GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

The Instability of GFM Fine-tuning

- Observations from SOTA GFMs: **huge variances** under few-shot fine-tuning

42.26 ± 10.18	**42.40 ± 9.26**	**49.82 ± 8.38**	**64.82 ± 10.53**	**49.77 ± 11.00**	**67.80 ± 6.84**
48.36 ± 11.34	**44.28 ± 10.16**	**54.34 ± 9.76**	**64.08 ± 9.85**	**48.29 ± 11.35**	**68.92 ± 7.32**
47.80 ± 11.88	36.38 ± 9.10	**50.25 ± 10.43**	**58.71 ± 8.69**	**48.22 ± 8.17**	**42.70 ± 8.73**	**33.36 ± 8.11**
55.35 ± 13.62	38.75 ± 9.40	48.69 ± 10.16	**58.75 ± 11.67**	**48.72 ± 11.18**	**43.71 ± 9.54**	**48.28 ± 9.72**
34.23±8.16	39.05±8.82	44.85±6.72	34.02±11.94	22.46±1.96	24.61±3.99	50.79±0.65
39.54±9.02	39.24±8.95	45.39±11.01	33.58±10.38	22.35±3.77	23.68±1.56	50.78±3.05
44.83±7.41	**42.18±6.41**	**46.84±7.31**	**40.77±5.96**	**24.30±3.26**	**28.36±3.65**	**52.36±0.86**
64.56±7.29	**61.24±4.82**	**63.50±5.81**	**49.56±6.92**	**23.00±4.39**	**30.54±2.87**	**53.58±0.83**

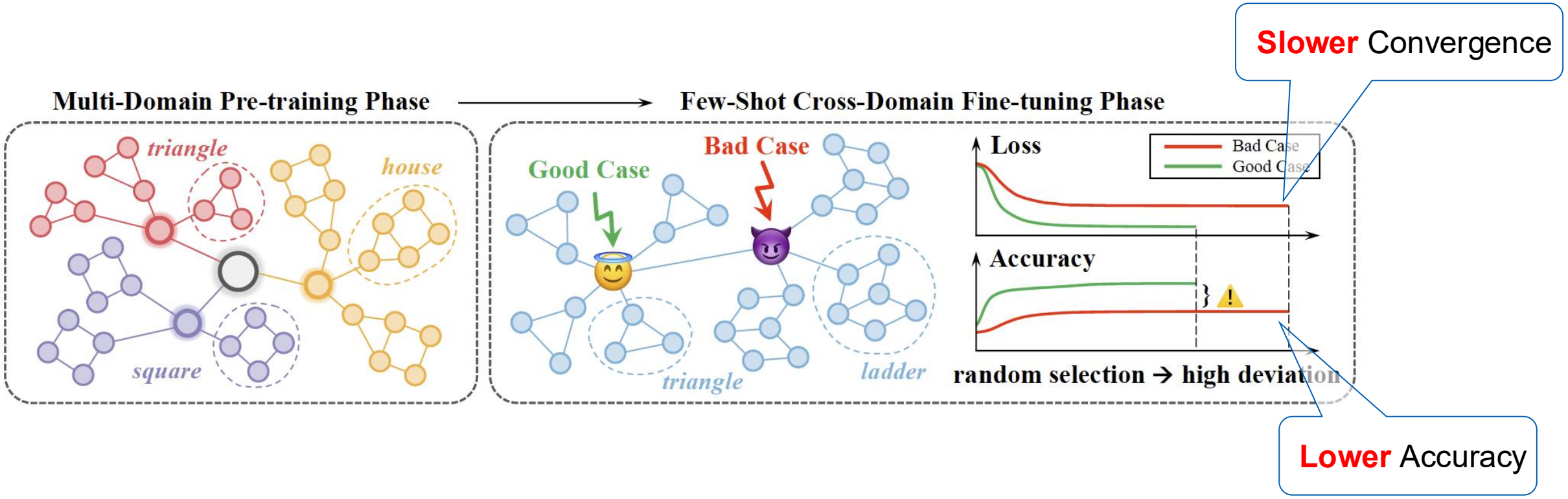
- **Why?** Random few-shot sample selection brings fine-tuning uncertainty.

Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

The Instability of GFM Fine-tuning

- How does fine-tuning instability happen?

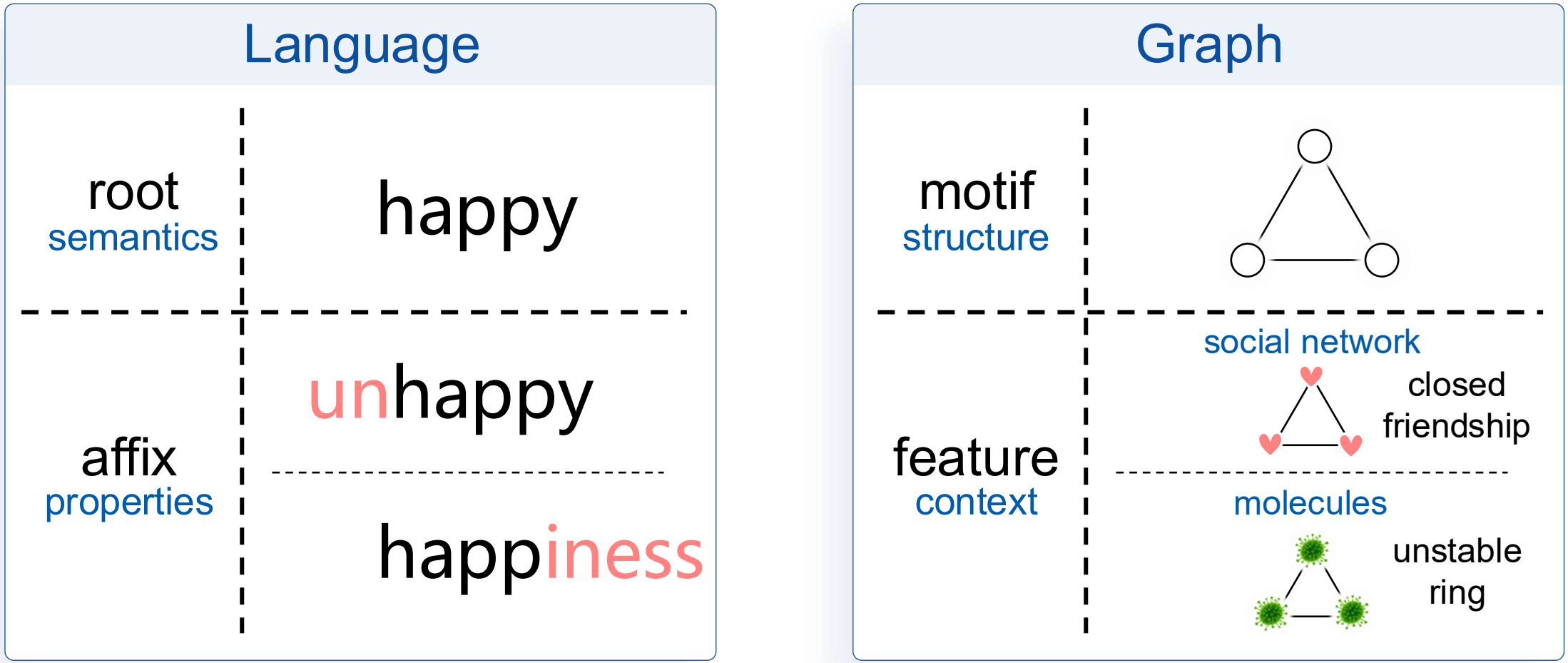


- **Challenge:** **mismatch** between support samples and pre-trained graph patterns.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Key Insight: Tokenize Graphs with Vocabulary

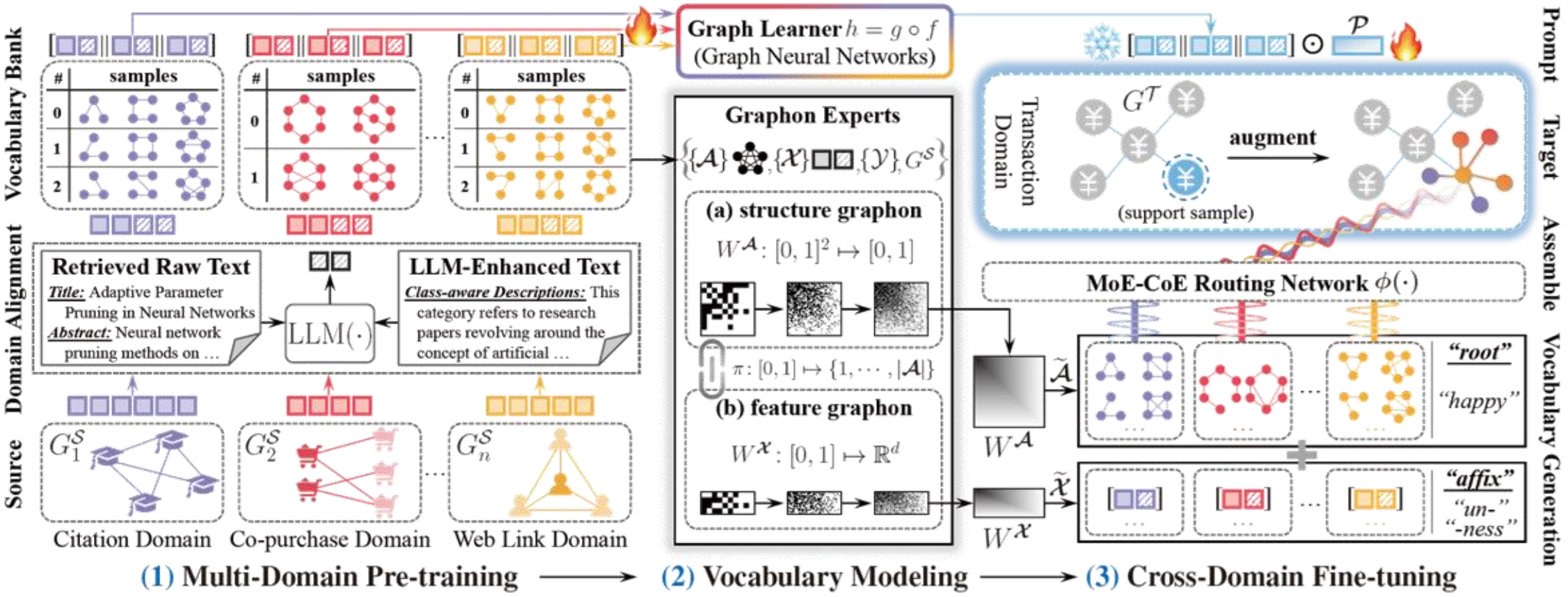
- Understanding graph data through the lens of language modeling



Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

- **GRAVER:** Pretrain-then-Finetune ➔ Pretrain-then-**Augment**-then-Finetune



Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

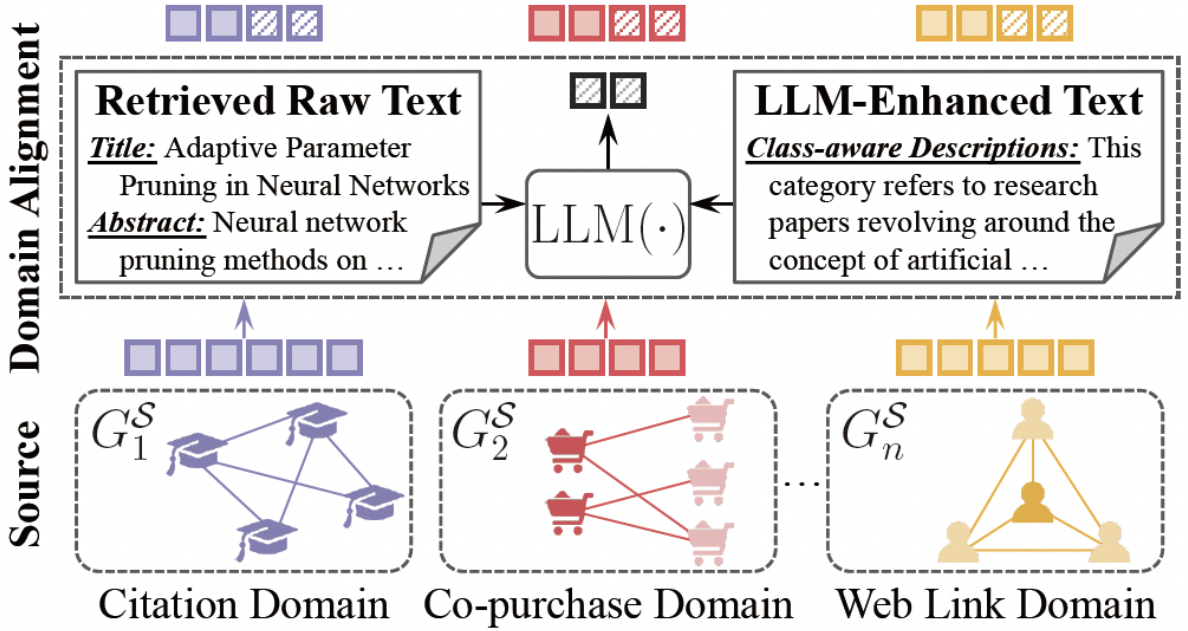
GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Phase 1: Pre-training with Transferable Graph Vocabulary

- **Goal:** identify subgraph (graph vocabulary) that generalize across tasks and domains.

- **Step 1:** Multi-Domain Alignment

$$\widehat{\mathbf{X}}_i^{\mathcal{S}} \in \mathbb{R}^{N_i \times d} = \mathbf{W}_i^{\top} \left(\mathcal{F}\left(\mathbf{X}_i^{\mathcal{S}} \middle\| \mathrm{LLM}\left(\mathbf{X}_i^{\mathcal{S}} \right) \right) \right), \quad \text{for all } G_i^{\mathcal{S}} \in \{G^{\mathcal{S}}\}$$



Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

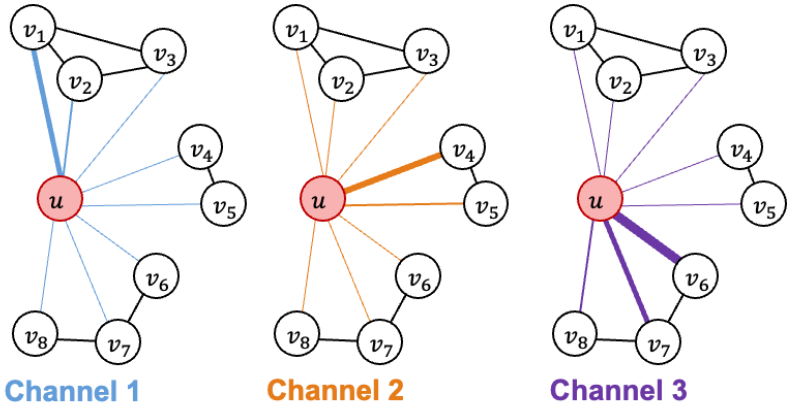
GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Phase 1: Pre-training with Transferable Graph Vocabulary

- **Goal:** identify subgraph (graph vocabulary) that generalize across tasks and domains.

- **Step 2:** Factor-aware Ego-Graph Disentanglement
  - multi-channel encoder:

$$f_{\mathbf{\Theta}}\colon \mathbf{g}_u^{\mathcal{S}} \mapsto \{\mathbf{h}_{u,k}^{\mathcal{S}} \in \mathbb{R}^d\}_{k=1}^{K}$$



  - neighbor attention:

$$\alpha_{v\to k}^{(t)} \propto \mathrm{Softmax}_k\left(\langle \mathbf{h}_{u,k}^{\mathcal{S}(t)}, \mathbf{h}_{v,k}^{\mathcal{S}(t)} \rangle / \tau\right), \quad \text{s.t.} \sum_{k=1}^{K} \alpha_{v\to k}^{(t)} = 1$$

  - updating:

$$\mathbf{h}_u^{\mathcal{S}(T)} = \Big\|_{k=1}^{K} \mathbf{h}_{u,k}^{\mathcal{S}(T)}, \quad \mathbf{h}_{u,k}^{\mathcal{S}(t+1)} := \mathbf{h}_{u,k}^{\mathcal{S}(t)} + \sum\nolimits_{v\in\mathcal{V}_{u,k}^{(t)}} \left[\alpha_{v\to k}^{(t)} \mathbf{h}_{v,k}^{\mathcal{S}(t)}\right]$$

Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Phase 1: Pre-training with Transferable Graph Vocabulary

- **Goal:** identify subgraph (graph vocabulary) that generalize across tasks and domains.

- **Step 3:** Semantic-Independence Promotion
  - regularizer: $\mathcal{R}_{\mathrm{MI}}^{u} = \sum_{i \neq j} I\left(\mathbf{h}_{u,i}^{\mathcal{S}}; \mathbf{h}_{u,j}^{\mathcal{S}}\right) \overset{\text{Proof C.1}}{\leqslant} \sum_{i \neq j} \mathbb{E}_{u \in \mathcal{B}}\left[-\log\left(\mathrm{Softmax}_v\left(\langle \mathbf{h}_{u,i}^{\mathcal{S}}, \mathbf{h}_{v,j}^{\mathcal{S}} \rangle / \tau\right)\Big|_{v=u}\right)\right]$

- **Why it work?**

Proposition 1 (Vocabulary Transferability). *Given any nodes u, v, and assume $\|\hat{\mathbf{x}}_u^{\mathcal{S}} - \hat{\mathbf{x}}_v^{\mathcal{S}}\|_2 \leqslant \epsilon$. The semantic discrepancies $\Delta = \|f(\mathbf{g}_u^{\mathcal{S}}) - f(\mathbf{g}_v^{\mathcal{S}})\|_2$ is upper bounded by:*

$$\Delta \overset{[a]\ \text{Proof C.2}}{\leqslant} \left[\min_{\Pi \in S_K} \sum_{k=1}^{K} \left\| \mathbf{h}_{u,k}^{\mathcal{S}} - \mathbf{h}_{v,\Pi(k)}^{\mathcal{S}} \right\|_2^2 + \psi(\mathcal{R}_{\mathrm{MI}}^{u,v})\right]^{\frac{1}{2}} \overset{[b]\ \text{Proof C.3}}{\leqslant} \epsilon\sqrt{K}\left(\frac{C_\sigma L_{\mathbf{w}} L_s}{4\rho\tau}\right)^{T}, \tag{7}$$

Key Takeaway: Graph vocabulary is the **smallest & indivisible** unit that preserves structural semantics.

Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

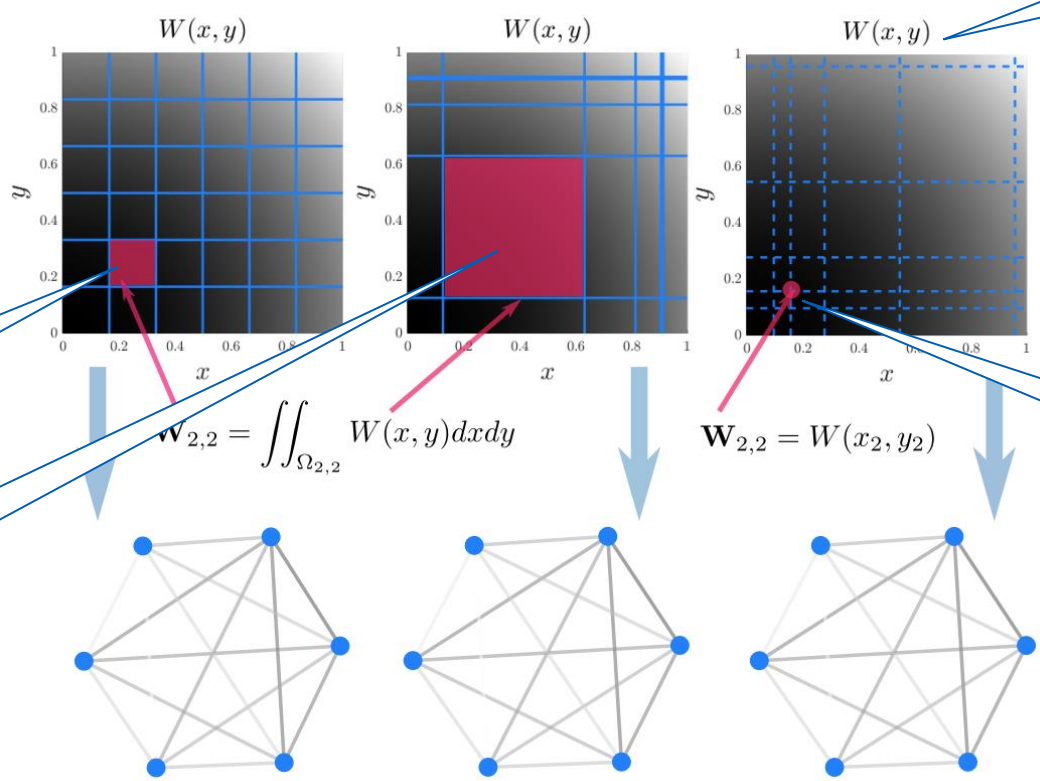
GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Phase 2: Vocabulary Modeling with Generative Graphon Experts

Goal: generate transferable, class-consistent graph vocabularies from structure-feature graphon distributions (establish a "vocabulary bank").

a continuous function that describes the connectivity pattern of a graph

graphon



average region sampling

block-wise sampling

coordinate (point) sampling

[Picture Credit] Graphon Pooling in Graph Neural Networks. *Alejandro Parada-Mayorga*, et al.

Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Phase 2: Vocabulary Modeling with Generative Graphon Experts

- **Goal:** generate transferable, class-consistent graph vocabularies from structure-feature graphon distributions (establish a "vocabulary bank").

- **Step 1:** Structure Token Modeling
  - structure token graphon approximation:

$$W_c^{\mathcal{A}}: [0,1]^2 \mapsto [0,1], \quad W_c^{\mathcal{A}}(u,v) = \frac{1}{N_c} \sum_{i=1}^{N_c} \mathcal{A}_i^{(c)}[\pi_i(u), \pi_i(v)]$$

$$\pi_i: [0,1] \mapsto \{1, \cdots, |\mathcal{A}_i|\}$$

shared

- **Step 2:** Feature Token Modeling
  - feature token graphon approximation:

$$W_c^{\mathcal{X}}: [0,1] \mapsto \mathbb{R}^d, \quad W_c^{\mathcal{X}}(u) = \frac{1}{N_c} \sum_{i=1}^{N_c} \mathcal{X}_i^{(c)}[\pi_i(u), :]$$

non-parametric

Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Phase 2: Vocabulary Modeling with Generative Graphon Experts

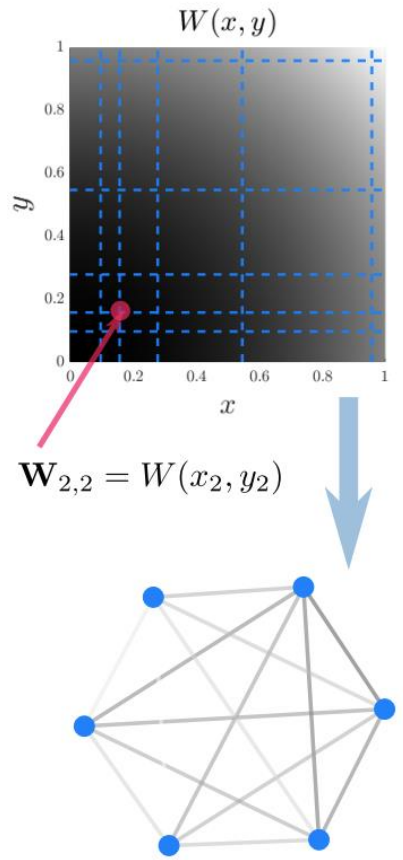
- **Goal:** generate transferable, class-consistent graph vocabularies from structure-feature graphon distributions (establish a "vocabulary bank").

- **Step 3:** Conditional Vocabulary Generation
  - **Principle: Root** (structures) + **Affix** (features)
  - sample and generation (coordinate sampling):

$$\widetilde{\boldsymbol{\mathcal{A}}}[i,j] \sim \text{Bern}\left(W_c^{\boldsymbol{\mathcal{A}}}(\mathbf{u}_i, \mathbf{u}_j)\right)$$

$$\widetilde{\boldsymbol{\mathcal{X}}}[i,:] = W_c^{\boldsymbol{\mathcal{X}}}(\mathbf{u}_i)$$

$$\mathbf{u}_1, \cdots, \mathbf{u}_{n'} \sim \mathcal{U}[0,1]$$



Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Phase 2: Vocabulary Modeling with Generative Graphon Experts

- **Goal:** generate transferable, class-consistent graph vocabularies from structure-feature graphon distributions (establish a "vocabulary bank").

- **Why it work?**

Proposition 2 (Generation Distributional Convergence). *Let* $\mathbf{g}_c^{\text{emp}}$ *be the empirical distribution over* n'*-node vocabulary subgraphs of class* c*, collected from the disentangled vocabulary bank:*

$$\mathbf{g}_c^{\text{emp}} := \frac{1}{N_c}\sum_{i=1}^{N_c}\delta_{(\boldsymbol{\mathcal{A}}_i^{(c)},\boldsymbol{\mathcal{X}}_i^{(c)})}, \quad \text{each } (\boldsymbol{\mathcal{A}}_i^{(c)},\boldsymbol{\mathcal{X}}_i^{(c)}) \in \{0,1\}^{n'\times n'} \times \mathbb{R}^{n'\times d}, \tag{12}$$

where $\delta_{(\cdot)}$ *denotes Dirac measure, representing a point mass probability distribution. Assume the true underlying structure and feature functions* $W_c^{\mathcal{A}\star}$, $W_c^{\mathcal{X}\star}$ *are bounded, Lipschitz, and satisfy permutation equivariance. If the estimators* $W_c^{\mathcal{A}}$, $W_c^{\mathcal{X}}$ *converge uniformly in* L_∞ *norm, then the total variation (TV) distance between* $\mathbf{g}_c^{\text{gen}} = (\widetilde{\boldsymbol{\mathcal{A}}}_c, \widetilde{\boldsymbol{\mathcal{X}}}_c)$ *and* $\mathbf{g}_c^{\text{emp}}$ *vanishes as* $N_c \to \infty$*:*

$$\|\mathbf{g}_c^{\text{gen}} - \mathbf{g}_c^{\text{emp}}\|_{\text{TV}} \to 0. \quad \text{(Proof C.4)} \tag{13}$$

TVD: the **strongest** statistical distance measurement (> KL, JS, and Wasserstein)

Key Takeaway: The generated graph vocabularies **match the true** distribution (their gap converges to 0).

Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Phase 3: Fine-tuning with Augmented Support Samples

- **Goal:** strengthen support samples with vocabulary-based structural augmentation.
  ("supervised augmentation")

- **MoE-CoE Network for Selective Augmentation**
  - MoE (Mixture-of-Experts): *where to route*
  - CoE (Collaboration-of-Experts): *how to compose*

$$\mathbf{S}_{\mathrm{M}} \in \mathbb{R}^{n} = \mathrm{Softmax}\left(\mathbf{W}_{\mathrm{M}}^{\top} \cdot \phi(\widehat{\mathbf{X}}_i^{\mathcal{T}})\right), \quad \mathbf{S}_{\mathrm{C}} \in \mathbb{R}^{\mathcal{C}} = \mathrm{Softmax}\left(\mathbf{W}_{\mathrm{C}}^{\top} \cdot (\phi(\widehat{\mathbf{X}}_i^{\mathcal{T}} \| \widetilde{\boldsymbol{\mathcal{X}}})\right)$$

  - vocabulary compose:

$$\widetilde{\mathbf{g}}_i^{\mathrm{gen}} \overset{\mathrm{def}}{=} (\widetilde{\boldsymbol{\mathcal{A}}}_i^{\mathrm{gen}}, \widetilde{\boldsymbol{\mathcal{X}}}_i^{\mathrm{gen}}) = \sum_{i=1}^{n} \mathbf{S}_{\mathrm{M},i}^{\top} \Big(\sum_{c=1}^{\mathcal{C}} \mathbf{S}_{\mathrm{C},c}^{\top} \cdot \mathbf{g}_c^{\mathrm{gen}} \Big), \quad \widetilde{G}_i^{\mathcal{T}} = G_i^{\mathcal{T}} \oplus \widetilde{\mathbf{g}}_i^{\mathrm{gen}}$$

  - optimize: $\mathcal{L}_{\text{MoE-CoE}}(\mathbf{S}_{\mathrm{M}}, \mathbf{S}_{\mathrm{C}}) = -\sum_{i=1}^{n} \mathbf{S}_{\mathrm{M},i}^{\top} \log(\mathbf{S}_{\mathrm{M},i}) - n \sum_{c=1}^{\mathcal{C}} \mathbf{S}_{\mathrm{C},c}^{\top} \log(\mathbf{S}_{\mathrm{C},c})$

Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Experiment: One-shot Cross-Dataset/-Domain Transfer

background color denotes different dataset domain

Table 1: Accuracy (% ± std for 20 runs) of **one-shot classification**. Best results are presented in **bold** and the runner-ups are underlined. CR = Cora, CS = CiteSeer, PM = PubMed, arXiv = ogbn-arXiv, Tech = ogbn-tech, Home = ogbn-home, Wiki = Wiki-CS. Color denotes domain.

	Cross-Dataset				Cross-Domain		
Source	CS PM Home Wiki	CR PM Home Wiki	CR CS Home Wiki	CR CS PM Home Wiki	CR CS PM Wiki	CR CS Home PM	Home Tech Wiki
Model / Target	CR	CS	PM	Tech	Home	Wiki	arXiv
	Node Classification						
Vanilla GNNs							
GCN (bb.) [40]	28.40 ± 4.62	29.25 ± 3.39	40.33 ± 6.90	61.59 ± 5.13	53.89 ± 3.35	36.74 ± 2.53	28.58 ± 5.39
GAT [91]	29.72 ± 5.17	29.31 ± 3.47	40.51 ± 3.95	61.98 ± 5.23	51.70 ± 3.96	36.24 ± 4.19	29.03 ± 5.75
Self-supervised Graph Pre-training							
GCC [69]	32.47 ± 4.55	32.78 ± 3.85	41.66 ± 3.27	64.74 ± 3.78	55.36 ± 5.86	37.66 ± 3.80	30.42 ± 5.54
DGI [92]	30.77 ± 3.92	31.41 ± 4.11	39.97 ± 5.90	63.76 ± 3.77	53.31 ± 2.58	39.20 ± 5.67	32.15 ± 4.91
GraphCL [116]	33.64 ± 5.75	28.20 ± 3.13	39.03 ± 8.67	62.44 ± 6.55	51.55 ± 8.20	38.05 ± 3.30	31.81 ± 5.04
DSSL [102]	29.76 ± 4.55	30.84 ± 7.62	39.99 ± 6.29	61.27 ± 5.21	51.90 ± 6.52	37.58 ± 6.78	28.14 ± 5.17
GraphACL [103]	35.92 ± 5.61	33.88 ± 6.69	42.73 ± 5.26	66.70 ± 4.29	58.01 ± 6.32	40.94 ± 6.65	35.57 ± 4.29
Prompt-based Graph Fine-tuning							
GPPT [82]	32.38 ± 5.74	31.78 ± 4.61	41.97 ± 5.48	65.81 ± 6.56	57.81 ± 5.36	40.97 ± 5.41	34.22 ± 6.51
GraphPrompt [58]	38.02 ± 6.52	34.57 ± 6.34	46.19 ± 7.63	70.11 ± 6.62	60.99 ± 7.19	44.02 ± 6.74	40.74 ± 6.56
GPF [14]	40.01 ± 8.21	40.07 ± 7.64	47.58 ± 5.83	70.07 ± 6.55	59.05 ± 5.85	44.62 ± 8.68	40.13 ± 5.07
ProNoG [121]	43.67 ± 6.33	40.34 ± 7.11	51.35 ± 7.16	73.49 ± 7.34	62.77 ± 6.10	45.65 ± 6.44	41.15 ± 4.21
Multi-domain GFMs							
GCOPE [136]	36.27 ± 3.93	40.42 ± 4.64	44.75 ± 4.67	71.26 ± 5.95	60.39 ± 6.08	41.80 ± 4.96	40.00 ± 6.53
MDGPT [123]	42.55 ± 6.84	37.92 ± 7.18	51.03 ± 8.99	72.10 ± 7.12	62.69 ± 7.29	45.93 ± 6.24	43.33 ± 5.75
SAMGPT [119]	46.79 ± 6.54	38.65 ± 6.35	51.92 ± 9.50	73.60 ± 7.55	64.32 ± 7.02	46.03 ± 6.98	45.27 ± 5.05
GRAVER (ours)	**48.00** ± 2.52	**41.13** ± 3.00	**55.30** ± 3.03	**77.62** ± 2.94	**67.12** ± 3.18	**49.33** ± 2.44	**49.25** ± 3.43

the best performance across all 7 target datasets

the **best average** score, the **lower variation** (more robust / stable)

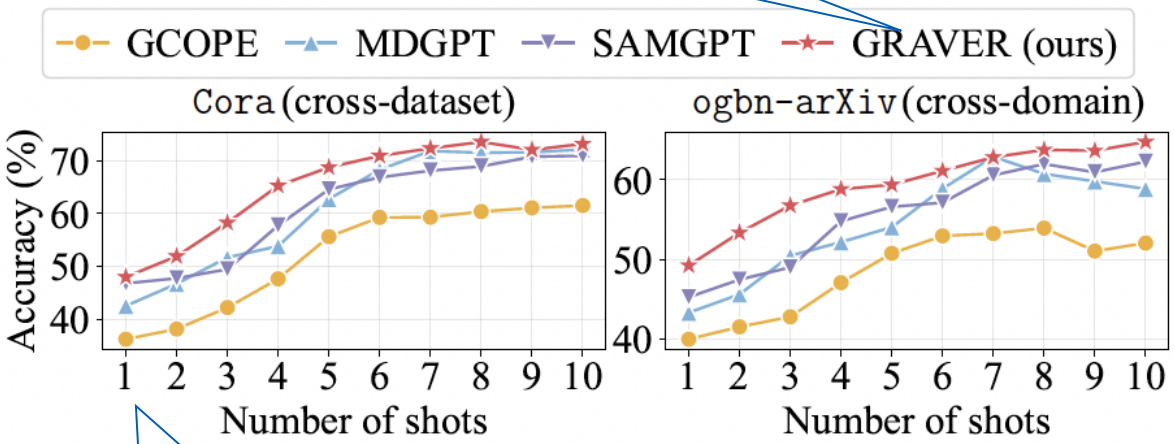
more **challenging** setting, remains **the best**

Haonan Yuan, Qingyun Sun, et al. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning, NeurIPS 2025.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Experiment: Few-shot Node and Graph Classification

consistently **outperforms** all baselines across different m-shot settings.



ure 3: m-shot node classification.

even performs **better** as the noise becomes **stronger**.

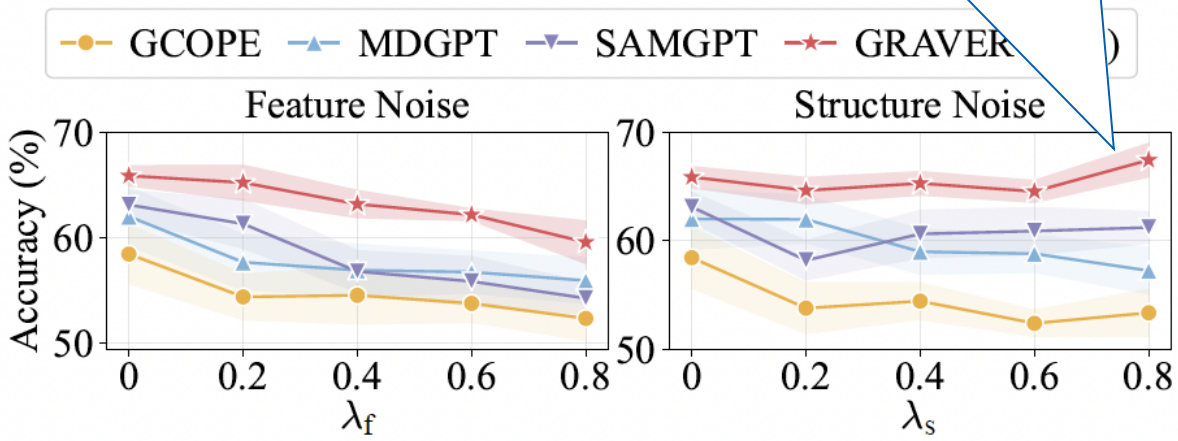


Figure 4: 5-shot graph classification (Wiki-CS).

the advantage is **more pronounced** when m is very small.

remains **robust** under both **feature** noise and **structure** noise.

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Experiment: Ablation Studies

promote vocabulary independence

vocabulary augmentation

the MoE-CoE selective network

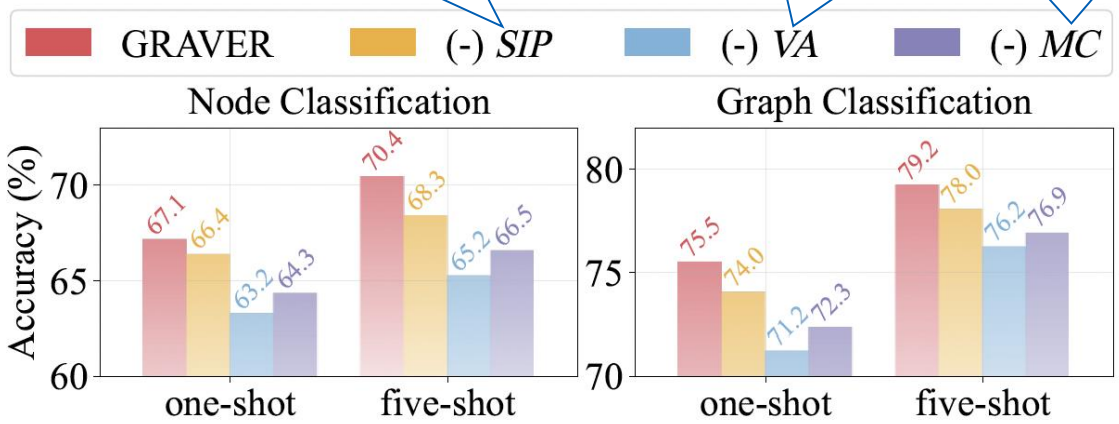


Figure 5: Ablation studies (ogbn-Home).

- All three components contributes to overall performance.
- **Importance:** *VA* > *MC* > *SIP*

raw text **significantly boost** improve zero-shot adaptability

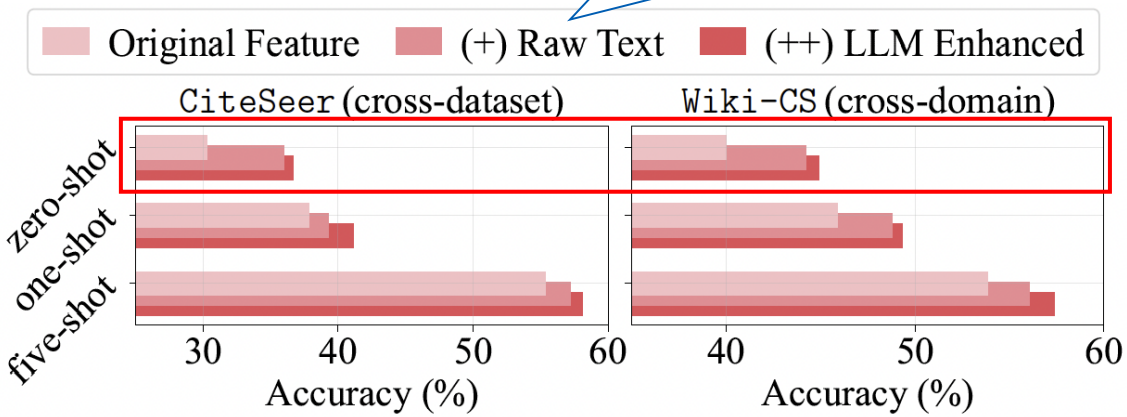


Figure 7: Analysis on LLM (node classification).

more effective under the cross-domain settings (more challenging)

GRAVER: Generative Graph Vocabularies for Robust GFM Fine-tuning

Experiment: Efficiency and Sensitivity

the **most time-efficient** among baseline SOTA GFMs

though requires a relative higher GPU memory, its **performance is better**.

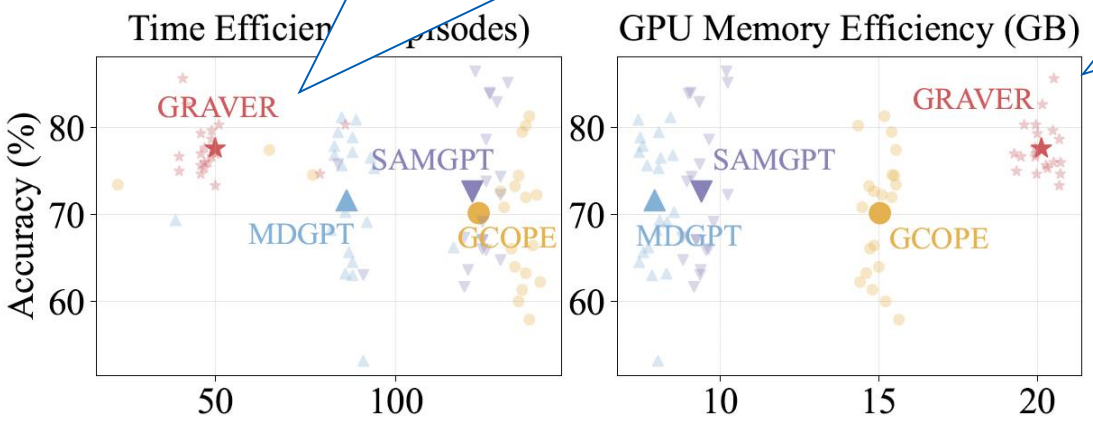


Figure 6: Efficiency analysis (ogbn-Tech).

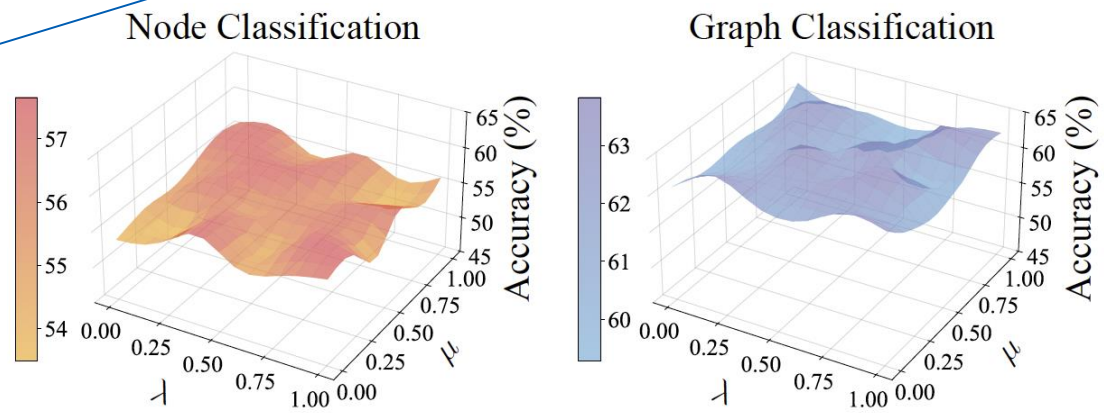


Figure 8: Hyperparameter sensitivity (PubMed).

(To ensure continuity, interpolation smo ing is applied.)

GRAVER offers the best **accuracy-memory trade-off**.

GRAVER exhibits low hyperparameter sensitivity
➔ **easy to fine-tune** in practice