

Relation-Aware Graph Foundation Model

Jianxiang Yu¹, Jiapeng Zhu¹, Yibo Zhao¹, Hao Qian², Ziqi Liu², Zhiqiang Zhang², Xiang Li¹

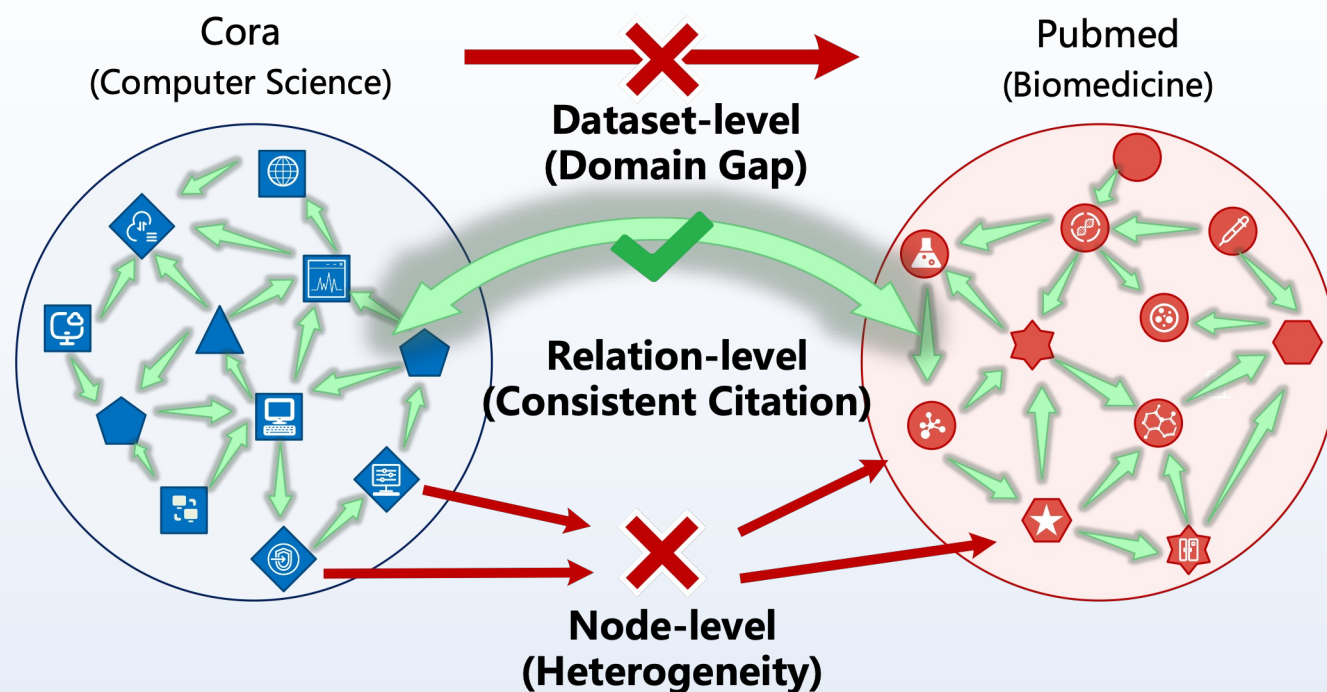
¹East China Normal University, ²Ant Group



Challenge: Graph foundation models (GFMs) aim to learn across diverse graphs, generalize across datasets, and transfer to unseen graphs.

- Node- and dataset-level approaches struggle with substantial variations in node features and semantics across graphs.

- **Relations** may exhibit more consistent semantics across graphs, offering a promising foundation for cross-graph knowledge transfer.



Introduction

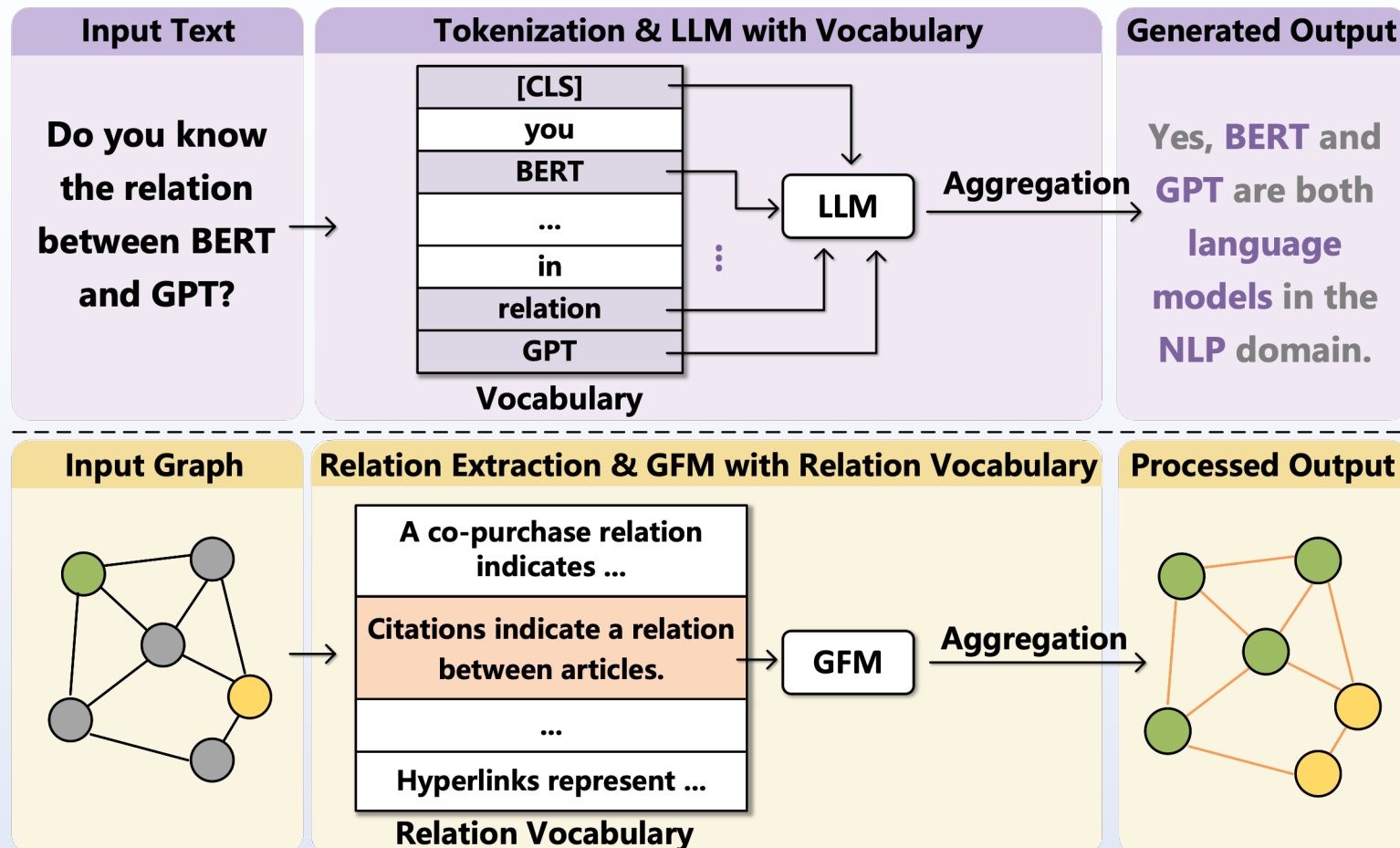
From Language Tokens to **Relation Tokens**

Inspiration:

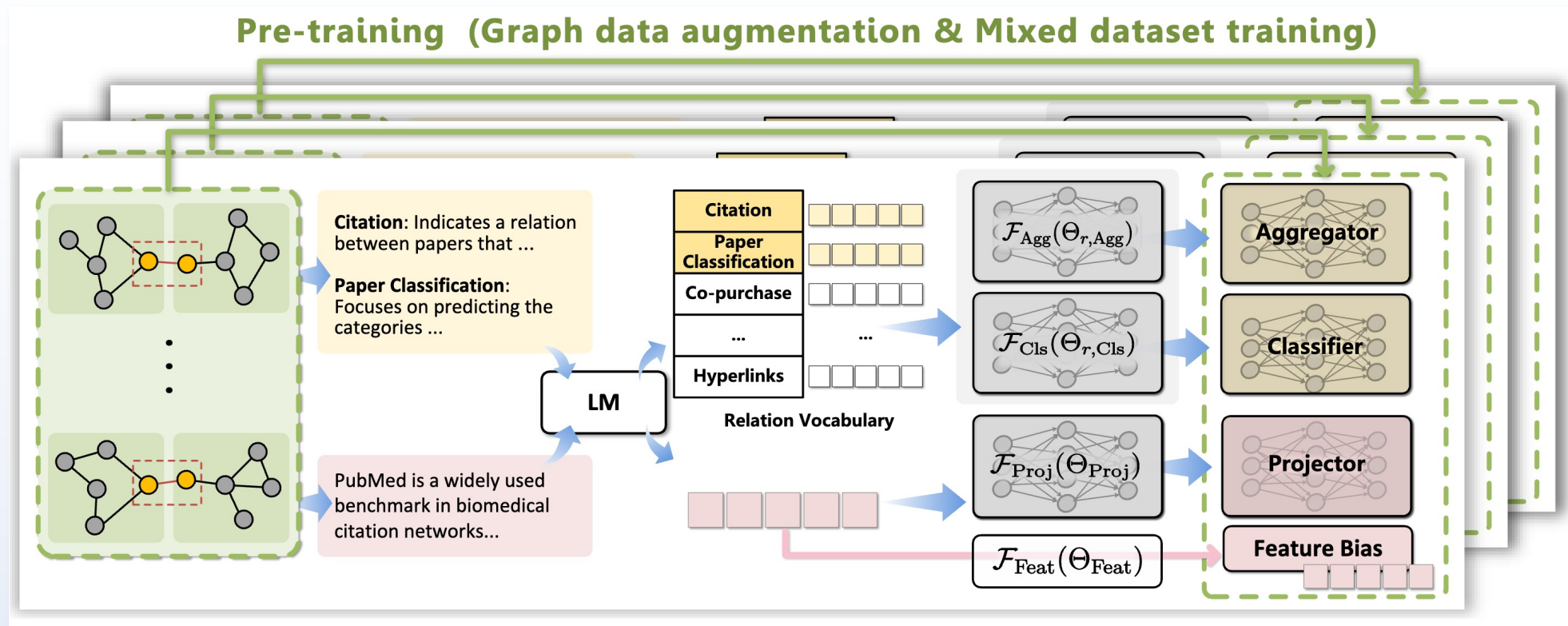
LLMs learn transferable knowledge through a **shared token vocabulary**, enabling generalization across diverse tasks.

Our Idea:

REEF introduces a relation vocabulary, treating graph relations as semantic tokens to enable knowledge sharing and generalization across graphs.



Relation-Aware Graph Foundation Model



Key Components:

1. Relation-aware Aggregator

2. Relation-aware Classifier

3. Dataset-specific Feature Adaptation

4. Pre-training Strategy

Relation-Aware Graph Foundation Model

Preprocessing: Feature Alignment

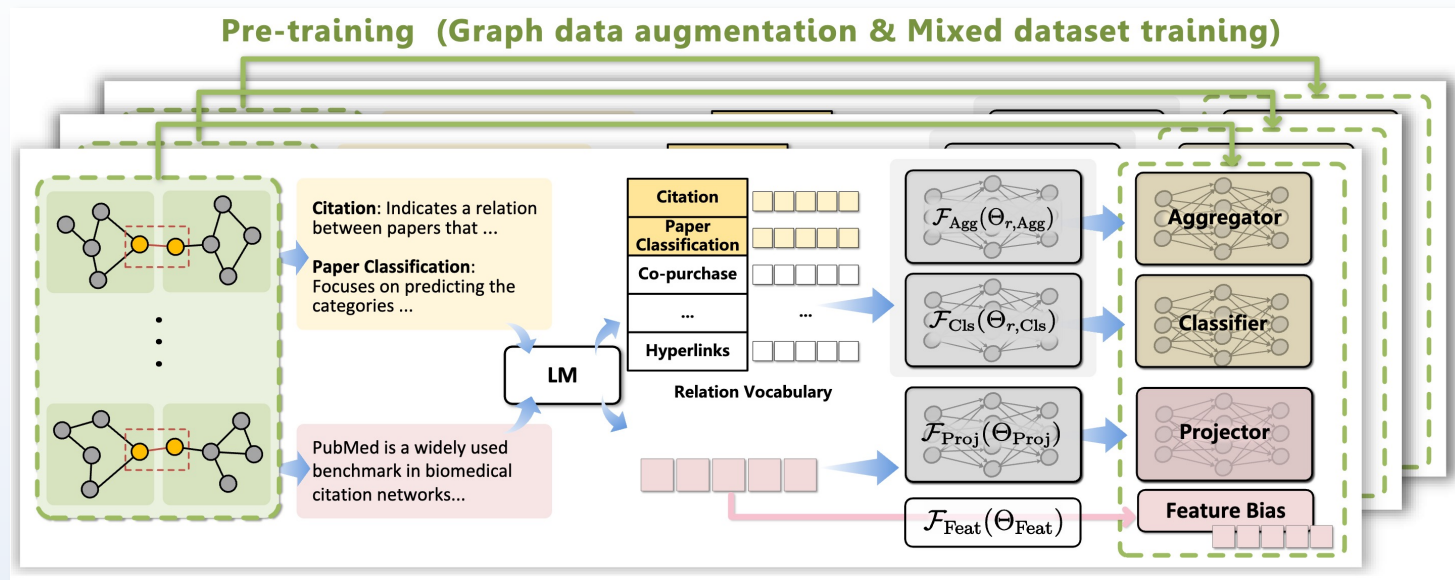
Heterogeneous node features are aligned into a unified representation space using SVD-based or LM-based methods, enabling cross-dataset learning.

$$\hat{x}_i = \mathcal{T}(x_i) \in \mathbb{R}^{d_x}, \quad x_i \in \mathbb{R}^{d_{x_i}},$$

Relation Vocabulary Construction

To capture transferable relational semantics, REEF constructs a **shared relation vocabulary** using natural-language and encodes relation tokens with a pre-trained language model.

$$h_r = \text{LM}(r), \quad r \in \mathcal{R}_{\text{source}}$$



Relation-Aware Graph Foundation Model

1. Relation-aware Aggregator

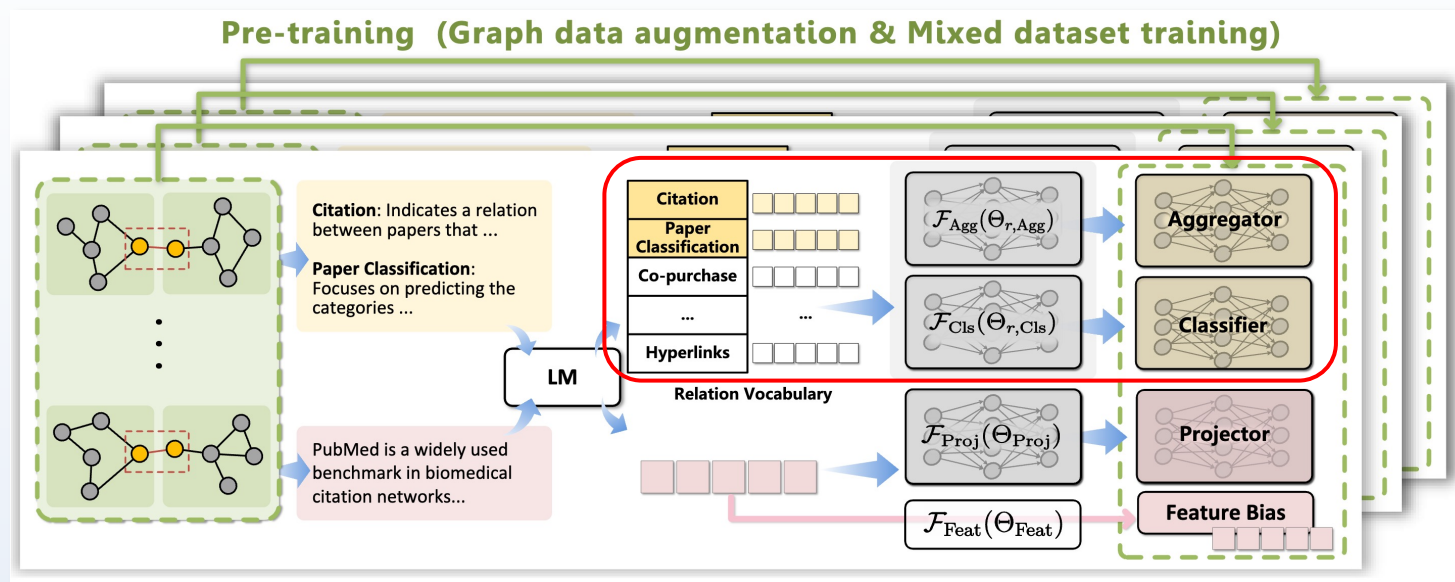
Different relations exhibit distinct structural patterns, motivating REEF to **generate relation-specific aggregation parameters** through a shared hypernetwork for adaptive message passing.

$$\Phi_r^{(l)} = \mathcal{F}_{\text{Agg}}(h_r; \Theta_{r, \text{Agg}}^{(l)}), \quad l = 1, 2 \dots, L$$

2. Relation-aware Classifier

Another shared hypernetwork **generates relation-specific classifiers conditioned on relation semantics**, allowing diverse graph learning tasks to be handled through unified relation prediction.

$$\Psi_r = \mathcal{F}_{\text{Cls}}(h_r; \Theta_{r, \text{Cls}})$$



Relation-Aware Graph Foundation Model

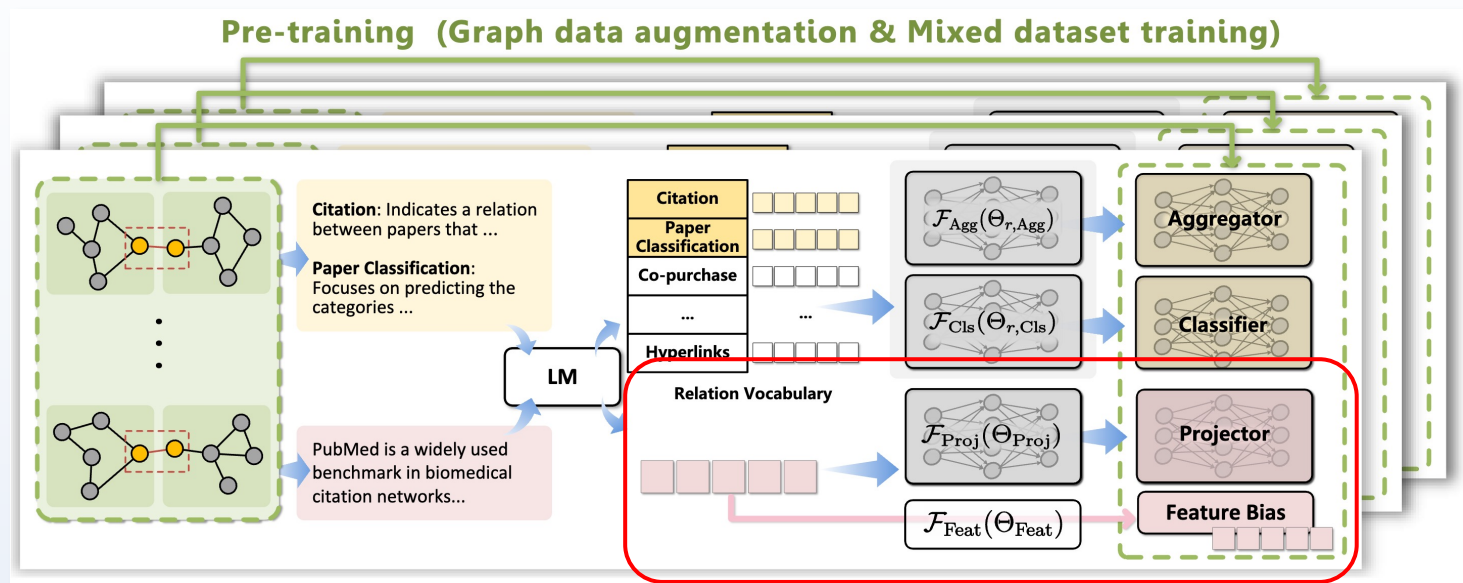
3. Dataset-specific Feature Adaptation

Motivation: Although relation semantics can be shared across graphs, node feature distributions and dataset characteristics may vary significantly.

Dataset-specific Projector: Dataset semantics guide the generation of dataset-specific projection parameters, enabling adaptive feature transformations across graphs.

$$h_{t_g} = \text{LM}(t_g), \quad \Phi_g^{(l)} = \mathcal{F}_{\text{Proj}}(h_{t_g}; \Theta_{g, \text{Proj}}^{(l)})$$

Dataset-level Feature Bias: A dataset-conditioned feature bias enriches node representations with domain-specific semantic information.



REEF integrates relation-aware neighbor aggregation with dataset-specific node transformations to learn transferable graph representations.

$$h_{v_i}^{(l+1)} = \Phi_g^{(l)} h_{v_i}^{(l)} + \sum_{r \in \mathcal{R}_{\text{source}}} \sum_{v_j \in \mathcal{N}_i^r} \frac{1}{|\mathcal{N}_i^r|} \Phi_r^{(l)} h_{v_j}^{(l)}$$

Relation-Aware Graph Foundation Model

4. Pre-training Strategy

Relation Token Prediction (RTP):

REEF unifies node classification and link prediction into a shared relation prediction task, allowing relational knowledge to be learned across heterogeneous graphs. Each training instance is represented as a relation triplet: $\langle s_i, r_{ij}, s_j \rangle$. The relation prediction is formulated as:

$$P(y = 1 \mid \langle s_i, r_{ij}, s_j \rangle) = \mathcal{F}(\langle s_i, r_{ij}, s_j \rangle; \Theta)$$

Mixed-Dataset Training:

Training samples from multiple graph datasets are jointly optimized, enabling REEF to learn shared relational knowledge across different domains.

Graph Data Augmentation:

Random edge dropping introduces structural variations during pre-training, improving robustness and generalization across graphs.

Experiments

Pre-training Performance

Experimental Setup: (1) Joint pre-training on seven graph datasets from four domains.
(2) Comparison with independently and jointly trained GNN baselines.

Table 1: Comparison of model performance on various pretraining datasets. The table reports accuracy (Acc %) and the performance gap ($\Delta = \text{Top} - \text{Score}$) to the best result on each dataset. The average “Acc” is the mean accuracy calculated across the five datasets presented in the table. The **best** result is bold and underlined, while the runner-up is underlined.

Domain	Citation		WebKB		Amazon	Average	
Dataset	Pubmed	Citeseer	Wisconsin	Texas	Photo	Acc $\uparrow$	Δ $\downarrow$
GCN (ind)	<u>88.42</u> ($\Delta_{0.00}$)	<u>76.50</u> ($\Delta_{0.05}$)	51.76 ($\Delta_{24.71}$)	55.14 ($\Delta_{25.81}$)	<u>93.02</u> ($\Delta_{0.00}$)	72.97	10.11
GAT (ind)	86.33 ($\Delta_{2.09}$)	<u>76.55</u> ($\Delta_{0.00}$)	49.41 ($\Delta_{27.06}$)	52.16 ($\Delta_{28.79}$)	92.73 ($\Delta_{0.29}$)	71.44	11.65
GCN (joint)	61.21 ($\Delta_{27.21}$)	64.86 ($\Delta_{11.69}$)	61.21 ($\Delta_{15.26}$)	57.14 ($\Delta_{23.81}$)	30.45 ($\Delta_{62.57}$)	54.97	28.11
GAT (joint)	52.93 ($\Delta_{35.49}$)	57.81 ($\Delta_{18.74}$)	55.81 ($\Delta_{20.66}$)	52.38 ($\Delta_{28.57}$)	28.52 ($\Delta_{64.50}$)	49.49	33.59
RGCN (joint)	80.30 ($\Delta_{8.12}$)	71.77 ($\Delta_{4.78}$)	<u>72.09</u> ($\Delta_{4.38}$)	<u>80.95</u> ($\Delta_{0.00}$)	79.03 ($\Delta_{13.99}$)	<u>76.83</u>	<u>6.25</u>
REEF	<u>86.38</u> ($\Delta_{2.04}$)	72.37 ($\Delta_{4.18}$)	<u>76.47</u> ($\Delta_{0.00}$)	<u>70.27</u> ($\Delta_{10.68}$)	<u>93.01</u> ($\Delta_{0.01}$)	<u>79.70</u>	<u>3.38</u>

Key Findings

Superior Performance: REEF achieves the highest average accuracy of 79.70%, outperforming RGCN by 2.87 pp.

Reduced Negative Transfer: Joint training substantially degrades the performance of conventional GNNs, while REEF maintains stronger performance across datasets.

Effective Knowledge Sharing: Relation-aware modeling facilitates knowledge transfer across heterogeneous graph datasets.

Few-shot Transfer Learning (1-shot, SVD-based)

Experimental Setup: Evaluate 1-shot transfer learning on three unseen graph datasets using SVD-aligned node features.

Method	Cora (Citation)			Cornell (WebKB)			Computers (Amazon)		
	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
GCN	0.3012 $\pm$.06	0.6444 $\pm$.04	0.2591 $\pm$.04	0.3263 $\pm$.04	0.5666 $\pm$.01	0.3151 $\pm$.03	0.2602 $\pm$.07	0.6773 $\pm$.02	0.2428 $\pm$.04
GAT	0.3646 $\pm$.04	0.6769 $\pm$.03	0.3108 $\pm$.04	0.3275 $\pm$.14	0.5306 $\pm$.03	0.1497 $\pm$.04	0.3482 $\pm$.07	0.6878 $\pm$.05	0.2397 $\pm$.05
GraphCL	0.2507 $\pm$.06	0.6350 $\pm$.03	0.2240 $\pm$.03	0.4175 $\pm$.04	0.6350 $\pm$.02	0.3500 $\pm$.04	0.2856 $\pm$.04	0.6467 $\pm$.03	0.1653 $\pm$.06
SimGRACE	0.2492 $\pm$.02	0.5765 $\pm$.03	0.1567 $\pm$.04	0.3438 $\pm$.13	0.5954 $\pm$.09	0.2168 $\pm$.09	0.2666 $\pm$.10	0.6286 $\pm$.01	0.1603 $\pm$.03
GCOPE+CL	0.3368 $\pm$.02	0.6971 $\pm$.04	0.2967 $\pm$.03	0.3975 $\pm$.10	0.6694 $\pm$.04	0.3120 $\pm$.04	0.3439 $\pm$.03	0.7023 $\pm$.01	0.2976 $\pm$.03
GCOPE+Sim	0.2525 $\pm$.05	0.5744 $\pm$.03	0.1722 $\pm$.06	0.3675 $\pm$.09	0.6045 $\pm$.04	0.2339 $\pm$.04	0.3230 $\pm$.01	0.6994 $\pm$.00	0.2515 $\pm$.00
MDGPT	0.4421 $\pm$.08	0.7912 $\pm$.05	0.4271 $\pm$.08	0.2400 $\pm$.08	0.5776 $\pm$.04	0.1723 $\pm$.04	0.3837 $\pm$.09	0.7950 $\pm$.04	0.4120 $\pm$.08
REEF	0.4866$\pm$.04	0.8035$\pm$.03	0.4800$\pm$.05	0.4865$\pm$.08	0.7016$\pm$.03	0.3649$\pm$.07	0.4772$\pm$.06	0.8480$\pm$.03	0.4403$\pm$.02
Improve.	+10.07%	+1.55%	+12.39%	+16.52%	+4.81%	+4.08%	+24.05%	+6.67%	+6.87%

Key Findings

- Superior Few-shot Transferability: REEF outperforms the strongest baselines by **4.45–9.35 pp** across three unseen datasets.
- Cross-Graph Generalization: Relation-level pre-training enables effective transfer with **only one labeled example** per class.

Few-shot Transfer Learning (3-shot, LM-based)

Experimental Setup: (1) 3-shot transfer learning on text-attributed graphs, with node features initialized using Sentence-BERT. (2) Two pre-training settings: S1 excludes E-commerce graphs, while S2 additionally includes Amazon Photo.

Key Findings

1. Strong LM-based Transferability:

REEF achieves the highest accuracy across all three target datasets using LM-based node features.

2. Cross-Domain Generalization:

REEF maintains strong performance on E-commerce datasets even **without E-commerce pre-training data**.

Method	Cora	History	Ratings
GraphMAE	72.49	39.15	31.68
LLaGA	60.70	36.45	23.45
OFA	52.49	39.36	29.08
OFA-FS	42.10	17.50	20.50
Prodigy (M)	23.40	12.71	20.16
Prodigy	40.59	19.47	20.84
GraphText	50.33	48.00	37.67
GraphAdapter	33.74	36.60	29.41
GraphLLM	68.00	56.67	34.33
GPPT	44.14	27.54	14.24
Gprompt	55.38	17.36	15.38
SBERT (w/o PT)	68.42	51.25	20.95
GIT-G	60.93	50.78	21.21
GIT-S	63.23	50.99	21.17
REEF (S1) [†]	75.95	57.40	37.20
REEF (S2)	74.58	62.82	38.24

Cross-Domain Generalization

Experimental Setup:

- (1) Evaluate cross-domain generalization on 26 graph datasets across six domains, covering diverse graph structures and semantics.
- (2) Follow the GraphAny evaluation protocol and compare REEF with GAT, GraphAny, and TS-Mean.

Domain	#Datasets	GAT	GraphAny	TS-Mean	REEF
Citation Networks	4	64.23	70.14	63.39	71.08
Co-authorship Networks	4	77.63	81.82	81.03	83.00
Social Networks	6	69.00	69.03	68.31	70.25
Co-purchase Graphs	3	64.12	71.58	71.27	71.00
Heterophilic Benchmarks	6	72.85	69.34	68.53	76.77
Air-traffic Graphs	3	39.14	40.38	39.15	42.64
Average (%) ↑	26	66.47	68.23	66.54	70.75

Key Findings

1. Strong Cross-Domain Generalization: REEF achieves the highest average accuracy (**70.75%**) across 26 datasets, ranking first in **5 out of 6 graph domains**.
2. Robustness to Diverse Graph Structures: REEF substantially outperforms GraphAny on heterophilic graphs (**+7.43 pp**), demonstrating strong adaptability to diverse neighborhood patterns.

Zero-shot Transfer to Unseen Domains and Tasks

Experimental Setup:

- (1) Evaluate zero-shot transfer on BBBP and HIV molecular graph classification, an unseen domain and task during pre-training.
- (2) Follow the GOFA evaluation setting, without molecular-domain pre-training or target-task fine-tuning for REEF.

Table 4: Zero-shot transfer performance on molecular graph benchmarks without molecular-domain pretraining or fine-tuning.

Method	FT scale	BBBP	HIV
OFA [24]	N/A	—	35.67
MoMu [37]	N/A	49.81	50.26
Galactica [41]	N/A	53.94	33.85
GIMLET [60]	400M QA	59.39	66.24
GOFA [21]	100k QA	54.91	53.02
REEF	None	54.80	63.17

Key Findings

1. Zero-shot Cross-Domain and Cross-Task Transfer

REEF achieves competitive performance on molecular graph classification without molecular-domain pre-training or target-task fine-tuning.

2. Strong Performance with Limited Supervision

REEF achieves **63.17% AUROC on HIV**, outperforming GOFA by **10.15 pp**, despite using no additional molecular QA fine-tuning.

Ablation Study

Experimental Setup:

- (1) Evaluate four ablated variants by modifying relation representations, dataset adaptation, and graph augmentation.
- (2) Compare pre-training and transfer performance to quantify the contribution of each component.

Method	Pre-training datasets							Transfer datasets			Avg.		
	FB15K237	WN18RR	Pubmed	Citeseer	Wisconsin	Texas	Photo	Cora	Cornell	Computers	Pre.	Trans.	All
REEF-LM	91.33	94.06	86.05	70.42	64.71	59.46	93.53	46.58	44.72	43.46	79.94	44.92	69.43
REEF-FB	89.90	92.18	86.23	70.27	70.59	54.05	92.16	40.07	44.51	42.09	79.34	42.22	68.21
REEF-FP	90.72	94.29	85.04	73.12	56.86	59.46	93.33	40.22	43.69	46.50	78.97	45.10	68.32
REEF-AGU	91.14	93.78	86.41	71.62	68.63	64.86	92.22	40.81	43.69	46.68	81.24	43.73	69.98
REEF	91.04	94.70	86.38	72.37	76.47	70.27	93.01	48.66	48.65	47.72	83.46	48.34	72.93

Key Findings

1. Importance of Relation Semantics: LM-initialized relation representations provide transferable semantic knowledge beyond randomly initialized relation embeddings.
2. Contribution of Dataset Adaptation: Both dataset-specific feature bias and projection improve REEF's overall performance, supporting adaptation across heterogeneous graphs.
3. Effectiveness of Graph Augmentation: Graph augmentation further enhances model performance, particularly in transfer learning.

Conclusion

- 1** *We introduce relation tokens as transferable semantic units to facilitate knowledge sharing and generalization across graphs from diverse datasets and domains.*
- 2** *We propose REEF, a relation-aware graph foundation model that leverages hypernetworks to generate relation-specific parameters for adaptive graph learning.*
- 3** *We develop a unified Relation Token Prediction (RTP) objective to support joint pre-training across diverse graph datasets and tasks.*
- 4** *Extensive experiments demonstrate strong cross-graph generalization, including multi-dataset pre-training, few-shot transfer, and zero-shot transfer to unseen domains and tasks.*

Relation tokens offer a promising direction for building transferable graph foundation models, while improving semantic robustness and computational efficiency remains an important future direction.

Thank you!

Contact: ffjasonyu@gmail.com

Paper: <https://arxiv.org/abs/2505.12027>

Model: <https://huggingface.co/ffjasonyu/REEF>

Github: <https://github.com/jianxiangyu/REEF>