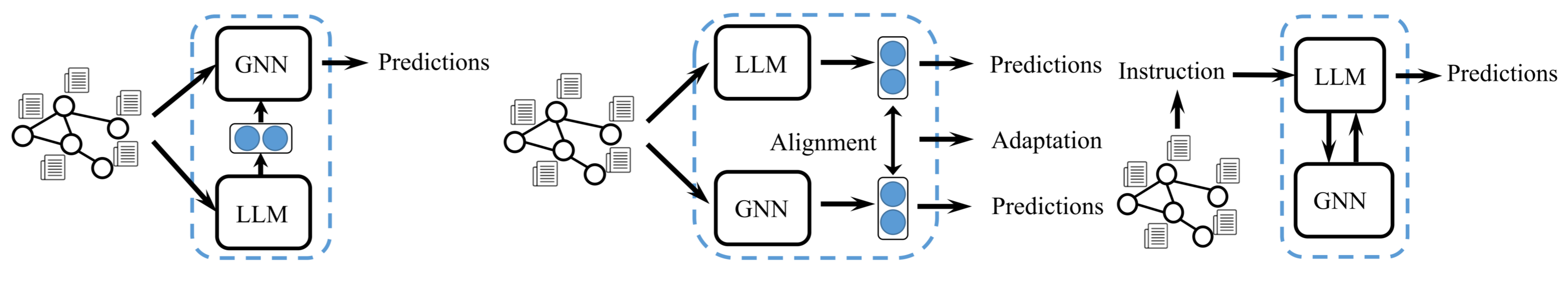
NAACL 2025

Exploring the Potential of Large Language Models for Heterophilic Graphs

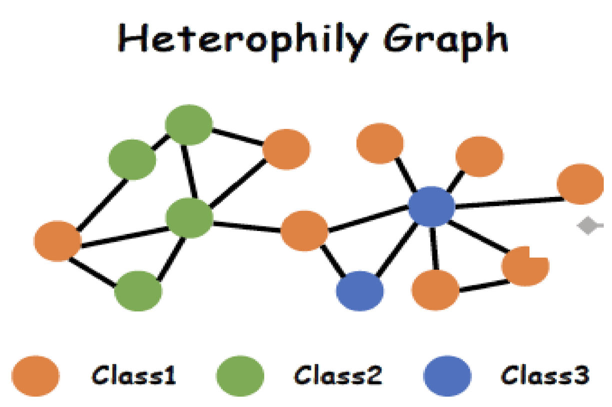
Yuxia Wu* — Singapore Management University
Shujie Li* — Beijing University of Post and Telecommunication
Yuan Fang — Singapore Management University
Chuan Shi — Beijing University of Post and Telecommunication

Motivation

❑ LLM achieve success on graph learning tasks. However, LLM for heterophilic graphs is largely unexplored.



(a) GNN-centric methods. (b) Symmetric methods. (c) LLM-centric methods.

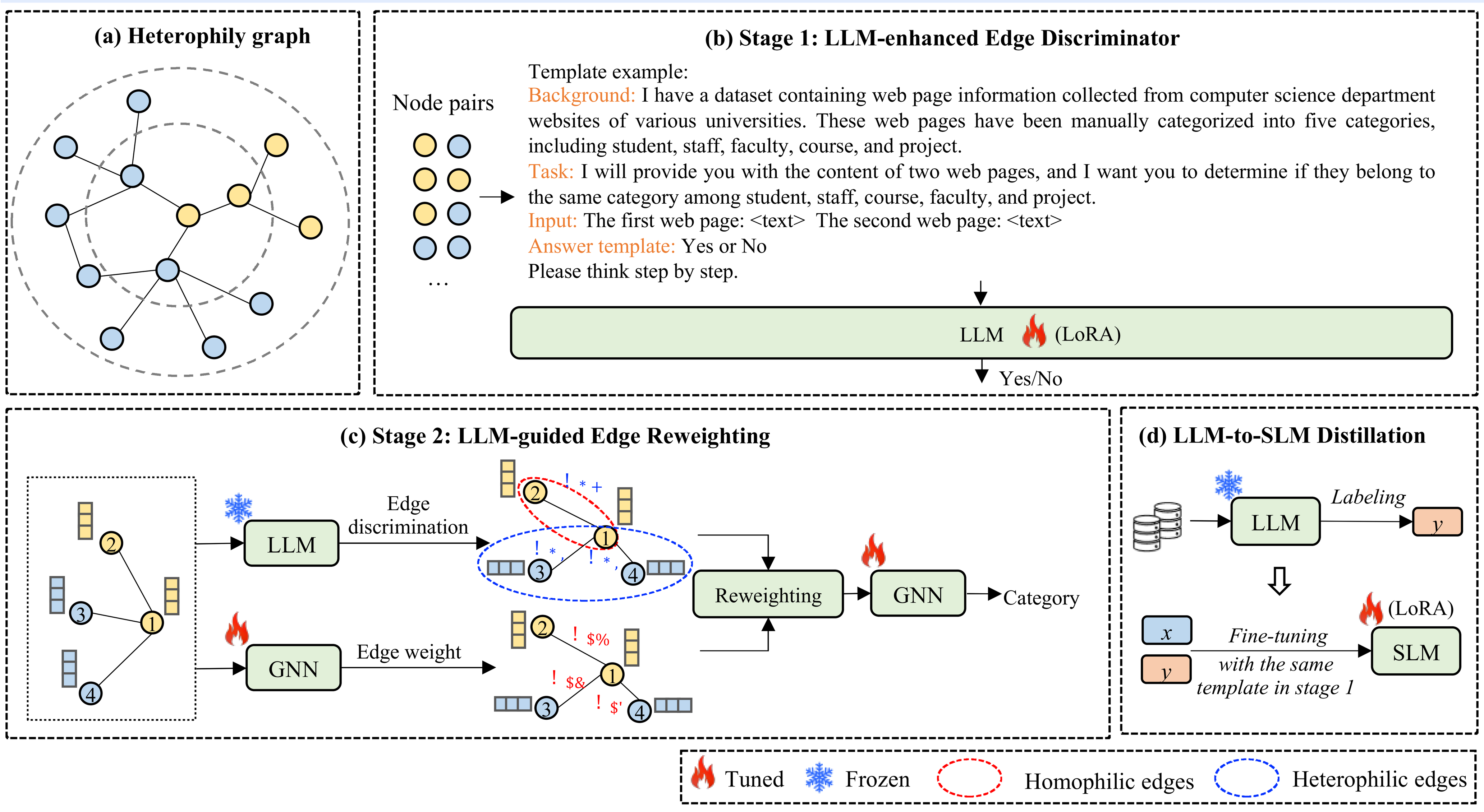


❑ Heterophily-specific GNNs overlook the rich textual content associate with the nodes (bag-of-words, shallow embedding)

Contribution:

❖ We are the first exploration of the LLMs for heterophilic graphs
❖ We propose a two-stage framework including an LLM-enhanced edge discriminator and an LLM-guided edge reweighting.
❖ We applied model distillation to create smaller models with faster inference and competitive performance

Framework:LLM4HeG



Stage 1: LLM-enhanced Edge Discriminator

- Construct the ground truth labels from the training set.
- Design a language template to describe the task of edge discrimination.
- Parameter-efficient fine-tuning LLM: LoRA

Stage 2: LLM-guided Edge Reweighting

❑ Edge weight from LLM:

$$w_{uv}^{\mathrm{LLM}} = \begin{cases} \tanh(w_{\mathrm{Ho}}) & \text{if } O_{\mathrm{LLM}}(u,v) = \mathit{Yes}, \\ \tanh(w_{\mathrm{He}}) & \text{if } O_{\mathrm{LLM}}(u,v) = \mathit{No}, \end{cases}$$

Learnable parameters for different edge types

❑ Reweighting:

$$w_{uv} = \frac{1}{2}\left(w_{uv}^{\mathrm{LLM}} + w_{uv}^{\mathrm{G}}\right).$$

Various GNN models for heterophilic graph

FAGCN: $w_{uv}^{\mathrm{G}} = \tanh\left(\mathbf{g}^{\top}[\mathbf{h}_u \parallel \mathbf{h}_v]\right),$

❑ GNN prediction:

Initial features from LLM

$$\mathbf{h}_v^{(0)} = \sigma(\mathrm{LLM}(x_v)\mathbf{W}_e),$$

$$\mathbf{h}_v^{(l)} = \epsilon \mathbf{h}_v^{(0)} + \sum_{u \in \mathcal{N}_i(v)} \frac{w_{uv}}{\sqrt{d_u d_v}} \mathbf{h}_u^{(l-1)},$$

$$\mathbf{h}_{\mathrm{out}} = \mathbf{W}_o \mathbf{h}_v^{(L)},$$

LLM-to-SLM Distillation

- Teacher model : fine-tuned LLM in Stage 1
- Expand label set: Pseudo-labels for additional node pairs + training set
- Fine-tune small language model (SLM) and inference

Experiment

Datasets

Dataset	Classes	Nodes	Edges	$\mathcal{H}(G)$
Cornell	5	195	304	0.13
Texas	5	187	328	0.12
Wisconsin	5	265	530	0.20
Actor	5	4,416	12,172	0.56
Amazon	5	24,492	93,050	0.38

Table 1: Dataset statistics.

Results

Methods	Cornell	Texas	Wisconsin	Actor	Amazon
Classic GNNs					
GCN	52.86±1.8	43.64±3.3	41.40±1.8	66.70±1.3	39.33±1.0
GraphSAGE	75.71±1.8	81.82±2.5	80.35±1.3	70.37±0.1	46.63±0.1
GAT	54.28±5.1	51.36±2.3	50.53±1.7	63.74±6.7	35.12±6.4
Heterophily-specific GNNs					
H2GCN	69.76±3.0	79.09±3.5	80.18±1.9	70.73±0.9	47.09±0.3
FAGCN	76.43±3.1	84.55±4.8	83.16±1.4	75.58±0.5	49.83±0.6
JacobiConv	73.57±4.3	81.80±4.1	76.31±11.3	73.81±0.3	49.43±0.5
GBK-GNN	66.19±2.8	80.00±3.0	72.98±3.3	72.49±1.0	44.90±0.3
OGNN	71.91±1.8	85.00±2.3	79.30±2.1	72.08±2.4	47.79±1.6
SEGSL	66.67±4.1	85.00±2.0	79.30±1.8	72.73±0.8	47.38±0.2
DisamGCL	50.48±2.0	65.00±1.2	57.89±0.0	67.78±0.3	43.90±0.4
LLM4HeG (fine-tuned LLM/SLMs and distilled SLMs)					
Vicuna 7B	**77.62**±2.9	**89.09**±3.3	86.14±2.1	**76.82**±0.5	51.53±0.4
Bloom 560M	75.48±2.1	80.00±4.0	86.49±1.9	76.16±0.6	51.52±0.5
Bloom 1B	75.71±1.4	83.86±2.8	83.86±1.7	74.99±0.5	**52.33**±0.6
7B-to-560M	75.00±4.0	88.18±2.2	**87.19**±2.5	75.78±0.2	51.51±0.4
7B-to-1B	77.38±2.7	88.18±4.0	86.14±1.5	75.37±0.9	51.58±0.4

Table 2: Accuracy for node classification of different methods. (Best results bolded; runners-up underlined.)

We use the initial node features derived from the Vicuna 7B model for all methods.

- Heterophily-specific GNNs generally outperform classic GNNs
- Our methods consistently achieve the best. Fine-tuned LLM > Fine-tuned SLMs
- Fine-tuned LLM ~= Distilled SLMs

FANG Lab@SMU: https://www.yfang.site/group Email: yieshah2017@gmail.com Homepage: https://yuxiawu.github.io/