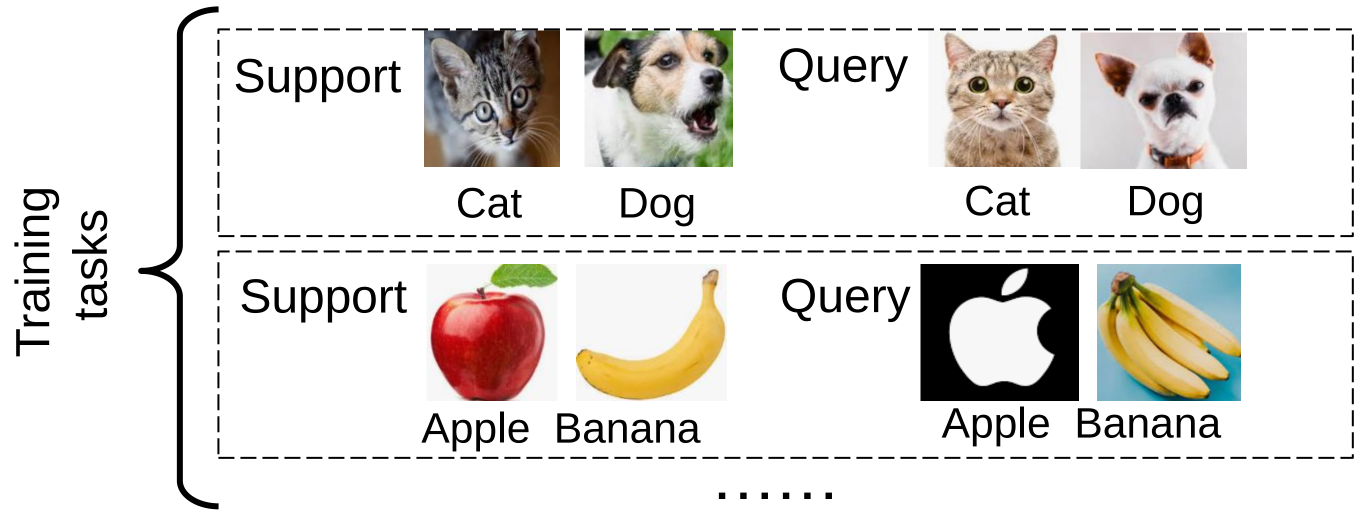


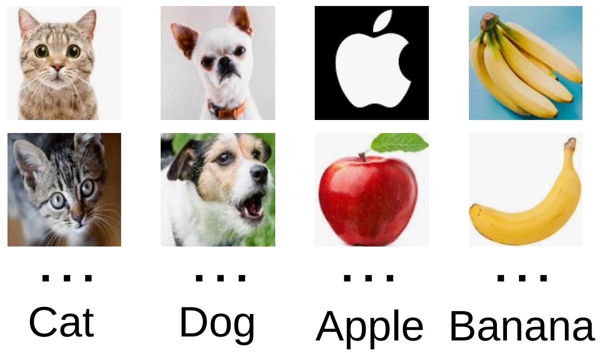
“Learn to learn”

[FAL17] Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. C. Finn *et al*. ICML 2017.

Self-supervised learning

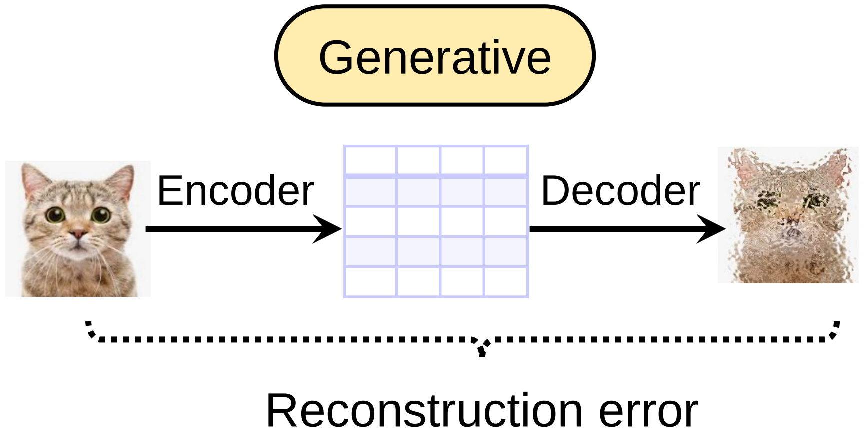
Meta-learning

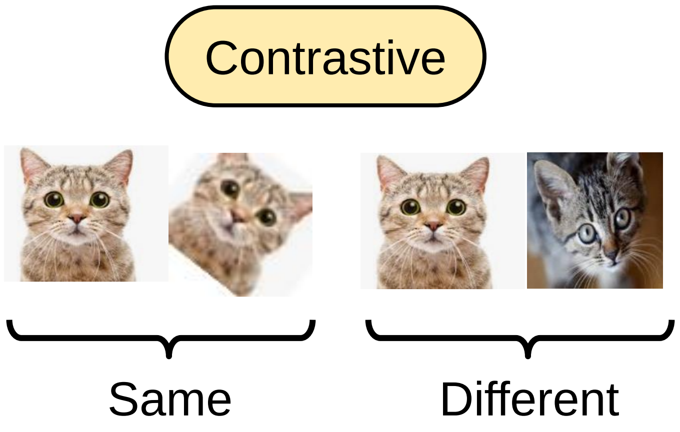




Still require many labels on these base classes to form training tasks

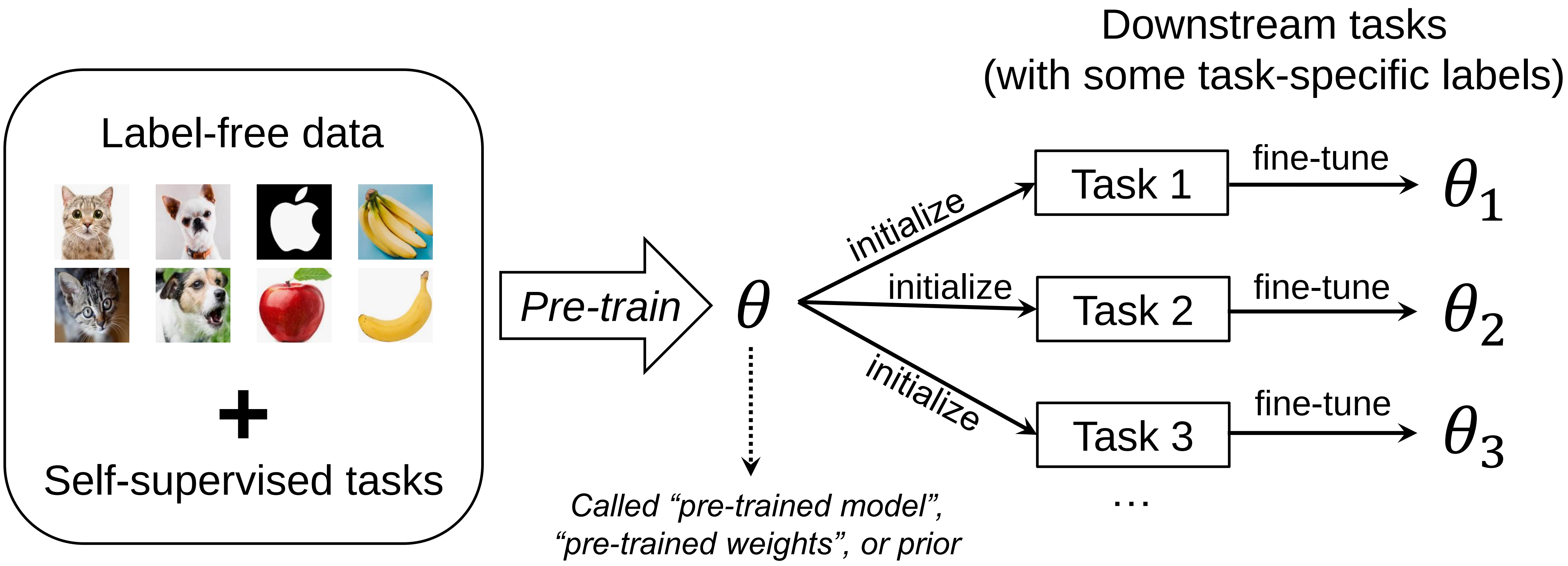
Self-supervised learning





“Free” supervision, no annotation cost!

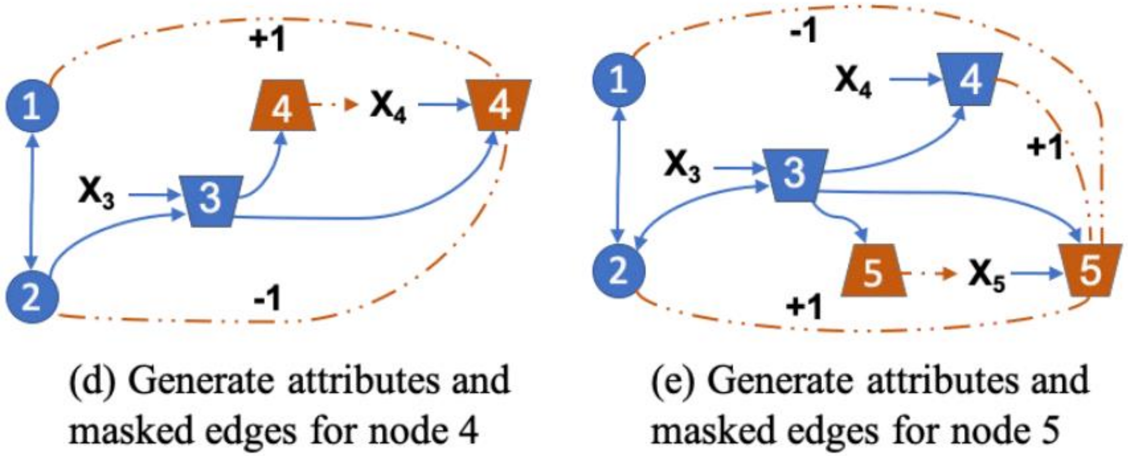
Self-supervised learning / Pre-training



Graph pre-training: Generative vs. contrastive

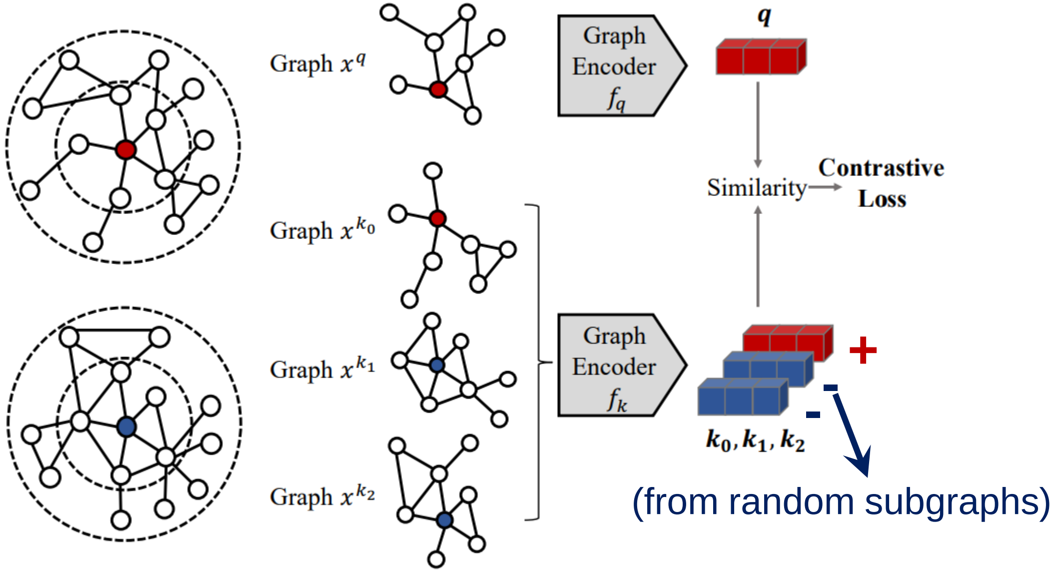
- Key: Design self-supervised pre-training tasks on graphs

Generative



[Image from HDW20]

Contrastive



[Image from QCD20]

[HDW20] GPT-GNN: Generative Pre-Training of Graph Neural Networks. Z. Hu *et al*. KDD 2020
[QCD20] GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training. J. Qiu *et al*. KDD 2020

Graph pre-training: Spatial vs. Spectral

Spatial

Explicit (local) structures and node features

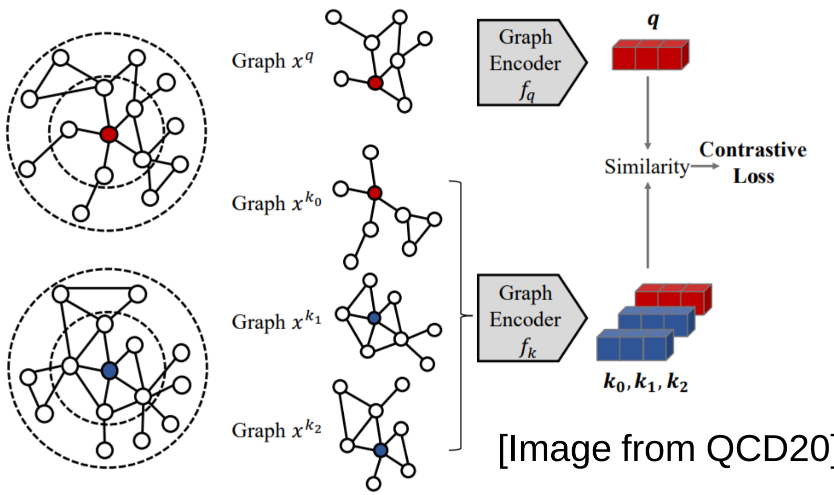
$$\mathbf{H}_a = f(\mathbf{A}, \mathbf{X})$$

Contrast

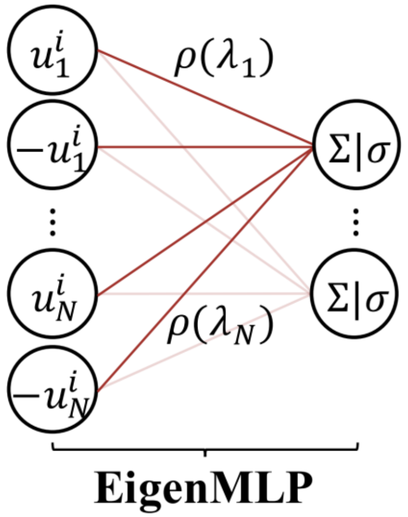
Spectral

Implicit node (global) positions on graph topology

$$\mathbf{H}_e = g(\mathbf{\Lambda}, \mathbf{U})$$



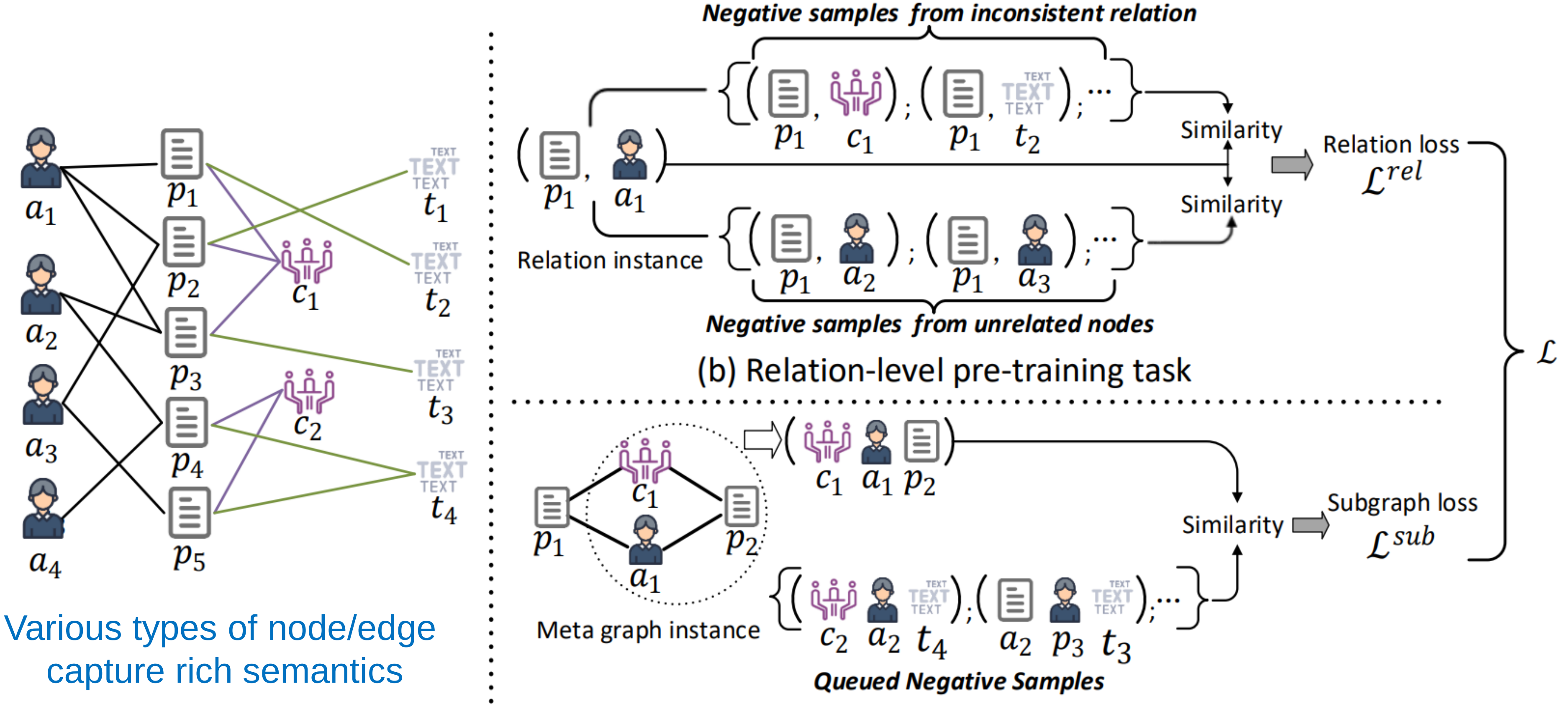
[Image from QCD20]



[NeurIPS23] Deyu Bo, Yuan Fang, Yang Liu, Chuan Shi. *Graph Contrastive Learning with Stable and Scalable Spectral Encoding*

Pre-training on heterogeneous graphs

- Pre-training tasks to capture relation- and subgraph-level semantics



(a) A heterogeneous graph

(b) Relation-level pre-training task

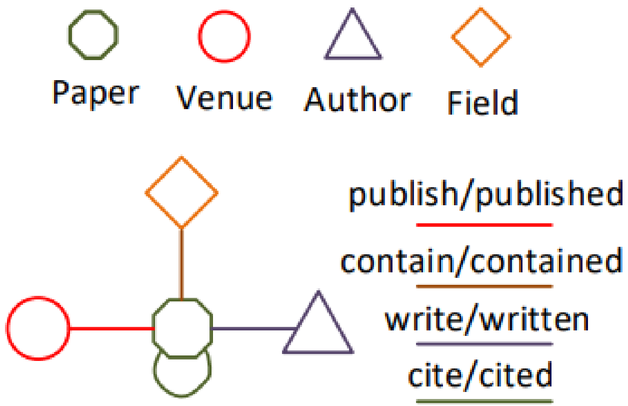
(c) Subgraph-level pre-training task

[CIKM21] X. Jiang, Y. Lu, Y. Fang and C. Shi. *Contrastive Pre-training of GNNs on Heterogeneous Graphs*

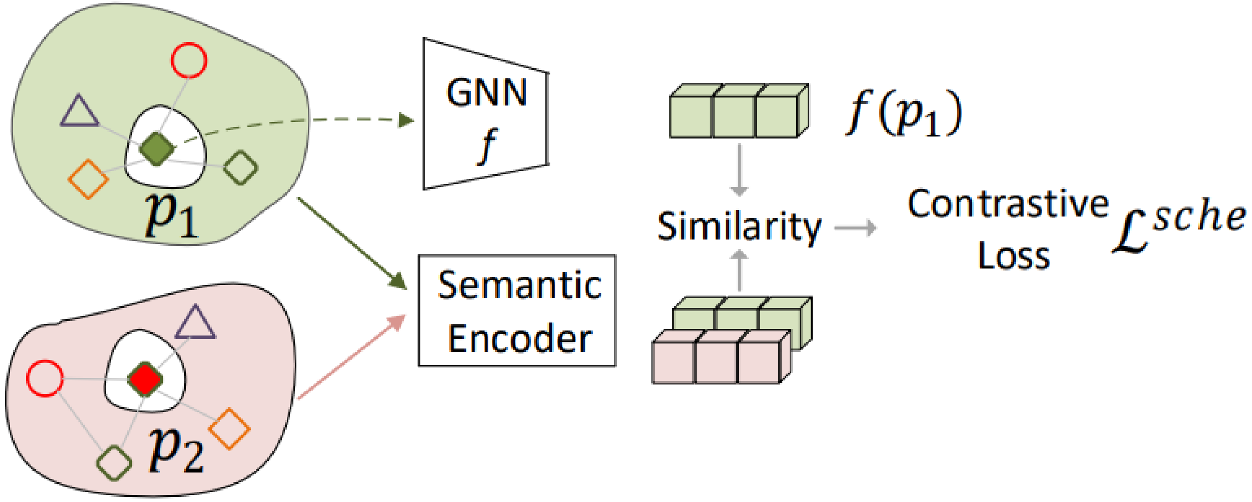
Pre-training on heterogeneous graphs

- Pre-training tasks to capture schema-level semantics

Schema



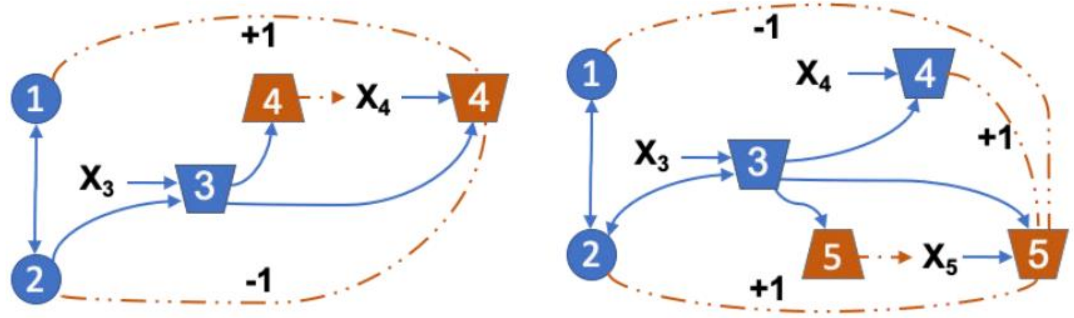
Schema-level task



[KDD21] X. Jiang, T. Jia, C. Shi, Y. Fang, Z. Lin and H. Wang. *Pre-training on Large-Scale Heterogeneous Graph.*

Problem with pre-training approaches

- The gap between pre-training and downstream objectives



(d) Generate attributes and masked edges for node 4

(e) Generate attributes and masked edges for node 5

- And the fine-tuning step..
  - Can be expensive for large pre-trained models
  - may overfit if there are very few labels from downstream tasks

Bridging the gap: Pre-train, prompt

- **Problem:** Gap between pre-training and downstream tasks
- **Prompt** [LYF23]**:** an alternative to "pre-train, fine-tune"
  - Originated in NLP, an instruction to reformulate the original task to unify with pre-trained model (e.g., masked language modeling)

<u>Task: Sentiment classification</u> **<u>+ Prompt</u>**

"I missed the bus today." **"I felt so ___"**

Ask pre-trained model to fill in the blank

happy +

unlucky –

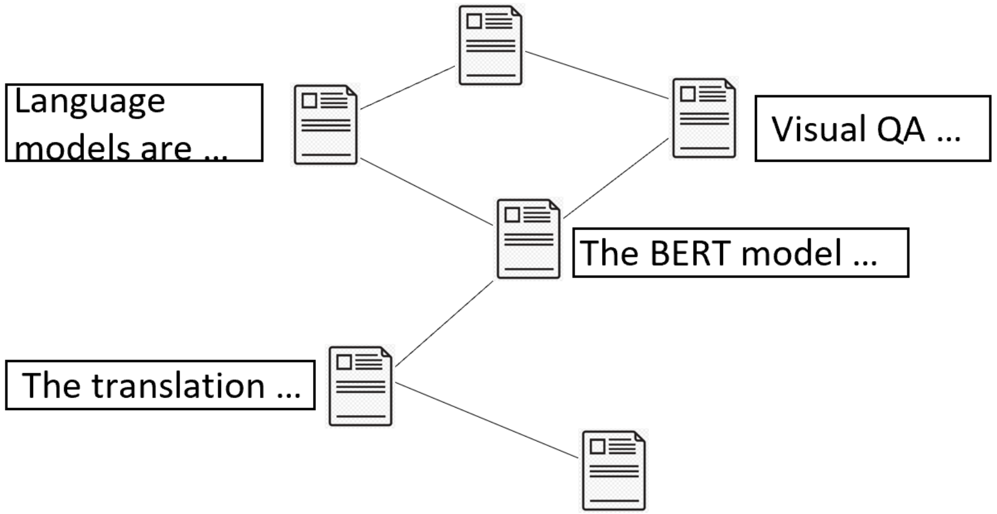
Zero-shot: Handcrafted (prompt engineering)
Few-shot: Learnable word vectors (prompt tuning)

[LYF23] Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. P. Liu, *et al*. ACM Computing Surveys: 55(9), 2023.

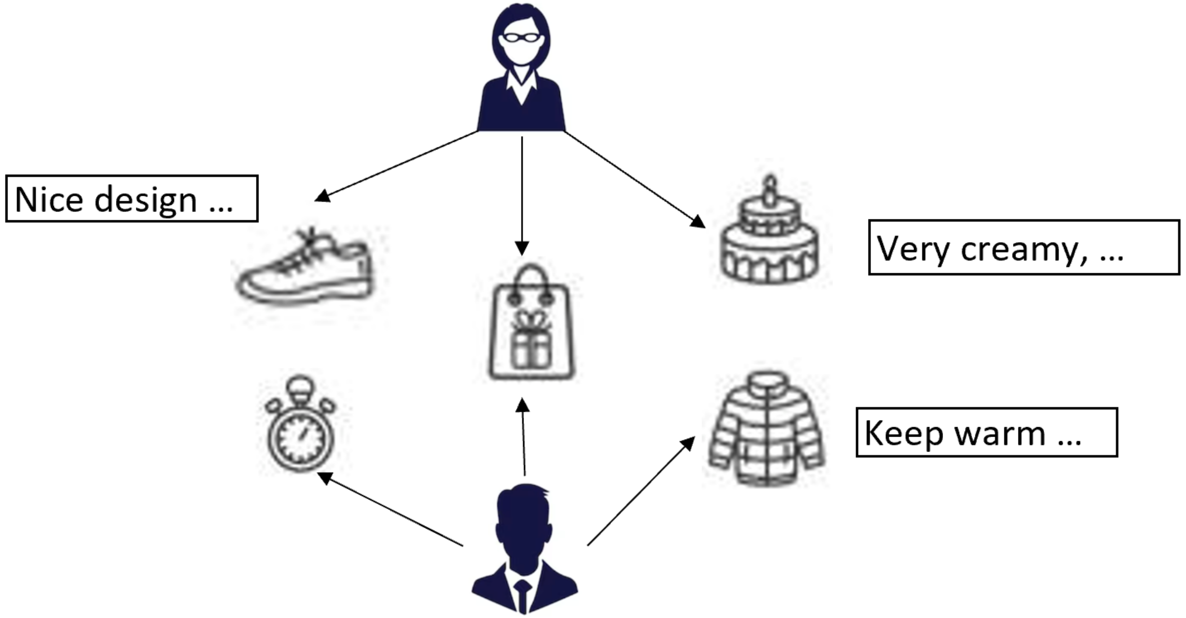
Outline

- Introduction: Data and problems
- Overview of few-shot methodologies
- **Can prompt work on graph + text?**
- Can prompt work on graph alone?
- Conclusion

Graph data often associate with texts



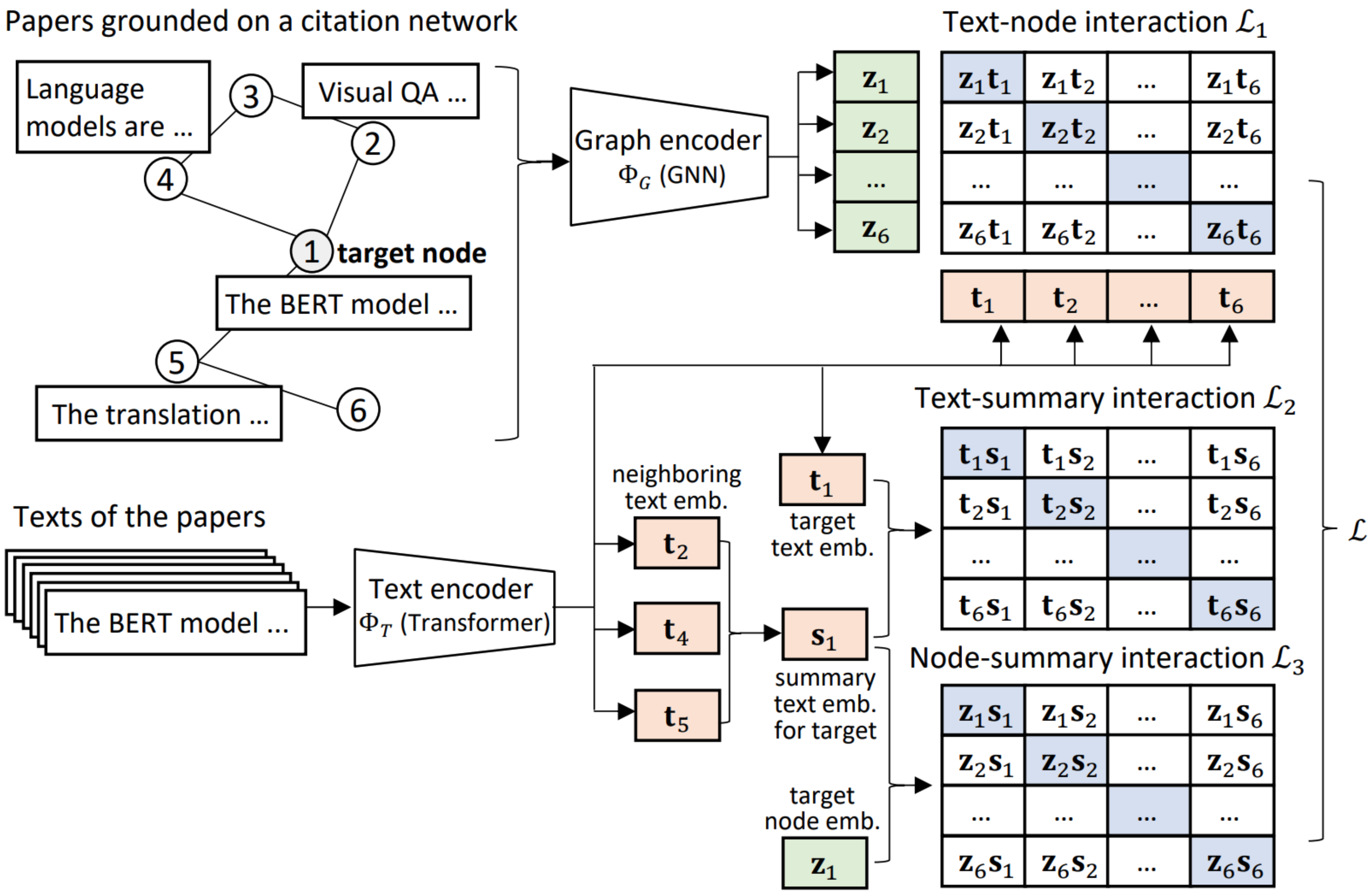
Citation graph for online articles



E-commerce item review graph

So, if there is a **jointly pre-trained graph-text model**, we can easily apply natural language-based prompts to graphs.

Graph-grounded **pre-training** and prompting (G2P2)



(a) Graph-grounded contrastive pre-training

Learns a dual-modal embedding space by jointly training a **text encoder** and **graph encoder**

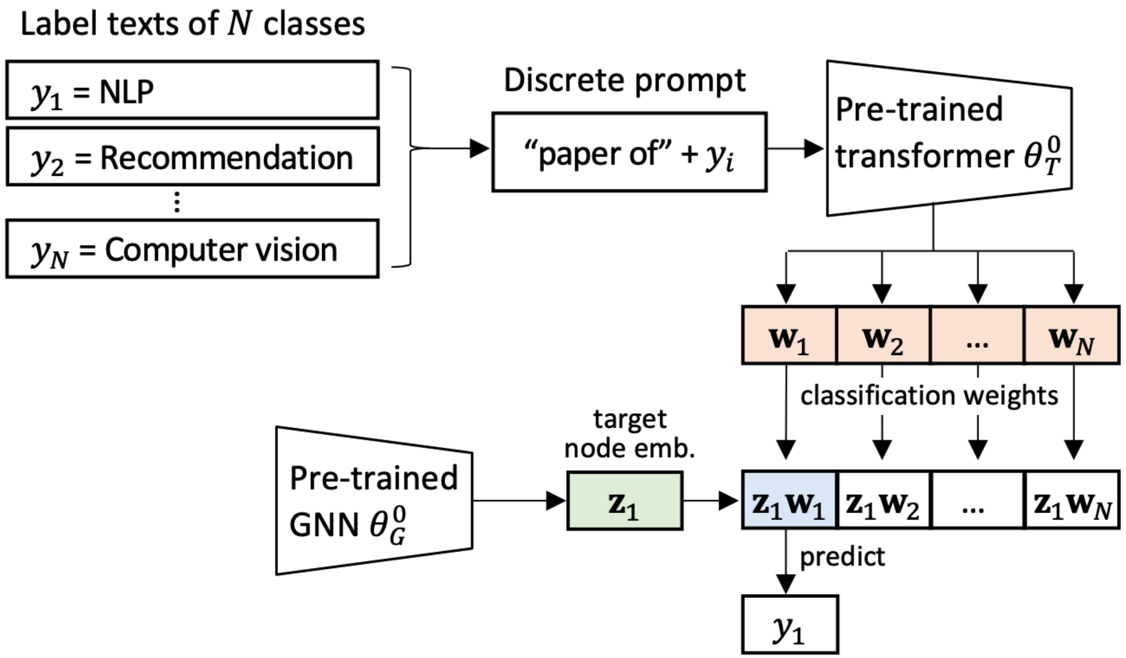
Exploits **three contrastive strategies**

- Text-node contrast
- Text-summary contrast
- Node-summary contrast

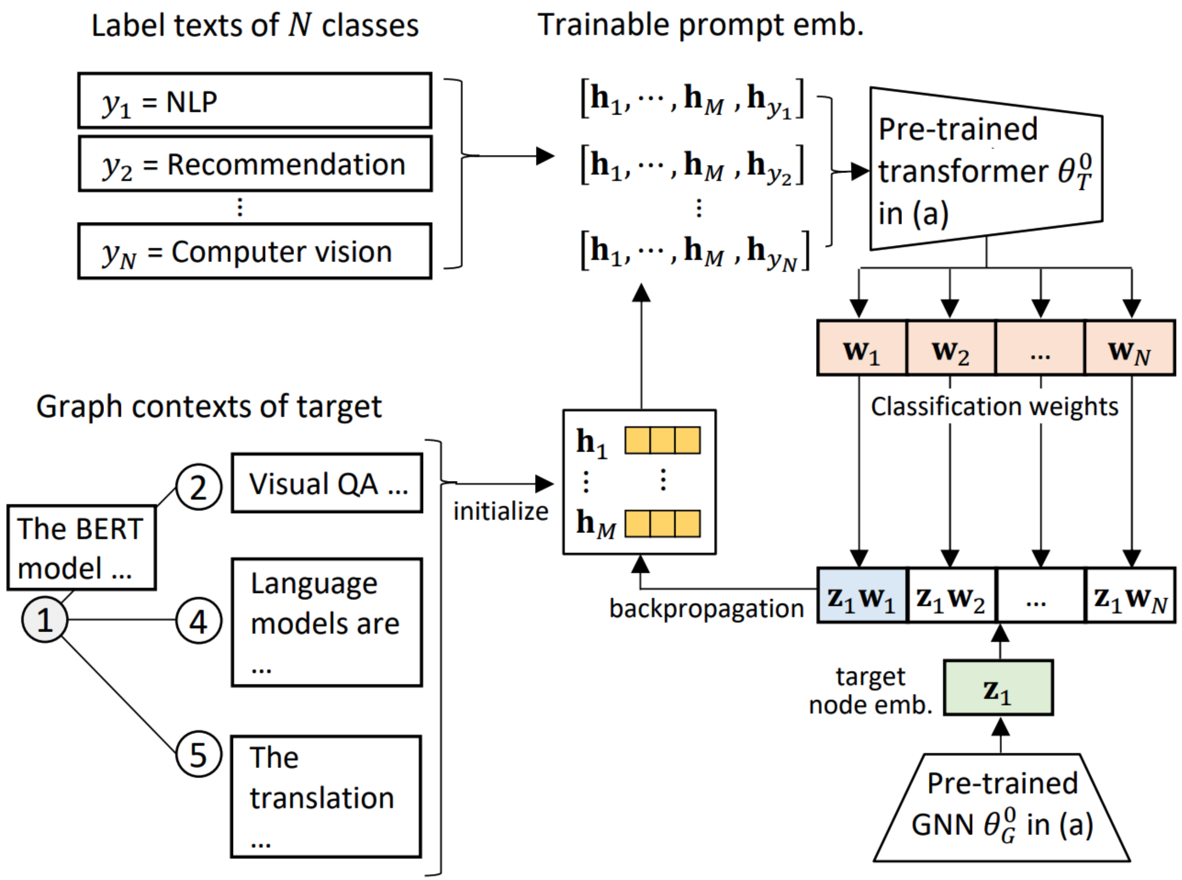
[SIGIR23] Z. Wen and Y. Fang. *Augmenting Low-Resource Text Classification with Graph-Grounded Pre-training and Prompting.*

Graph-grounded pre-training and **prompting** (G2P2)

Zero-shot node classification with discrete prompts

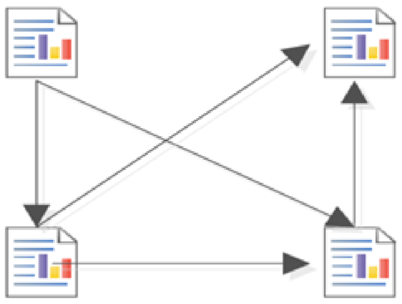


Few-shot node classification with continuous prompt tuning

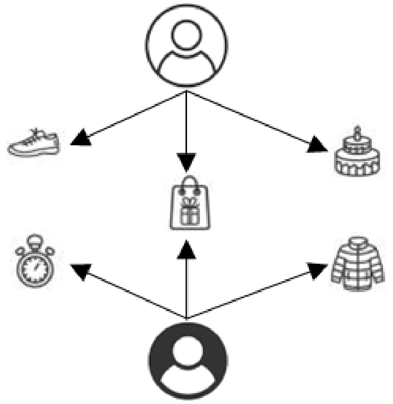


Datasets to evaluate G2P2

Dataset	Cora	Art	Industrial	M.I.
# Documents	25,120	1,615,902	1,260,053	905,453
# Links	182,280	4,898,218	3,101,670	2,692,734
# Avg. doc length	141.26	54.23	52.15	84.66
# Avg. node deg	7.26	3.03	2.46	2.97
# Classes	70	3,347	2,462	1,191



Cora is a collection of research papers with citation links



Art, **Industrial** and **Music Instruments** (M.I.) are three Amazon review datasets

Empirical performance of G2P2

		Cora		Art		Industrial		M.I.	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
End-to-end GNN	GCN	41.15±2.41	34.50±2.23	22.47±1.78	15.45±1.14	21.08±0.45	15.23±0.29	22.54±0.82	16.26±0.72
	SAGE_{sup}	41.42±2.90	35.14±2.14	22.60±0.56	16.01±0.28	20.74±0.91	15.31±0.37	22.14±0.80	16.69±0.62
	TextGCN	59.78±1.88	55.85±1.50	43.47±1.02	32.20±1.30	53.60±0.70	45.97±0.49	46.26±0.91	38.75±0.78
Pre-trained GNN	GPT-GNN	76.72±2.02	72.23±1.17	65.15±1.37	52.79±0.83	62.13±0.65	54.47±0.67	67.97±2.49	59.89±2.51
	DGI	78.42±1.39	74.58±1.24	65.41±0.86	53.57±0.75	52.29±0.66	45.26±0.51	68.06±0.73	60.64±0.61
	$\text{SAGE}_{\text{self}}$	77.59±1.71	73.47±1.53	76.13±0.94	65.25±0.31	71.87±0.61	65.09±0.47	77.70±0.48	70.87±0.59
Pre-trainned Transformers	BERT	37.86±5.31	32.78±5.01	46.39±1.05	37.07± 0.68	54.00±0.20	47.57±0.50	50.14±0.68	42.96±1.02
	BERT*	27.22±1.22	23.34±1.11	45.31±0.96	36.28±0.71	49.60±0.27	43.36±0.27	40.19±0.74	33.69±0.72
	RoBERTa	62.10±2.77	57.21±2.51	72.95±1.75	62.25±1.33	76.35±0.65	70.49±0.59	70.67±0.87	63.50±1.11
	RoBERTa*	67.42±4.35	62.72±3.02	74.47±1.00	63.35±1.09	77.08±1.02	71.44±0.87	74.61±1.08	67.78±0.95
Prompt tuning	P-Tuning v2	71.00±2.03	66.76±1.95	76.86±0.59	66.89±1.14	79.65±0.38	74.33±0.37	72.08±0.51	65.44±0.63
	G2P2-p	79.16±1.23	74.99±1.35	79.59±0.31	68.26±0.43	80.86±0.40	74.44±0.29	81.26±0.36	74.82±0.45
	G2P2	**80.08***±1.33	**75.91***±1.39	**81.03***±0.43	**69.86***±0.67	**82.46***±0.29	**76.36***±0.25	**82.77***±0.32	**76.48***±0.52
	(improv.)	(+2.12%)	(+1.78%)	(+5.43%)	(+4.44%)	(+3.53%)	(+2.7%)	(+6.53%)	(+7.92%)

G2P2 outperforms the best baseline by around 3–7%.

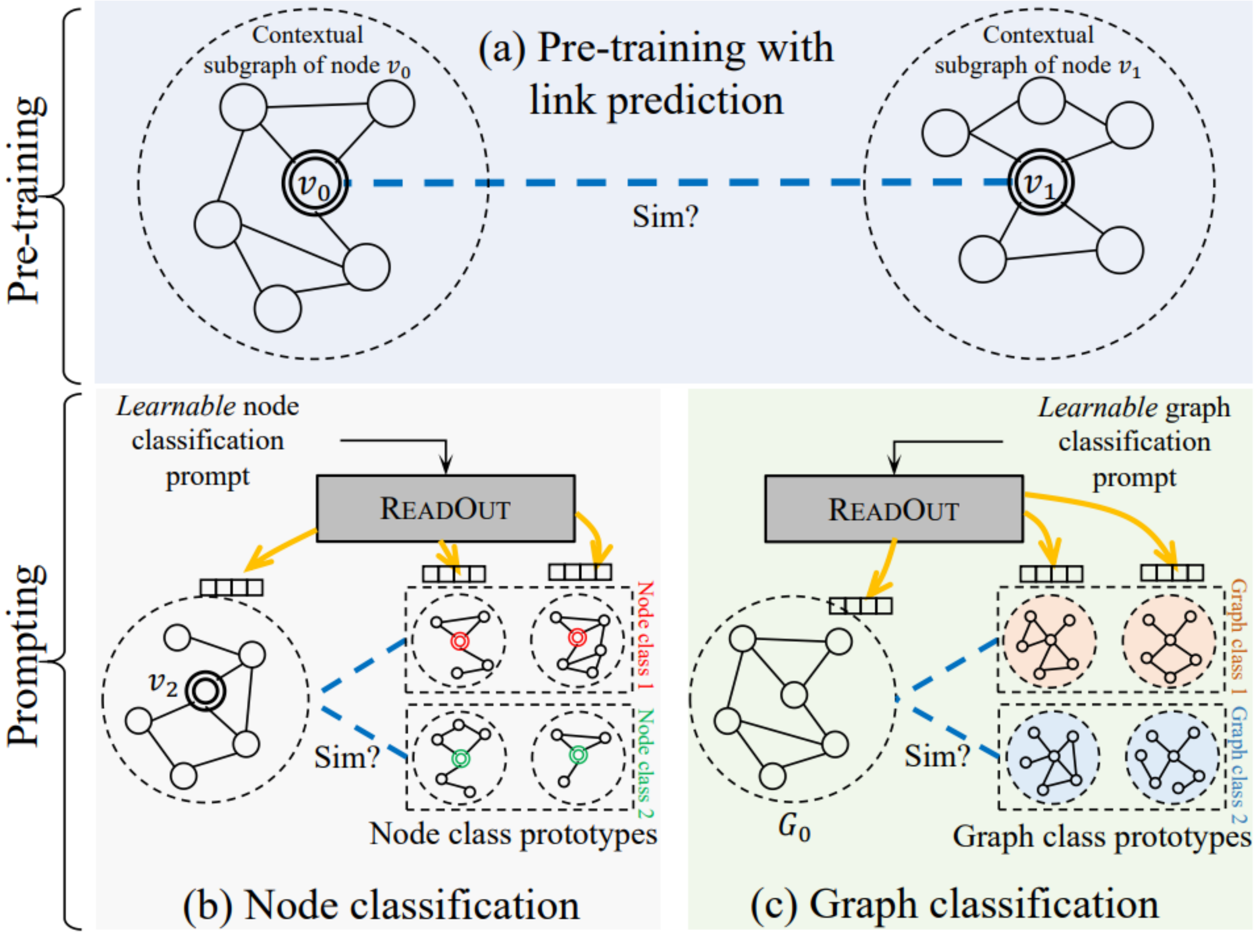
Outline

- Introduction: Data and problems
- Overview of few-shot methodologies
- Can prompt work on graph + text?
- **Can prompt work on graph alone?**
- Conclusion

GraphPrompt: Pre-train, prompt on **graph only**

- Two challenges
  - How to unify various pre-training and downstream tasks on graph?
  - How to design prompts on graph?

- Insights
  - A **unified task template** based on subgraph similarity computation
  - Use a **learnable prompt** to guide graph readout for different tasks



[WWW23] Z. Liu, X. Yu, Y. Fang and X. Zhang. *GraphPrompt: Unifying Pre-Training and Downstream Tasks for Graph Neural Networks.*

GraphPrompt: Pre-train, prompt on **graph only**

Unified task template

Link prediction

Triplet (v, a, b), s.t. v is linked to a, but not b:

$$\mathrm{sim}(\mathbf{s}_v, \mathbf{s}_a) > \mathrm{sim}(\mathbf{s}_v, \mathbf{s}_b)$$

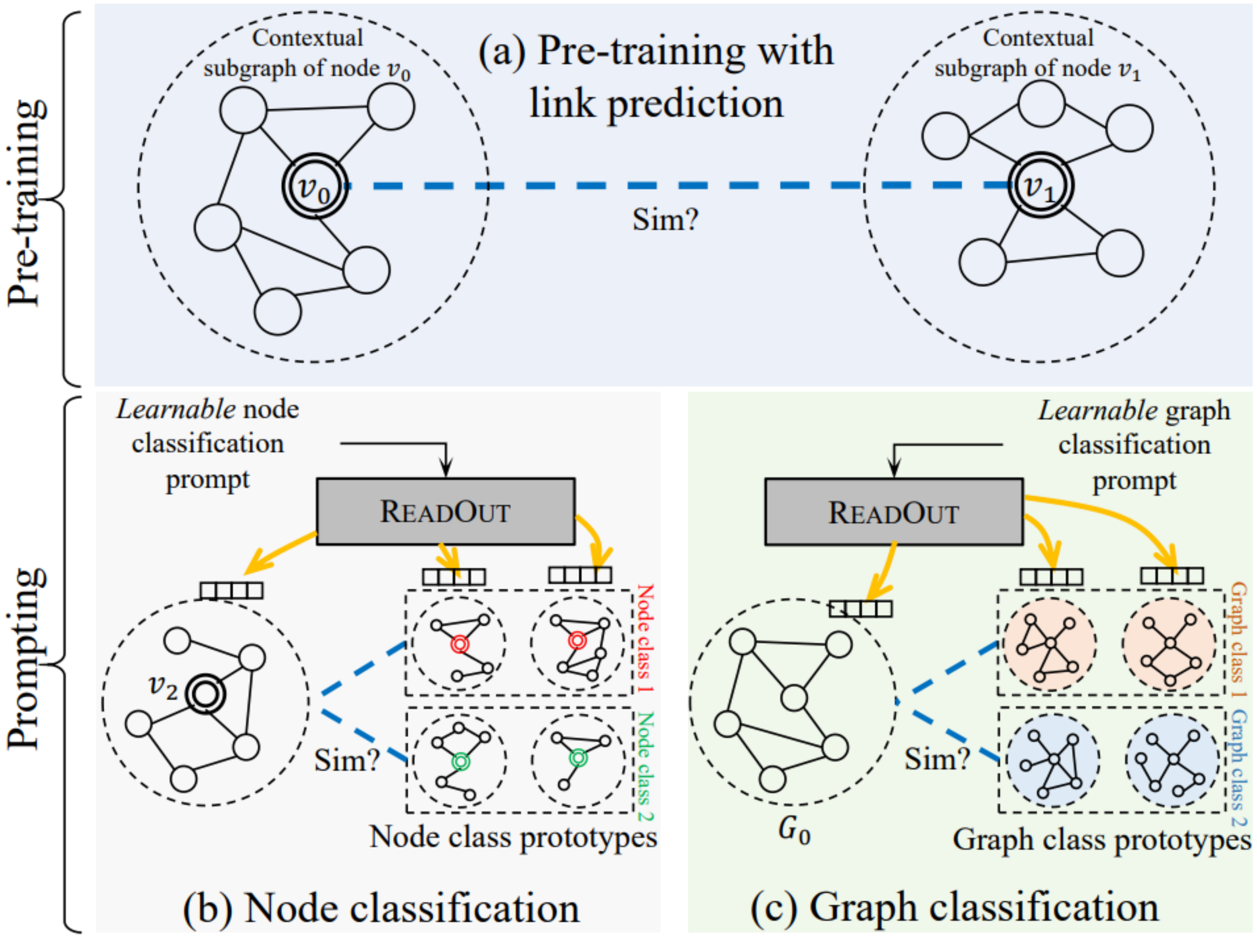
Node classification

$$\ell_j = \arg\max_{c \in C} \mathrm{sim}(\mathbf{s}_{v_j}, \tilde{\mathbf{s}}_c)$$

Graph classification

$$L_j = \arg\max_{c \in C} \mathrm{sim}(\mathbf{s}_{G_j}, \tilde{\mathbf{s}}_c)$$

All tasks converted to subgraph similarity computation!



$\mathbf{s}_x$: (sub)graph embedding of x (x is a node or graph)

$\tilde{\mathbf{s}}_c$: class c's prototype (a virtual subgraph, by aggregates all subgraph embeddings in the class)

GraphPrompt: Pre-train, prompt on graphs

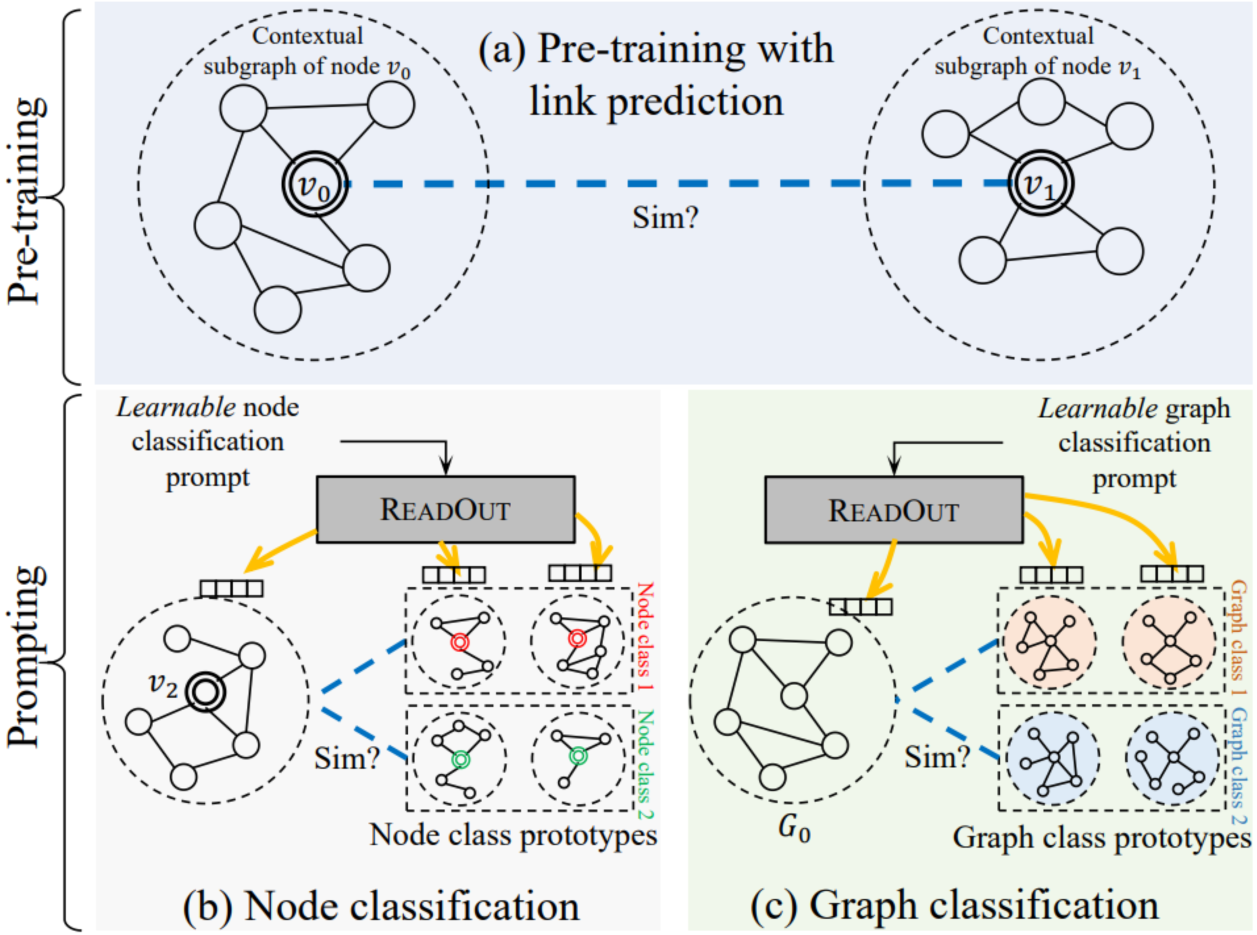
Prompt design

Different downstream tasks require different subgraph readout
→ Use task-specific learnable prompts

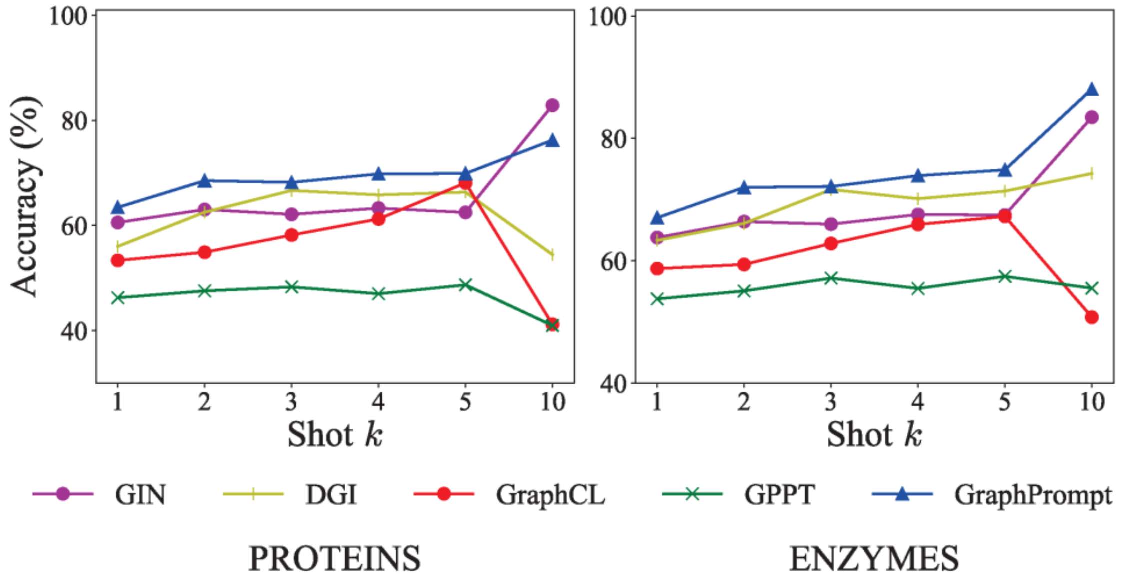
Prompt vector added to the readout layer of the pre-trained GNN

$$\mathbf{s}_{t,x} = \textsc{ReadOut}(\{\mathbf{p}_t \odot \mathbf{h}_v : v \in V(S_x)\})$$

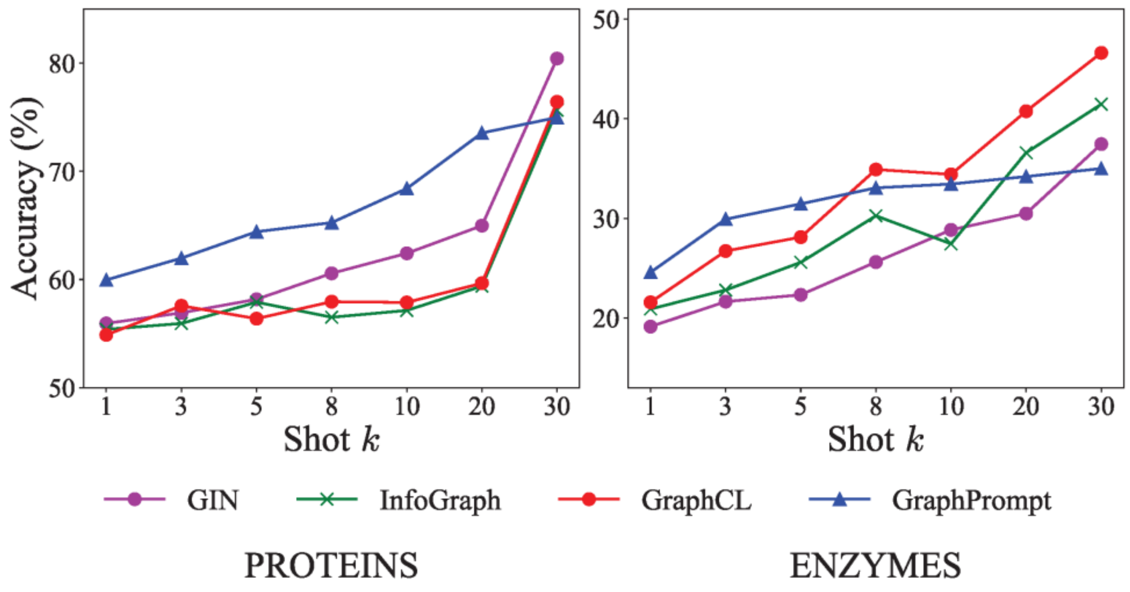
$\mathbf{s}_{t,x}$: (sub)graph embedding of x for a task t
$\mathbf{h}_v$: node v's embedding vector
$\mathbf{p}_t$ or $\mathbf{P}_t$: learnable prompt vector or matrix for task t



GraphPrompt: Pre-train, prompt on graphs



Impact of shots on few-shot node classification.



Impact of shots on few-shot graph classification.

<u>Few-shot:</u> Significantly better

<u>10-shot:</u> Still competitive
(as graphs are small – 10 shots are a lot)

<u>Few-shot:</u> Significantly better

<u>On ENZYMES:</u> worse performance on ≥20 shots
(only 600 graphs – 20 shots/class ~ 20% labels)

GraphPrompt: Pre-train, prompt on graphs

Comparison of parameter efficiency

Significantly fewer parameters/FLOPs than:
- *Supervised model (GIN [XHL19]),*
- *“Pretrain, fine-tune” model (GraphPrompt-ft),*
- *Existing prompt model (GPPT [SZH22])*

Methods	Flickr	
	Params	FLOPs
GIN	22,183	240,100
GPPT	4,096	4,582
GraphPrompt	96	96
GraphPrompt-ft	21,600	235,200

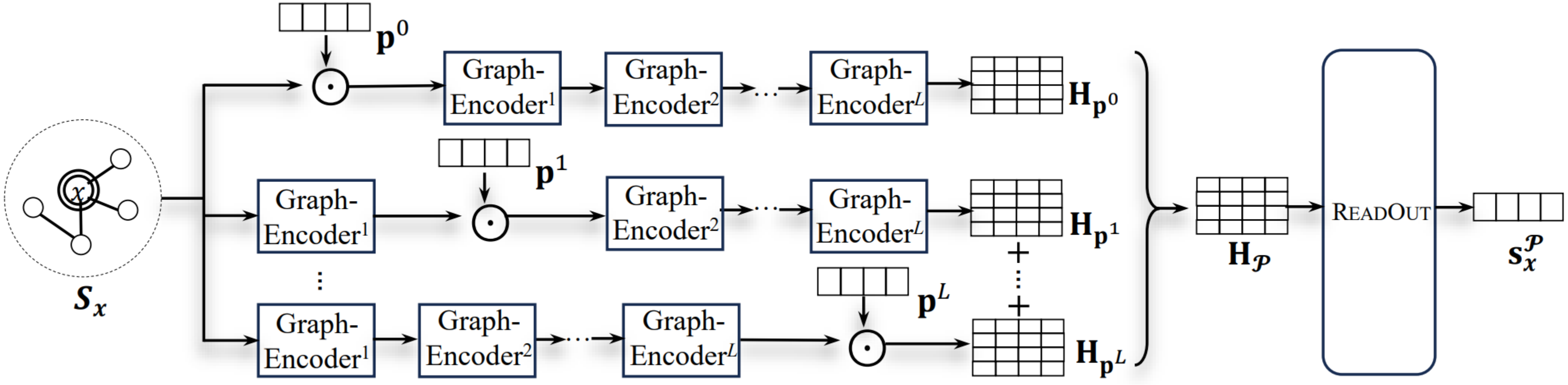
Methods	PROTEINS		ENZYMES	
	Params	FLOPs	Params	FLOPs
GIN	5,730	12,380	6,280	11,030
GPPT	1,536	1,659	1,536	1,659
GraphPrompt	96	96	96	96
GraphPrompt-ft	6,176	13,440	6,176	10,944

[XHL19] How Powerful are Graph Neural Networks? K. Xu *et al*. ICLR 2019
[SZH22] GPPT: Graph Pre-training and Prompt Tuning to Generalize Graph Neural Networks. M. Sun *et al*. KDD 2022

Generalized Graph Prompt

- Support more pre-training tasks beyond link prediction
  - DGI, InfoGraph, GraphCL, GCC, …

- Layer-wise prompts



Generalized Graph Prompt: Toward a Unification of Pre-Training and Downstream Tasks on Graphs. Xingtong Yu, Zhenghao Liu, Yuan Fang, Zemin Liu, Sihong Chen, Xinming Zhang. https://arxiv.org/pdf/2311.15317.pdf

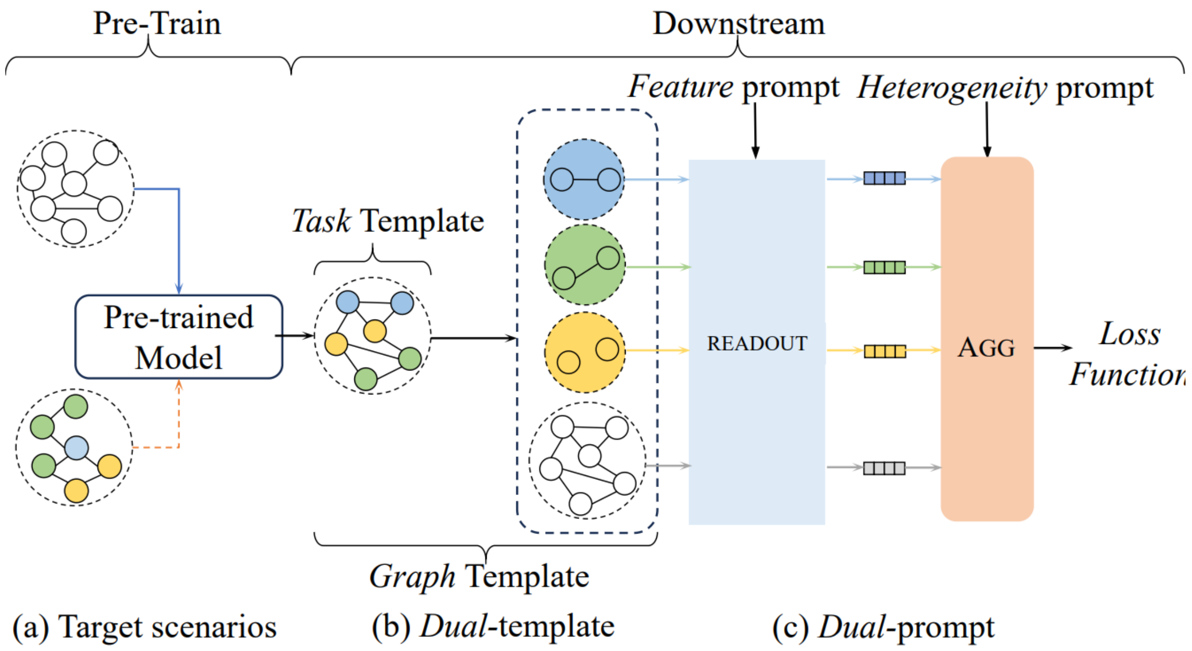
HGPrompt: Extending to heterogeneous graphs

Two challenges

- Gap between homogeneous and heterogeneous graph
- Different downstream tasks focus on heterogeneous aspect

Insights

- Dual-template:
  Task + Graph template
- Dual-prompt:
  Feature + Heterogeneity prompt



[AAAI24] Xingtong Yu, Yuan Fang, Zemin Liu and Xinming Zhang. *HGPrompt: Bridging Homogeneous and Heterogeneous Graphs for Few-shot Prompt Learning.* https://arxiv.org/pdf/2312.01878.pdf

Outline

- Introduction: Data and problems
- Overview of few-shot methodologies
- Can prompt work on graph + text?
- Can prompt work on graph alone?
- **Conclusion**

Conclusion

- Few-shot learning on graphs: different kinds of graphs/tasks
- Learning and transferring/using prior is the key
- Prompt is a promising paradigm…

Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S. Yu, Chuan Shi. **Towards Graph Foundation Models: A Survey and Beyond.** https://arxiv.org/pdf/2310.11829.pdf

Acknowledgement

Student/post-doc co-authors


Zemin Liu


Chenghao Liu


Zhihao Wen


Xingtong Yu


Deyu Bo

Main collaborators

Prof. Steven Hoi, Singapore Management University
Prof. Chuan Shi, Beijing University of Posts and Telecommunications
Prof. Xinming Zhang, University of Science and Technology of China

Main funding sources

- *One-shot learning: A crucial learning paradigm towards human-like learning.* National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG-RP-2018-001).
- *Learning with less data*. Agency for Science, Technology and Research (A*STAR) under its AME Programmatic Funds (Grant No. A20H6b0151).
- *Universal pre-training of graph neural networks*. Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (Proposal ID: T2EP20122-0041).
- Lee Kong Chian Fellowship, 2021, Singapore Management University.

Thank you

Questions?

Email: yfang@smu.edu.sg

Full publications, codes and data are available at
http://www.yfang.site/