PRODIGY: Enabling In-context Learning Over Graphs

Qian Huang[1*]
qhwang@cs.stanford.edu

Hongyu Ren[1*]
hyren@cs.stanford.edu

Peng Chen[1]
pengc@stanford.edu

Gregor Kržmanc[2]
gregor.krzmanc@ijs.si

Daniel Zeng[1]
dzeng@cs.stanford.edu

Percy Liang[1]
pliang@cs.stanford.edu

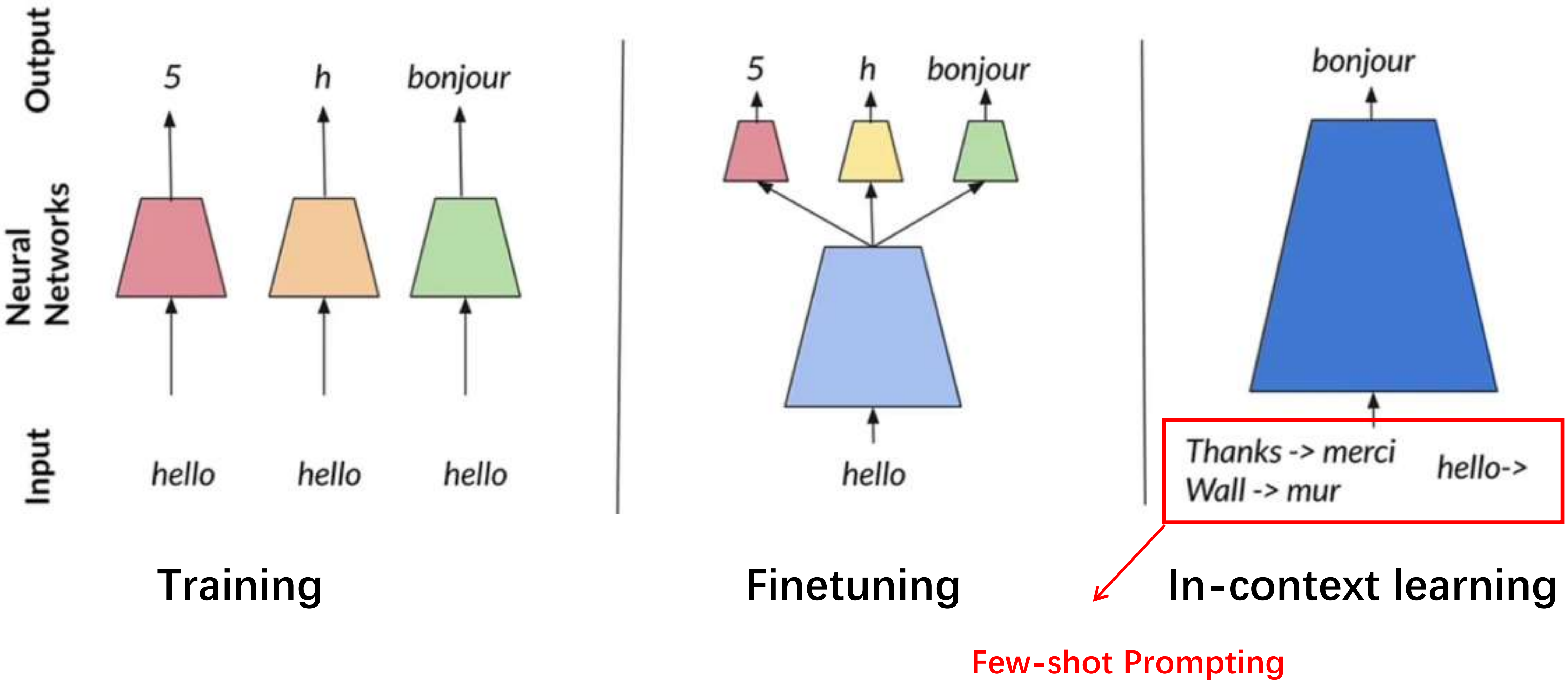
Jure Leskovec[1]
jure@cs.stanford.edu

[1] Stanford University
[2] University of Ljubljana

Conference: NeurIPS 2023

Year: 2023

Different Learning Paradigms



In-context Learning Over Graphs

Thanks -> merci
Wall -> mur

hello->?

Few-shot Prompting over text

What is In-context learning over graphs?

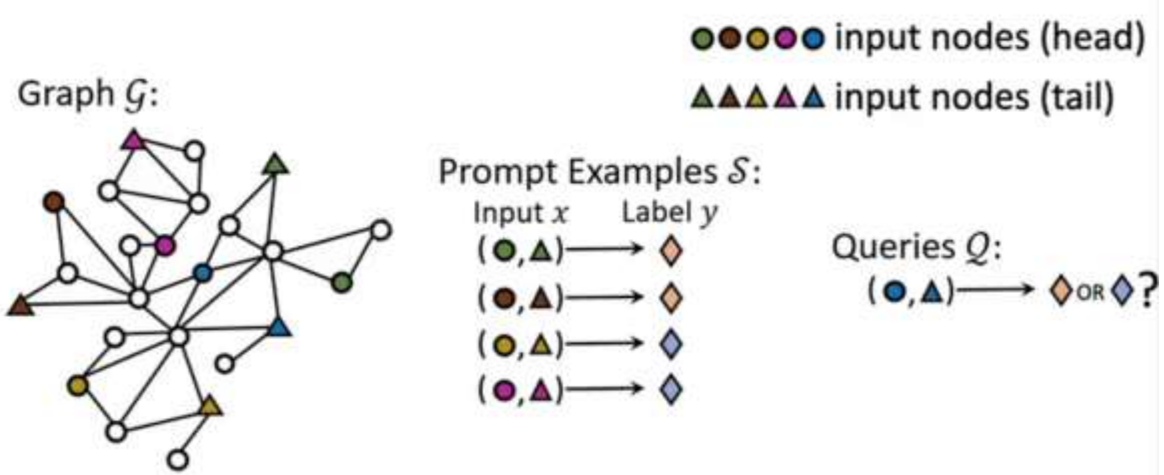
In-context Learning Over Graphs :Link Prediction Example

Thanks -> merci
Wall -> mur

hello->?

different tasks

Few-shot Prompting over text



different graphs

Few-shot Prompting over graph (for link classification)

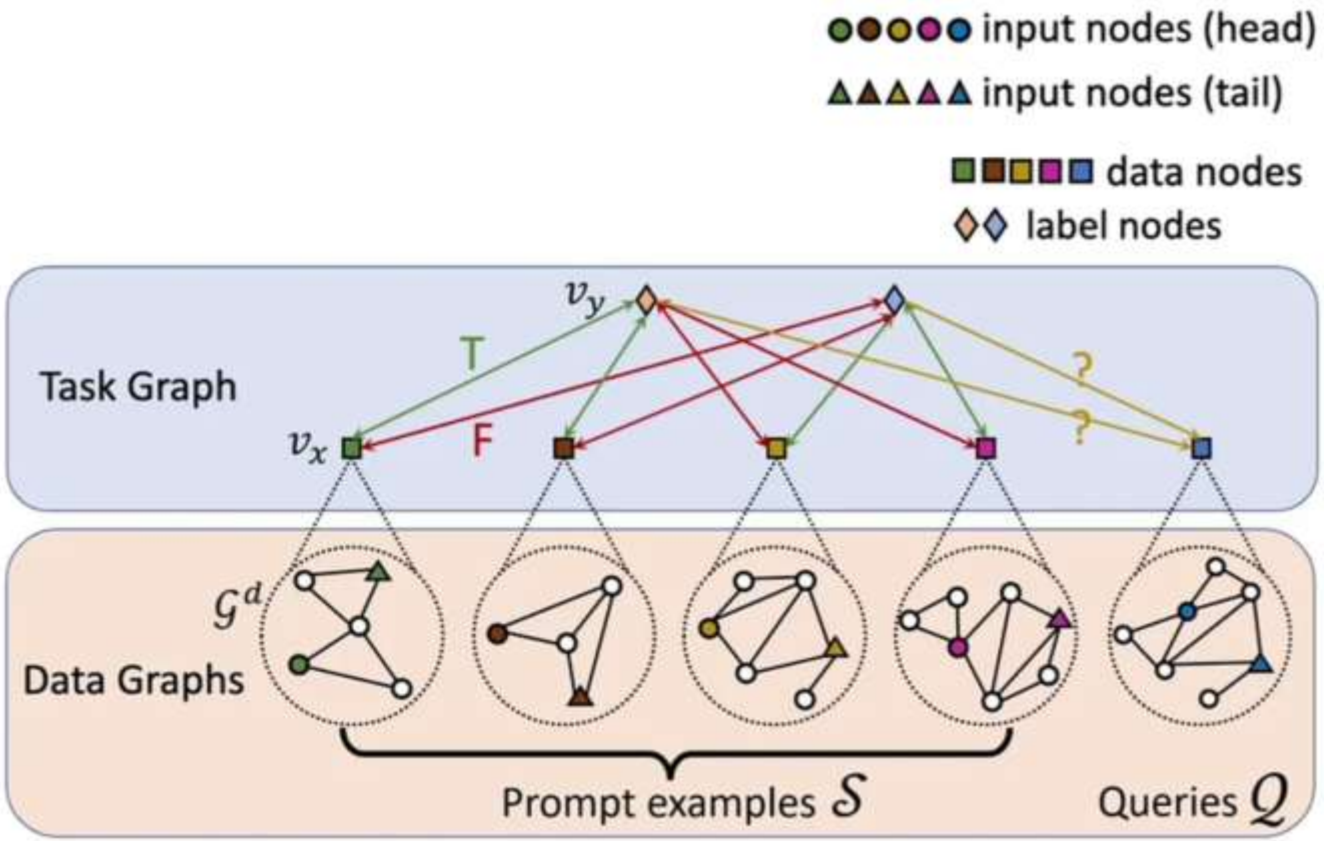
But,how to achieve this?Two Challenges:

1.How to formulate and represent node-, edge- and graph-level tasks over graphs with a unified task representation that allows the model to solve diverse tasks without the need for retraining or parameter tuning.

2. How to design model architecture and **pretraining** objectives that enable models to achieve in-context learning capability across diverse tasks and diverse graphs in the unified task representation.

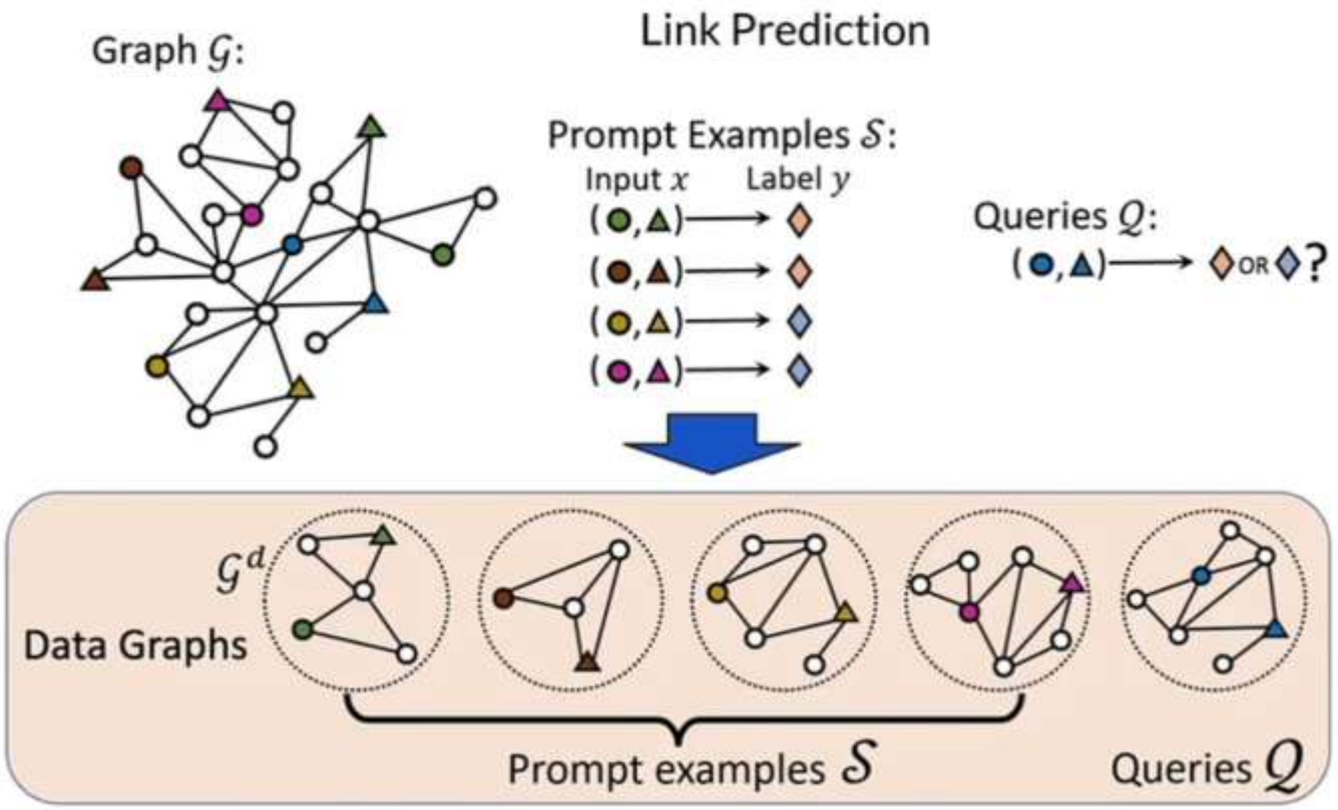
Prompt Graph

Prompt Graph is a unified representation of few-shot prompts over graph for diverse tasks



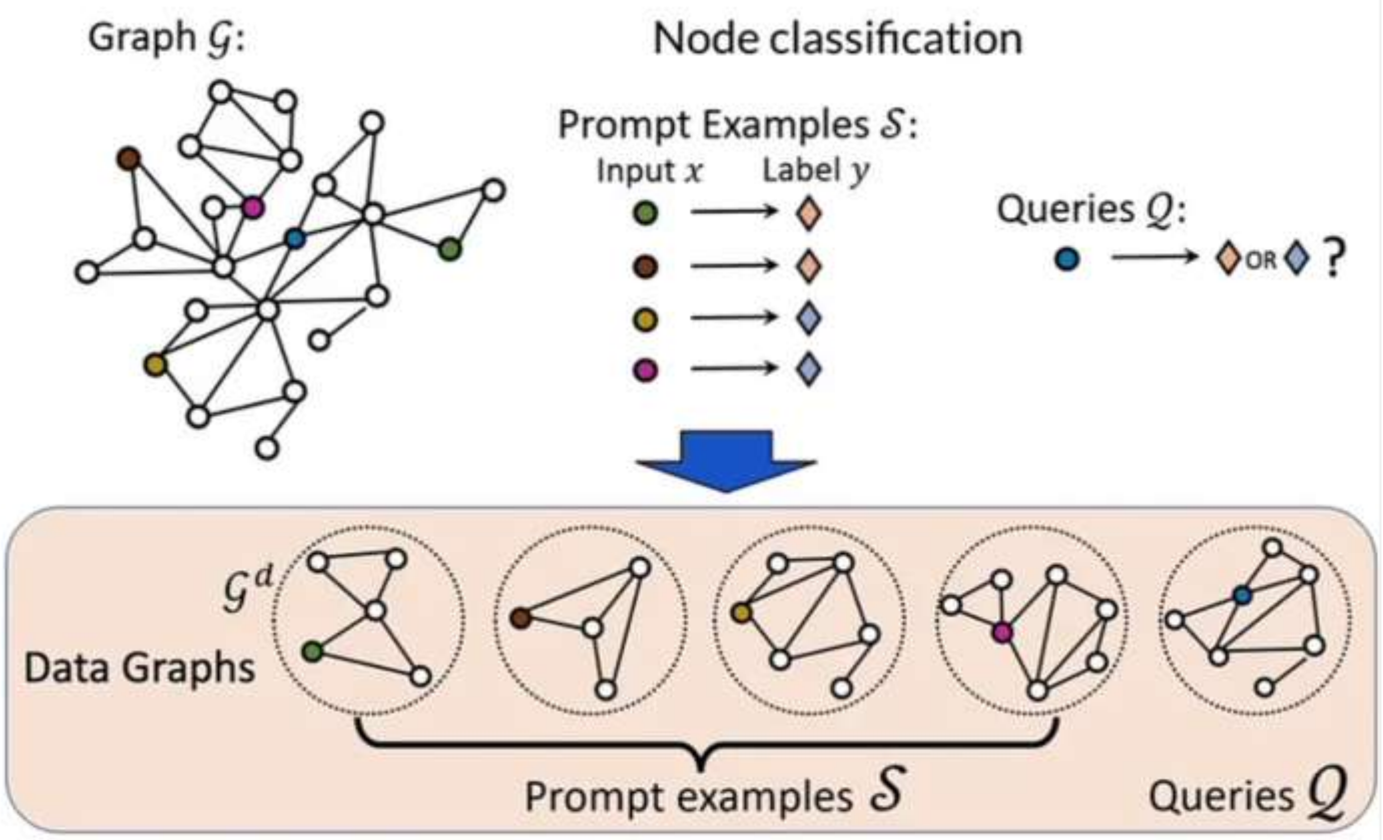
Step1:Data Graph - Link Prediction

Data Graph contextualizes each input ***x*** in the graph **G** (e.g. by **subgraph extraction**)



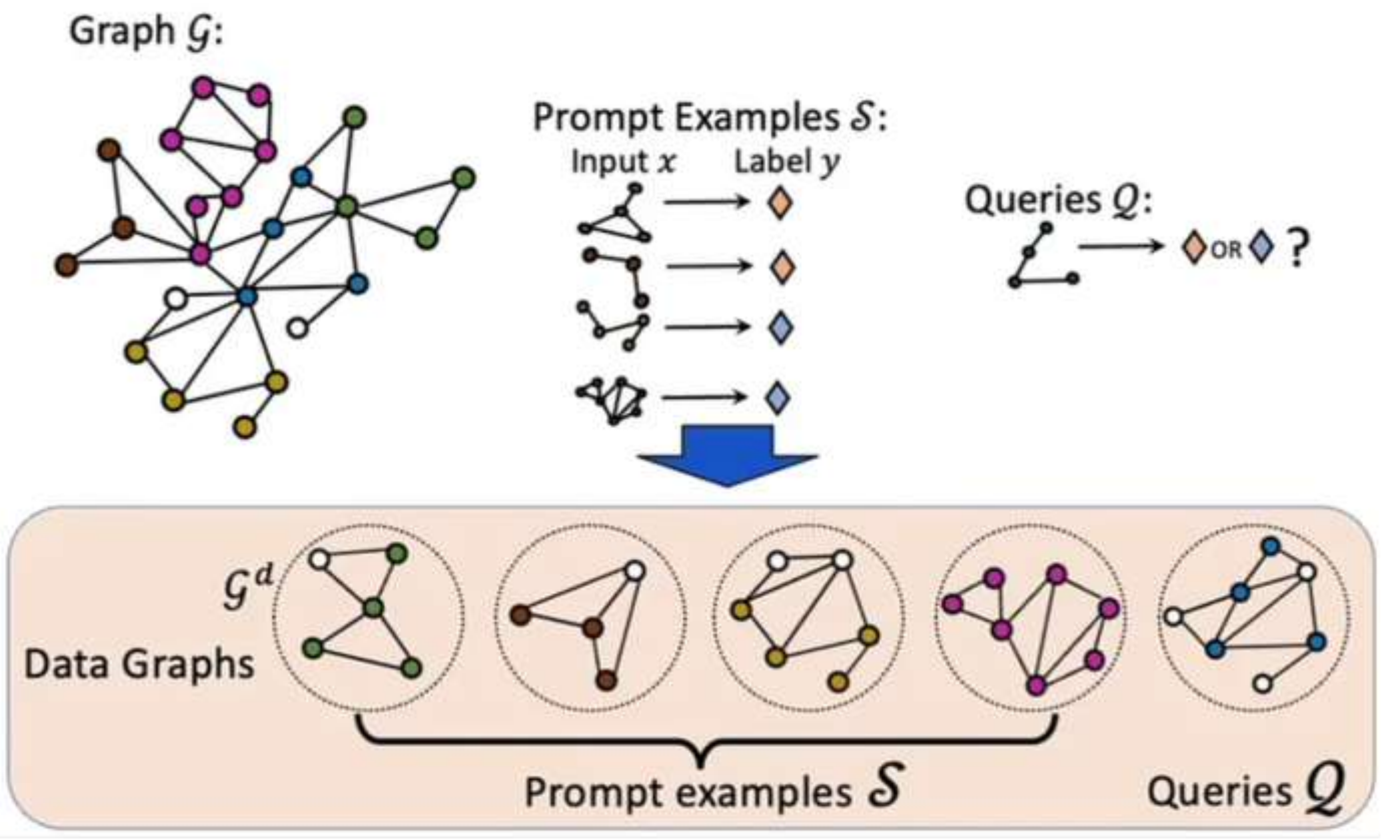
Step1:Data Graph - Node Classification

Data Graph contextualizes each input ***x*** in the graph **G** (e.g. by **subgraph extraction**)



Step1:Data Graph - Graph Classification

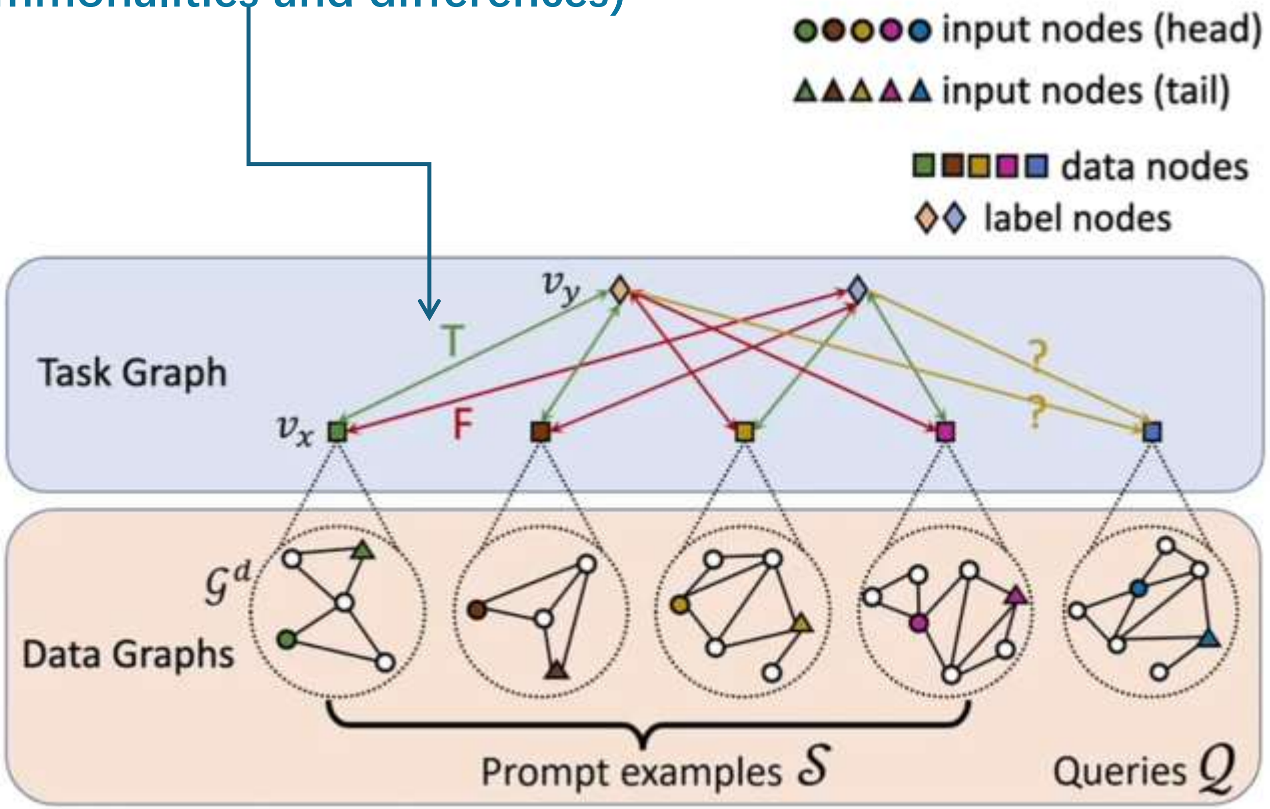
Data Graph
contextualizes each input
x in the graph **G**
(e.g. by **subgraph extraction**)



Step2:Task Graph

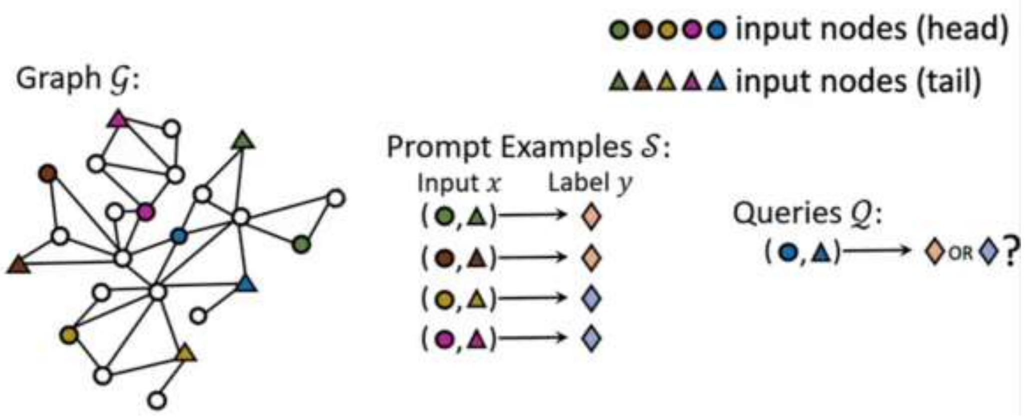
Task Graph
interconnects inputs and labels across examples to form context for queries

Reflects what is the task using examples (commonalities and differences)

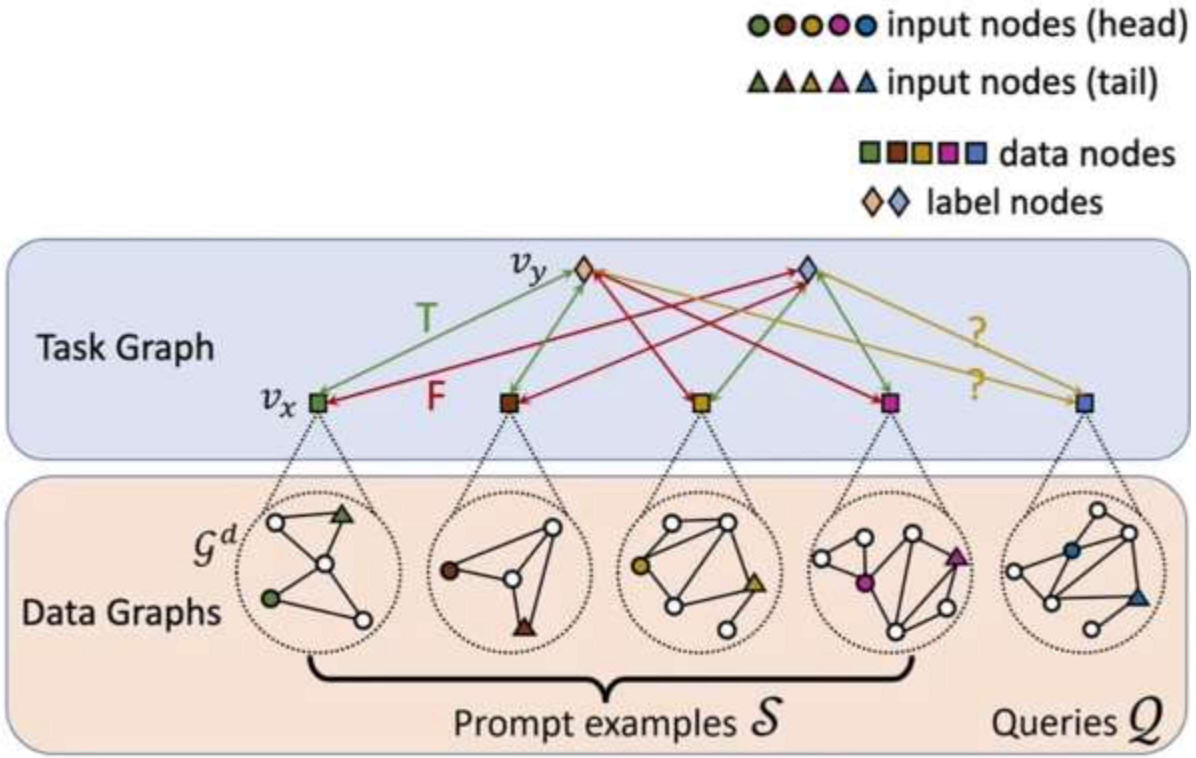


Step2:Task Graph

How to use PromptGraph for in-context learning?



In-context learning over graph



inductive **link prediction** over hierarchical graph

Pretraining to Enable In-context Learning

framework:

model architecture over prompt graph

in-context pretraining objectives

1.Data graph Message Passing
2.Task graph Message Passing

In-context Pretraining Objectives

Two stages:

pretraining task generation

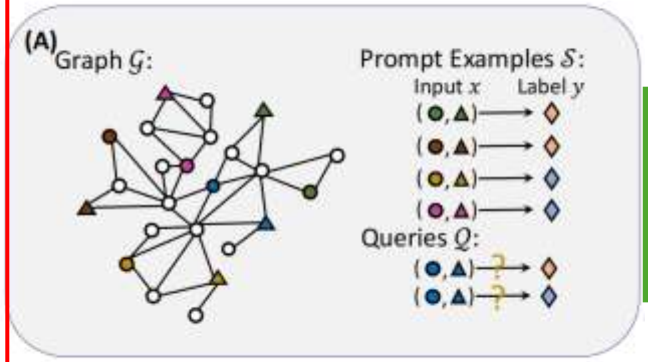
convert to PromptGraph

Pretraining Data/Graph

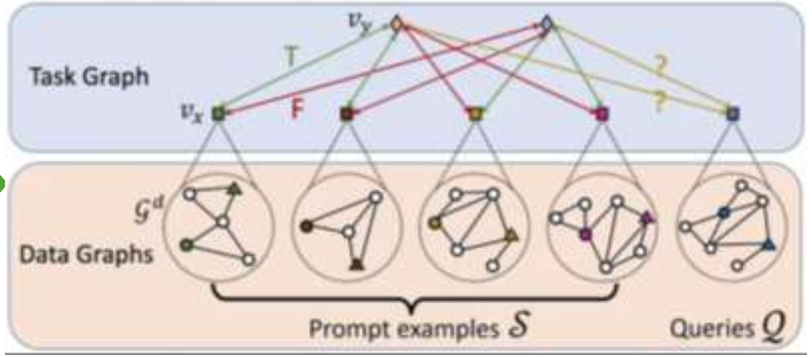
Neighbor Matching

\+

MultiTask



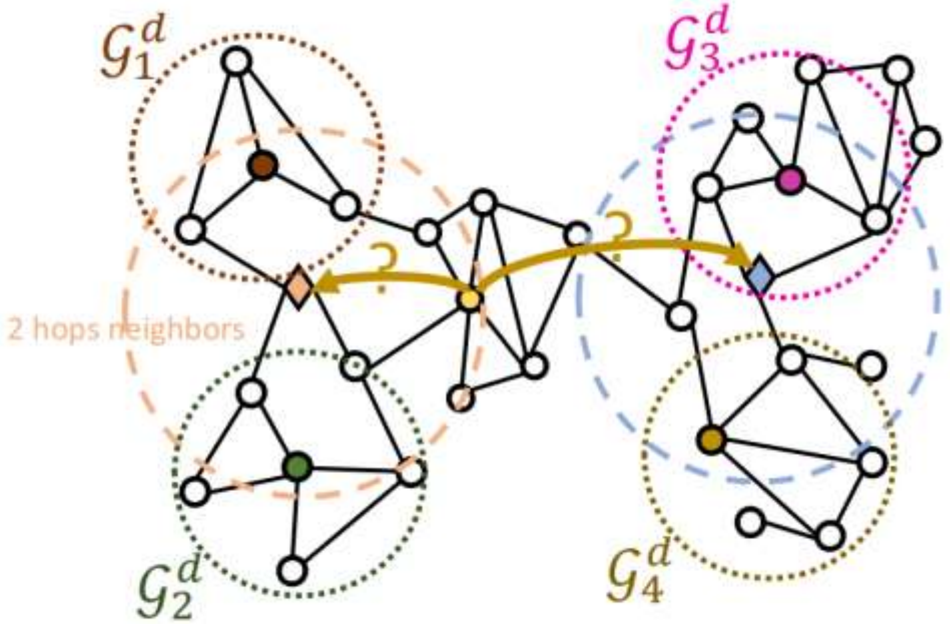
PromptGraph with Augmentation



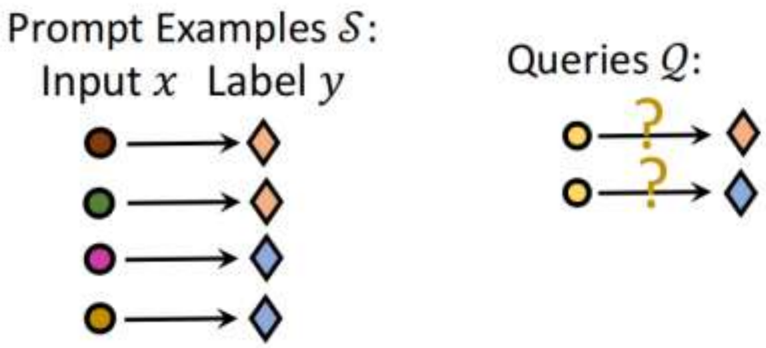
1.node dropping
2.node feature masking

Self-supervised Task Example:Neighbor Matching

Idea:the task is to classify which neighborhood a node is in,where each neighborhood is defined by other nodes in it.



$$(\mathcal{G}_{\texttt{pretrain}}, \mathcal{S}_{\text{NM}}, \mathcal{Q}_{\text{NM}}) \sim \text{NM}_{k,m}(\mathcal{G}_{\texttt{pretrain}})$$



Multi-task: $(\mathcal{G}_{\texttt{pretrain}}, \mathcal{S}_{\text{MT}}, \mathcal{Q}_{\text{MT}}) \sim \text{MT}_{k,m}(\mathcal{G}_{\texttt{pretrain}}, f)$

In-context Learning Results

Paper category classification:pretrain on MAG->in-context learning on Arxiv

Classes	NoPretrain	Contrastive	PG-NM	PG-MT	PRODIGY	Finetune
3	33.16 $\pm$0.30	65.08 $\pm$0.34	72.50 $\pm$0.35	65.64 $\pm$0.33	**73.09** $\pm$**0.36**	65.42 $\pm$5.53
5	18.33 $\pm$0.21	51.63 $\pm$0.29	61.21 $\pm$0.28	51.97 $\pm$0.27	**61.52** $\pm$**0.28**	53.49 $\pm$4.61
10	9.19 $\pm$0.11	36.78 $\pm$0.19	46.12 $\pm$0.19	37.23 $\pm$0.20	**46.74** $\pm$**0.20**	30.22 $\pm$3.77
20	4.72 $\pm$0.06	25.18 $\pm$0.11	33.71 $\pm$0.12	25.91 $\pm$0.12	**34.41** $\pm$**0.12**	17.68 $\pm$1.15
40	2.62 $\pm$0.02	17.02 $\pm$0.07	23.69 $\pm$0.06	17.19 $\pm$0.08	**25.13** $\pm$**0.07**	8.04 $\pm$3.00

Result1:PRODIGY improves performance in all cases,up to 48% over Contrastive.

Result2:PRODIGY can induce strong in-context learning with very different self-supervised pretraining tasks.

Pretraining has no idea about classifying paper categorie!

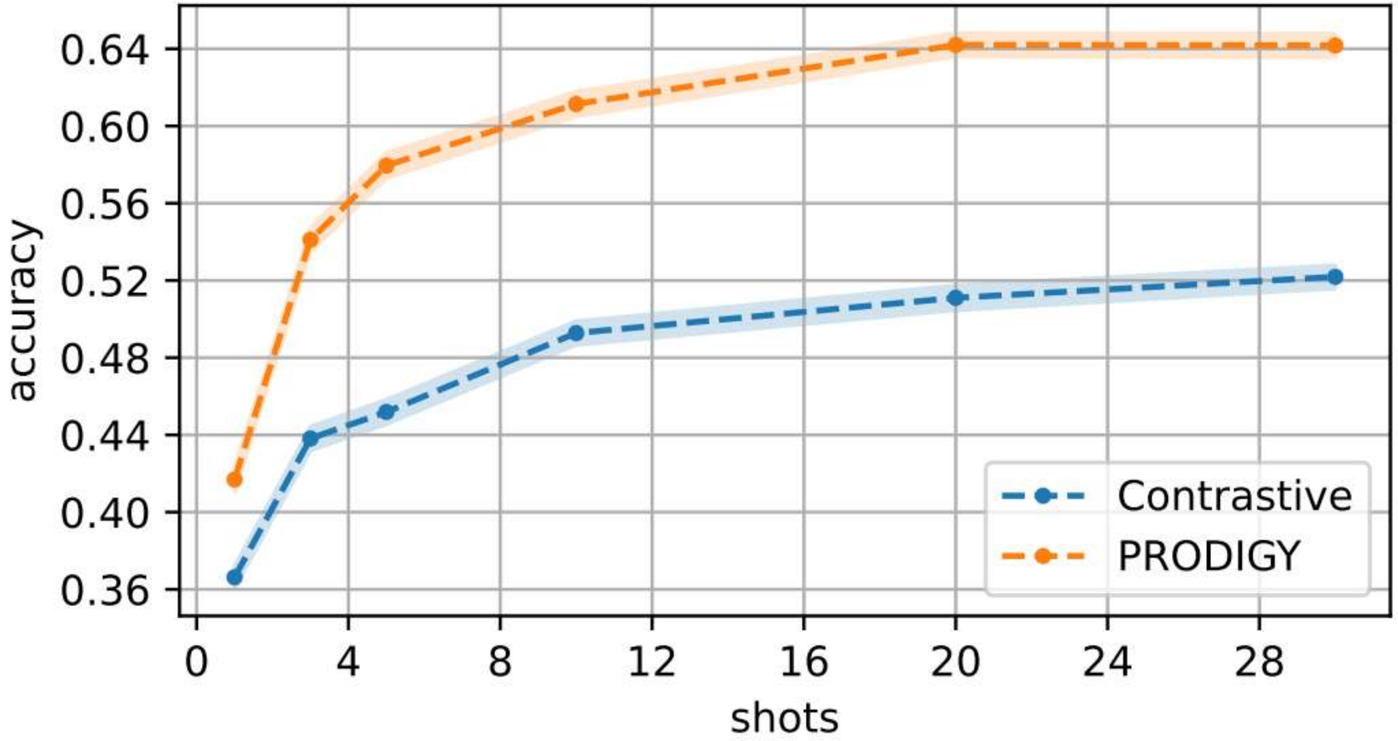
In-context Learning Results

KG Completion:pretrain on Wiki KG->in-context learning on ConceptNet,FB 15K-237,NELL

Classes	NoPretrain	Contrastive	PG-NM	PG-MT	PRODIGY	Finetune
4	30.4 $\pm$ 0.63	44.01 $\pm$ 0.61	46.94 $\pm$ 0.61	51.78 $\pm$ 0.63	**53.97** $\pm$ **0.63**	53.85 $\pm$ 9.29
5	33.54 $\pm$ 0.61	81.35 $\pm$ 0.58	80.35 $\pm$ 0.57	89.15 $\pm$ 0.46	**88.02** $\pm$ **0.48**	82.01 $\pm$ 12.83
10	20.0 $\pm$ 0.35	70.88 $\pm$ 0.48	71.68 $\pm$ 0.45	82.26 $\pm$ 0.40	**81.1** $\pm$ **0.39**	71.97 $\pm$ 6.16
20	9.2 $\pm$ 0.18	59.8 $\pm$ 0.35	59.9 $\pm$ 0.35	73.47 $\pm$ 0.32	**72.04** $\pm$ **0.33**	64.01 $\pm$ 4.66
40	2.5 $\pm$ 0.08	49.39 $\pm$ 0.23	46.82 $\pm$ 0.21	58.34 $\pm$ 0.22	**59.58** $\pm$ **0.22**	57.27 $\pm$ 3.33
5	20.95 $\pm$ 0.52	83.38 $\pm$ 0.5	82.39 $\pm$ 0.53	85.26 $\pm$ 0.48	**88.09** $\pm$ **0.43**	87.22 $\pm$ 12.75
10	11.0 $\pm$ 0.26	74.54 $\pm$ 0.46	75.14 $\pm$ 0.43	78.15 $\pm$ 0.41	**82.47** $\pm$ **0.39**	71.90 $\pm$ 5.90
20	5.34 $\pm$ 0.13	65.68 $\pm$ 0.34	65.68 $\pm$ 0.34	68.38 $\pm$ 0.33	**74.72** $\pm$ **0.31**	66.19 $\pm$ 8.46
40	2.5 $\pm$ 0.06	56.7 $\pm$ 0.23	54.91 $\pm$ 0.22	51.24 $\pm$ 0.25	**60.04** $\pm$ **0.23**	55.06 $\pm$ 4.19

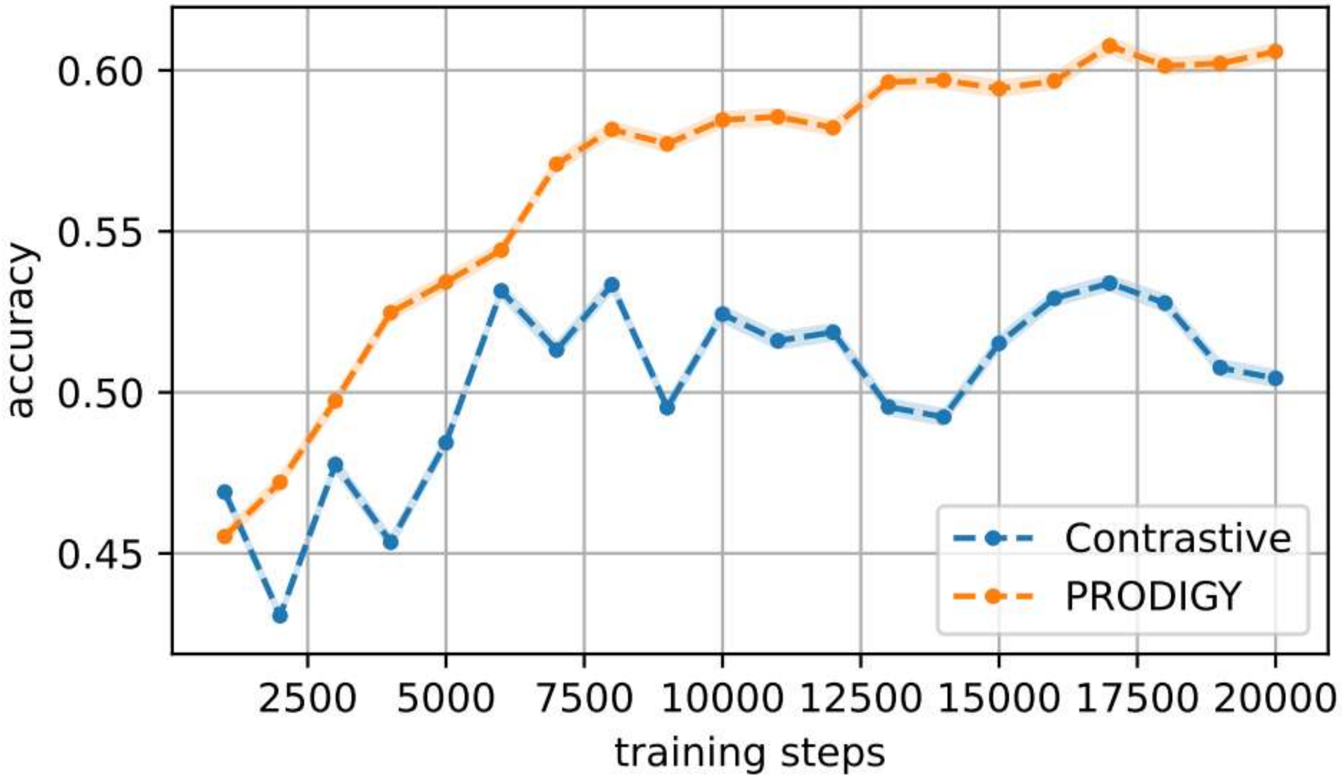
Performance Scaling with More Shots

ConceptNet -4 way



Pretraining with only 3 shots=> model can learn from context from much more than 3 shots

Data Scaling



Hard and diverse pretraining tasks generation pipeline allows the model to keep scaling with more training data