MAGIC
MAchine learning for Graph Intelligence Computing

ICML
International Conference On Machine Learning

How Much Can Transfer? BRIDGE: Bounded Multi-Domain Graph Foundation Model with Generalization Guarantees

Haonan Yuan[1], Qingyun Sun[1*], Junhua Shi[1], Xingcheng Fu[2], Bryan Hooi[3], Jianxin Li[1], Philip S. Yu[4]

[1]Beihang University [2]Guangxi Normal University [3]National University of Singapore [4]University of Illinois, Chicago

Email: yuanhn@buaa.edu.cn

Paper

Code

Why Graph Foundation Models?

- Graph-structured data is ubiquitous: social networks, e-commerce, biology...
- Like GPTs in NLP, Graph Foundation Models (GFMs) aim to learn general-purpose graph knowledge across domains.

Recent Trends:

- “Pretrain-then-finetune” shows promise for cross-domain transfer.
- “Pretrain-then-prompt” further reduces computation and adapts faster.

But …

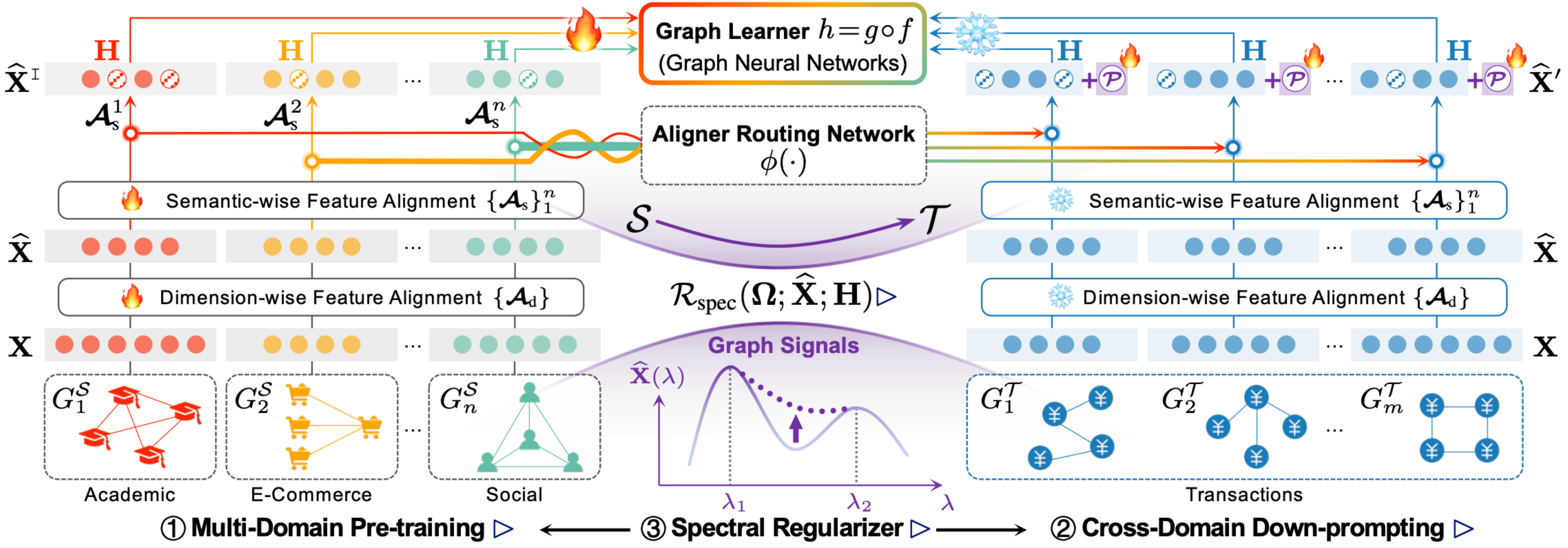
- Over-rely on text-attributed graphs, which many datasets lack.
- Lack theoretical understanding of why prompt transfer works.
- Can’t quantify how much knowledge is transferable.

We need GFMs that are generalizable, scalable, and theoretically grounded.

Key Challenges that Limit Transferability

- **Heterogeneous Feature Alignment**
  - Graphs across domains differ in both dimensions and semantics.
  - Naïve alignment (SVD, MLP) often fails to retain useful domain knowledge.
- **Inefficient Prompt Initialization**
  - Random prompts lead to slow convergence in few-shot settings.
  - We need smarter, self-supervised prompt initialization.
- **Lack of Theoretical Guarantees**
  - No current prompt-based GFM explains how much can be transferred.
  - Absence of a generalization error bound makes it hard to understand transfer limitations.

■ BRIDGE: Bounded gRaph foundatIon model pre-trained on multi-Domains with Generalization GuarantEes



Key Technical Contributions

- **Contribution 1: Domain-Invariant Feature Aligners**
  - Combine SVD + MLP for dimension alignment.
  - Use learnable semantic masks (As) to filter domain-specific noise.
  - Theoretically grounded by causal invariance assumptions.

- **Contribution 2: Self-Supervised MoE Prompting**
  - No retraining of experts is needed.
  - Routing network learns to assemble the most relevant knowledge for the target graph.

- **Contribution 3: Transfer Bound via Spectral Theory**
  - Derive an optimizable upper bound of generalization error.
  - Use this bound as a regularizer during prompt fine-tuning to control transfer quality.

Performance in Few-shot Learning

- **Task Setup:** Evaluate on 6 datasets across 3 domains (Academic, E-Commerce, Social).
- **Key Results**
  - BRIDGE outperforms 15 strong baselines, including GPPT, ProNoG, MDGPT.
  - Avg gain: +3.41% (node) and +3.25% (graph) over SOTA.

Table 1. Accuracy (% ± standard deviation for five runs) of **one-shot node** classification. CR =Cora, CS =CiteSeer, PM =PubMed, Ph =Photo, Com =Computers, Rdt =Reddit. The best results are shown in **bold** and the runner-ups are underlined.

Source (Cross-Dataset | Cross-Domain)	CS PM Ph Com	CR PM Ph Com	CR CS Ph Com	CR CS PM Com	CR CS PM Ph	CR CS PM Ph Com
Model / Target	CR	CS	PM	Ph	Com	Rdt
GCN (bb.) (Kipf & Welling, 2022)	29.36 ± 4.20	30.99 ± 4.85	40.94 ± 7.05	40.05 ± 7.25	34.62 ± 8.95	52.33 ± 7.26
GAT (Velickovic et al., 2017)	29.00 ± 5.33	29.46 ± 3.32	40.09 ± 5.21	35.60 ± 6.52	33.15 ± 6.59	50.70 ± 7.06
GCC (Qiu et al., 2020)	31.67 ± 5.23	32.55 ± 2.69	41.66 ± 4.58	42.10 ± 5.99	35.91 ± 5.68	55.67 ± 6.01
DGI (Veličković et al., 2019)	30.82 ± 4.41	31.85 ± 4.36	40.08 ± 6.22	47.23 ± 6.03	37.05 ± 6.40	61.95 ± 6.22
GraphCL (You et al., 2020)	33.28 ± 6.03	29.12 ± 4.26	39.31 ± 7.05	42.98 ± 6.54	42.87 ± 5.32	60.06 ± 5.29
DSSL (Xiao et al., 2022)	30.65 ± 5.24	31.20 ± 6.33	40.89 ± 6.94	45.32 ± 7.28	38.41 ± 6.54	57.23 ± 6.61
GraphACL (Xiao et al., 2024)	35.26 ± 4.65	34.09 ± 6.40	43.54 ± 5.58	49.20 ± 7.00	41.86 ± 6.20	60.97 ± 3.39
GPPT (Sun et al., 2022)	32.18 ± 5.02	31.33 ± 4.20	41.27 ± 5.60	46.98 ± 4.04	35.18 ± 8.09	55.24 ± 6.87
GraphPrompt (Liu et al., 2023b)	37.95 ± 6.31	34.92 ± 6.75	45.85 ± 8.58	50.42 ± 8.05	42.58 ± 7.21	64.71 ± 5.21
GraphPrompt+ (Yu et al., 2024b)	36.06 ± 6.93	33.85 ± 7.62	45.08 ± 7.99	52.28 ± 7.51	43.39 ± 6.90	64.12 ± 6.25
GPF (Fang et al., 2024)	40.26 ± 8.33	40.20 ± 7.10	47.33 ± 6.63	51.48 ± 5.34	40.09 ± 6.19	63.52 ± 6.37
ProNoG (Yu et al., 2025)	44.57 ± 6.57	39.96 ± 7.99	50.48 ± 7.06	63.30 ± 4.92	50.29 ± 6.32	64.87 ± 5.18
GCOPE (Zhao et al., 2024)	35.29 ± 4.29	40.75 ± 4.28	44.55 ± 6.07	51.55 ± 5.36	43.74 ± 7.05	64.21 ± 4.22
MDGPT (Yu et al., 2024c)	42.29 ± 7.75	37.32 ± 7.01	50.89 ± 7.74	63.63 ± 7.23	49.78 ± 8.77	63.11 ± 3.01
BRIDGE (ours)	**46.44 ± 8.01**	**42.18 ± 8.89**	**56.35 ± 7.22**	**67.87 ± 7.45**	**54.04 ± 8.52**	**68.55 ± 3.49**

Component Effectiveness & Efficiency

Ablation Study (Figure 4):

- Removing semantic aligner → performance drops significantly.
- Removing MoE → hurts especially in 1-shot (no expert routing).
- Removing $\boldsymbol{R}_{\text{spec}}$ → degrades fine-tuning in 5-shot settings.

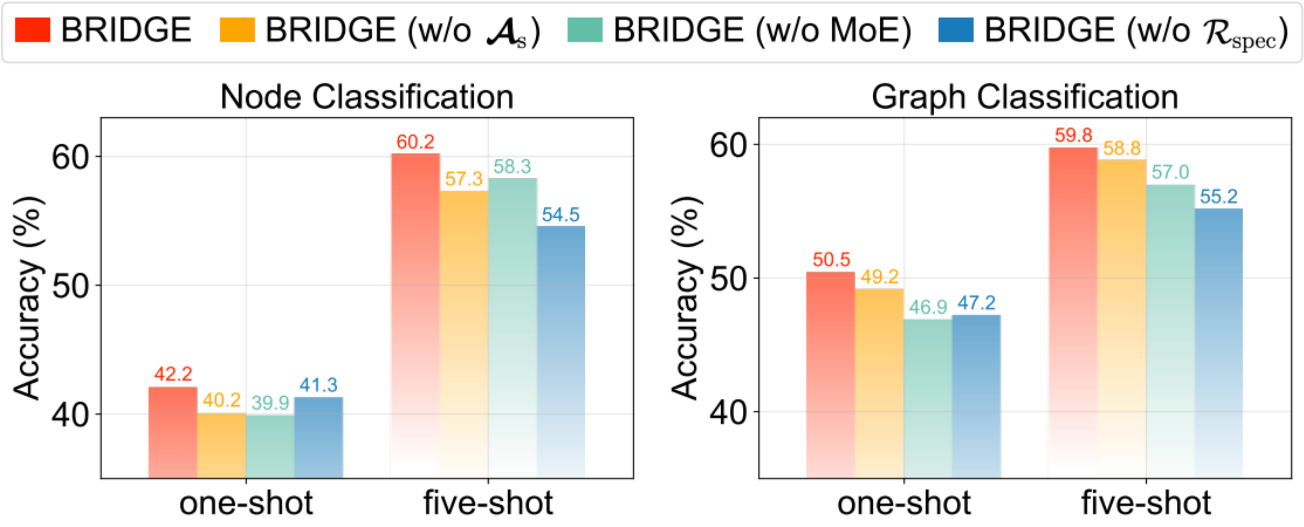


Figure 4. Ablation studies on `CiteSeer`.

Component Effectiveness & Efficiency

Prompt Efficiency (Figure 5):

- BRIDGE converges faster, with lower loss and higher accuracy in fewer epochs.
- Achieves 81.95% accuracy in 5-shot in only 200 epochs, much faster than other baselines.

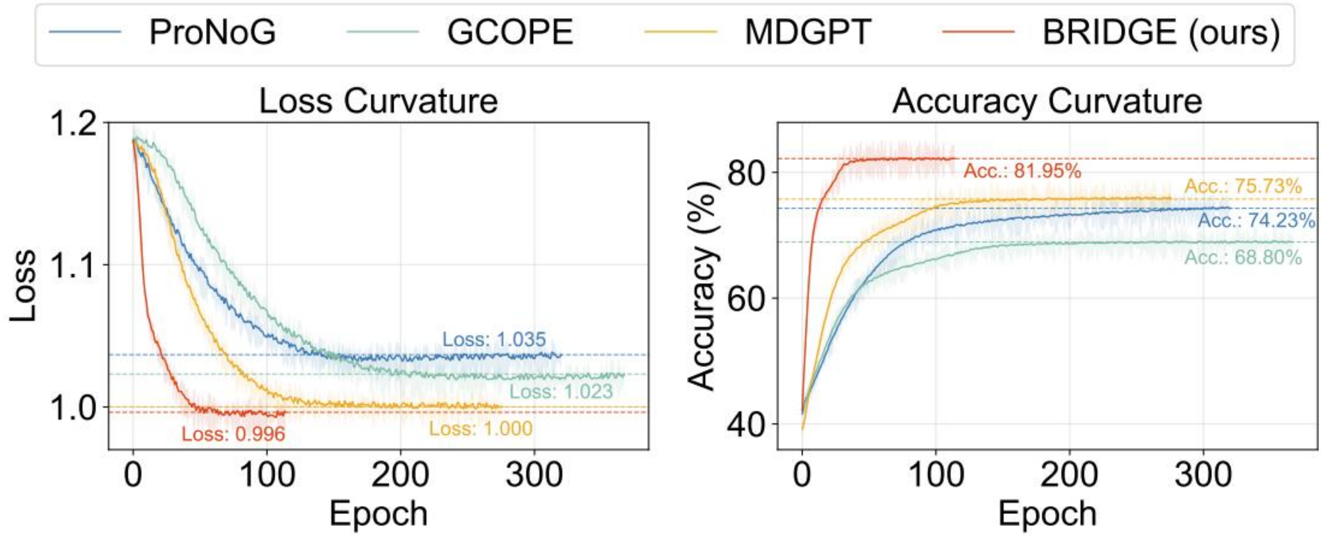


Figure 5. Down-prompt Fine-tuning on `Photo` ($m = 5$).

Robustness & LLM Extension

Hyperparameter Sensitivity

- BRIDGE shows stable performance across K, α, β, γ.
- Optimal transferability achieved at moderate regularization levels.

LLM-Enhanced Results

- Adding LLM-generated descriptions improves performance even for non-text graphs.
- BRIDGE + LLM achieves +2% gain on zero-/few-shot Reddit tasks.

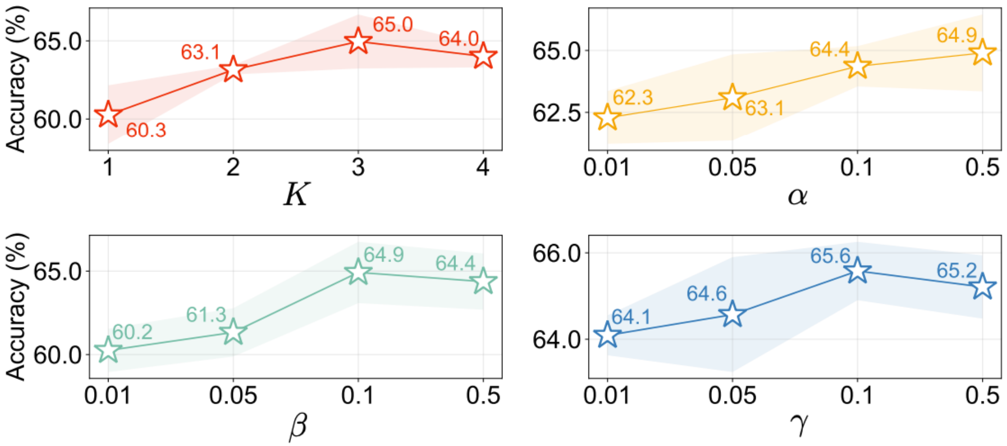


Figure 6. Hyperparameter studies on `Computers` ($m = 5$).

Table 3. Accuracy (%) of LLM-enhanced few-shot performance targeted on `Reddit` under the cross-domain setting.

Source | **Target**	CR CS PM Ph Com | Rdt					
Task	Node Classification			Graph Classification		
Model / m	0	1	5	0	1	5
MDGPT	49.69	63.11	75.29	64.80	73.14	80.29
MDGPT + LLM	51.08	65.56	76.34	66.07	75.35	82.02
BRIDGE	55.72	68.55	80.12	67.18	75.76	84.42
BRIDGE + LLM	57.16	70.70	82.31	69.14	77.31	86.87

Key Takeaways

✅ **Unified Multi-Domain Pretraining**

Domain-invariant aligners for stable feature generalization

✅ **Self-Supervised Prompting**

Lightweight MoE routing enables adaptive, fast transfer

✅ **Theory-Guided Fine-tuning**

Spectral regularizer from generalization bound ensures controllability

Results

🌍 Strong transfer performance across domains

📊 Outperforms 15 SoTA baselines in few-shot settings

⚡ Fast convergence + Robust adaptation

MAGIC
MAchine learning for **G**raph **I**ntelligence **C**omputing

ICML
International Conference On Machine Learning

How Much Can Transfer? BRIDGE: Bounded Multi-Domain Graph Foundation Model with Generalization Guarantees

Haonan Yuan[1], Qingyun Sun[1*], Junhua Shi[1], Xingcheng Fu[2], Bryan Hooi[3], Jianxin Li[1], Philip S. Yu[4]

[1]Beihang University [2]Guangxi Normal University [3]National University of Singapore [4]University of Illinois, Chicago

Email: yuanhn@buaa.edu.cn

Paper

Code