39th Annual Conference on Neural Information Processing Systems (NeurIPS 2025)

HubGT: Fast Graph Transformer with Decoupled Hierarchy Labeling

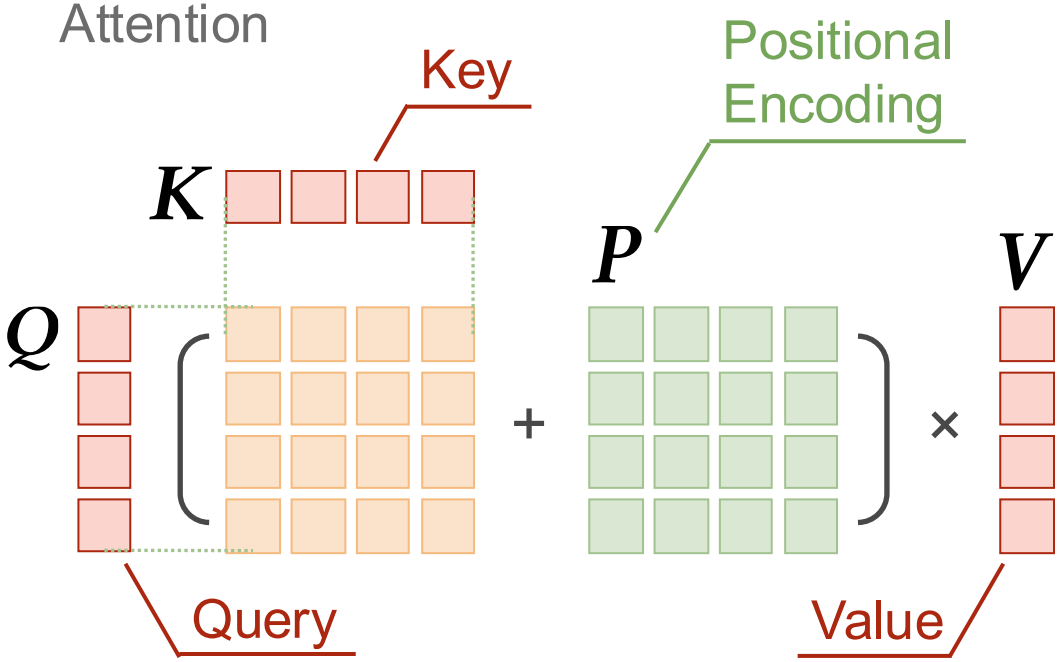
Ningyi Liao* Zihao Yu* Siqiang Luo Gao Cong

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

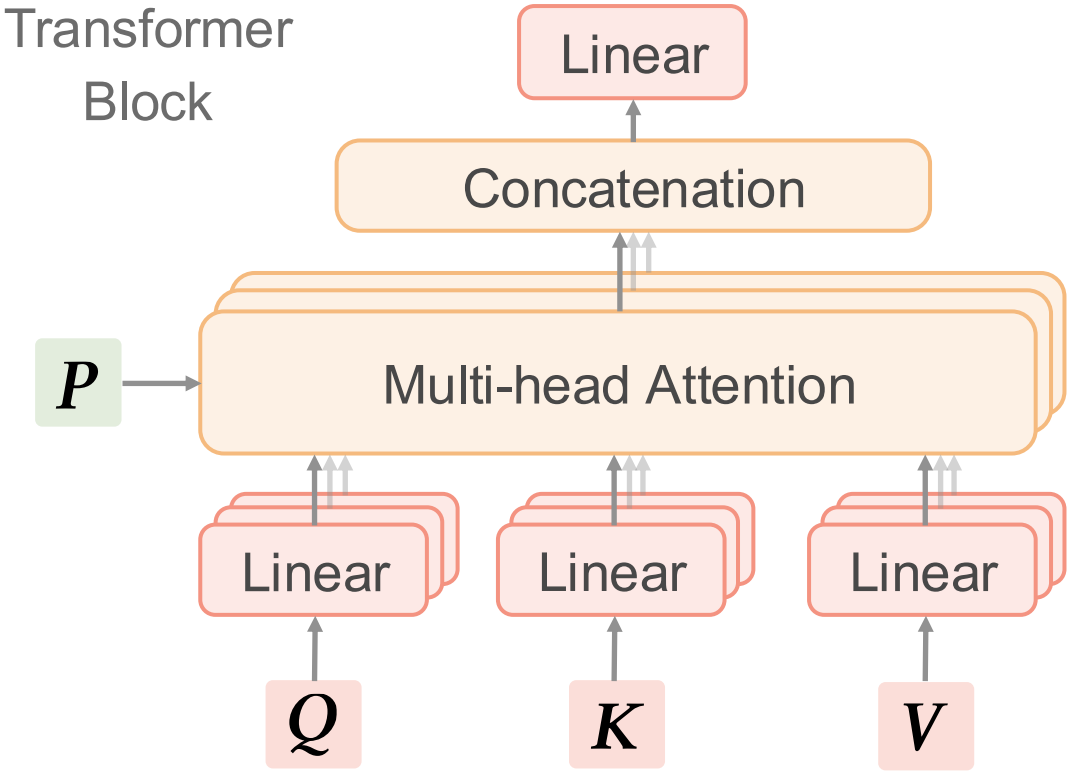
Background: Transformer PE

- Positional Encoding (PE): relational bias in attention

Dot-product Attention



Transformer Block



A Vaswani et al. “Attention Is All You Need”. NeurIPS 2017.

Motivation: Graph Transformer PE

- Positional Encoding: introducing pair-wise relation beyond adjacency

$$\tilde{\boldsymbol{H}}_i = \text{softmax}\left(\frac{\boldsymbol{Q}_i \boldsymbol{K}_i^\top}{\sqrt{F_K}} + \boldsymbol{P}\right)\boldsymbol{V}_i,$$

Positional Encoding

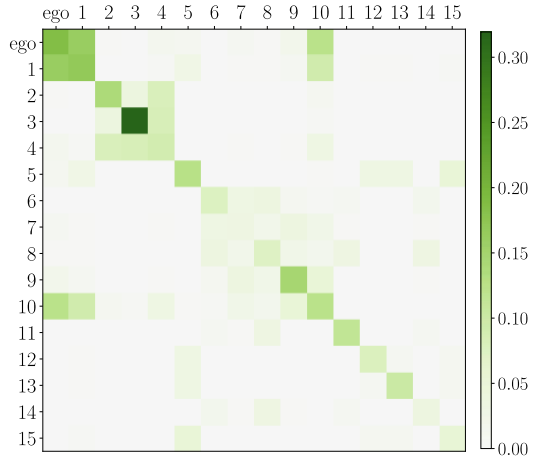
Node Feature

Transformer Attention

- Complexity:

$O(n^2)$ Full graph

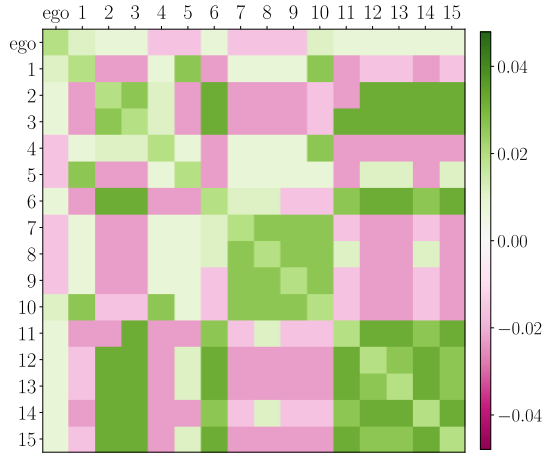
$O(s^2)$ Subgraph



Adjacency PE

☹

Too sparse



SPD Encoding

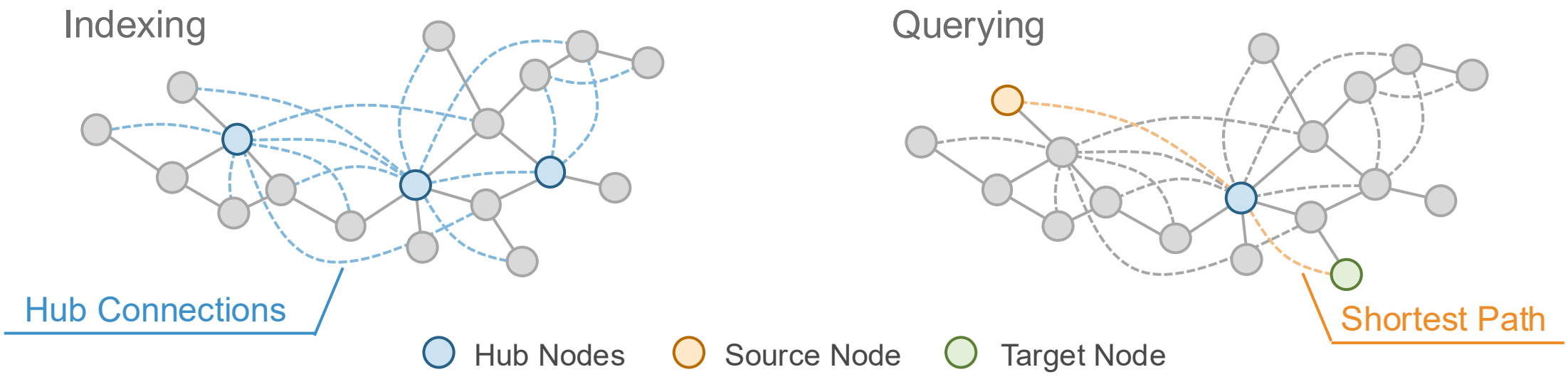
☺

Dense & Informative

C Ying et al. “Do Transformers Really Perform Bad for Graph Representation?”. NeurIPS 2021.

Method: Hub Labeling

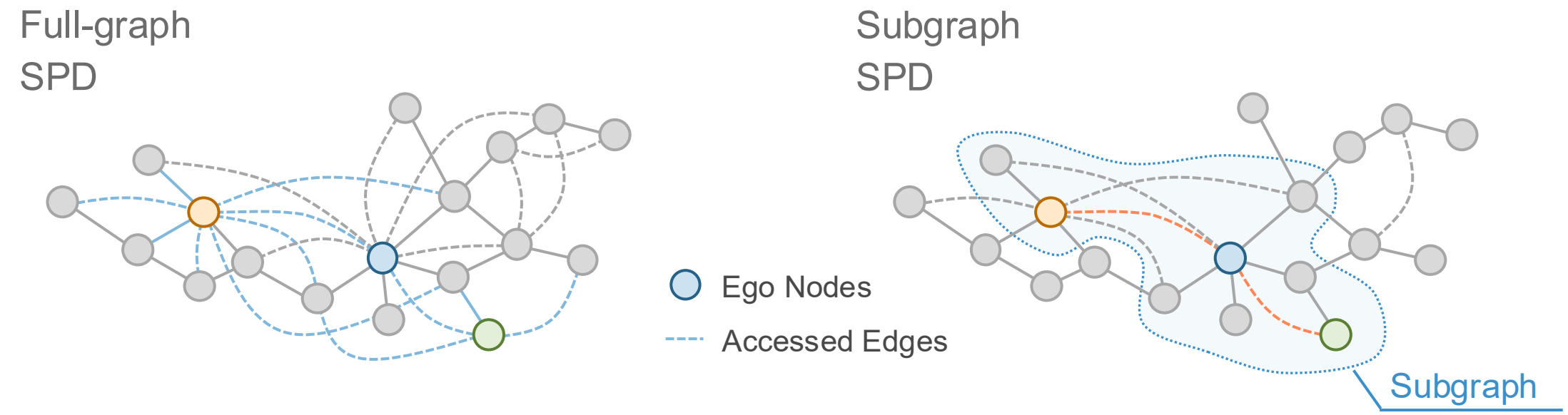
- Hub Labeling: adding links to selected hub nodes → distant node pairs can be easily reached through hubs
- Shortest Path Distance (SPD) can be computed through hubs



T Akiba et al. “Fast exact shortest-path distance queries on large networks by pruned landmark labeling”. SIGMOD 2013.

Method: Hub Labeling

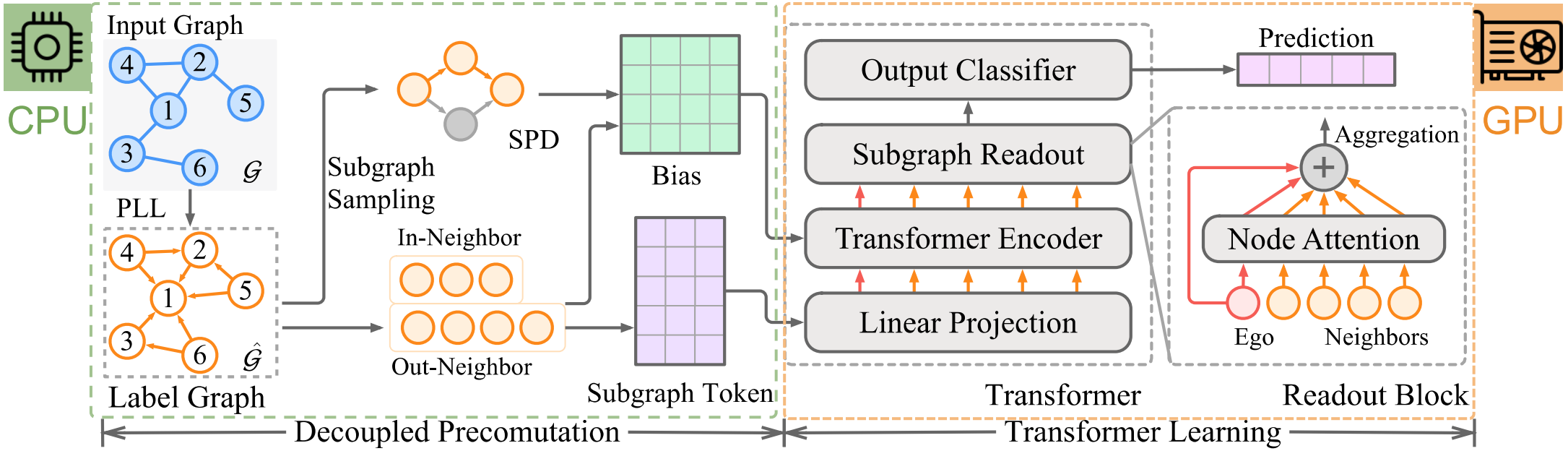
- Subgraph SPD (ours): querying any node pairs within a subgraph



Query Time	$O(s)$	$O(1)$
Index Size	$O(ns^2)$	$O(ns^2)$

Framework: Two-stage Processing

- Hub node → hierarchical subgraph → input token
- SPD value → positional encoding → attention bias



Results: Complexity Analysis

- Only linear time & memory complexity

☹ *Not scalable to large m and n* ☹ *Insufficient memory*

Taxonomy	Model	Precompute Time	Train Time	RAM Mem.	GPU Mem.	Scale
Vanilla (FB)	Graphormer [2]	$O(n^3)$	$O(Ln^2F)$	$O(n^2)$	$O(Ln^2F)$	0.3K/0.6K
	GRPE [18]	$O(n^2)$	$O(Ln^2F)$	$O(n^2)$	$O(Ln^2F)$	0.3K/0.6K
Kernel-based (NS)	GraphGPS [7]	$O(n^3)$	$O(LnF^2 + LmF)$	$O(nF + n^2)$	$O(Ln_bF + m)$	1.0K/3.0K
	Exphormer [19]	$O(dn)$	$O(LnF^2 + LmF)$	$O(nF + m)$	$O(Ln_bF + m)$	0.2M/1.2M
	NodeFormer [6]	–	$O(LnF^2 + LmF)$	$O(nF + m)$	$O(Ln_bF + m)$	2.4M/60M
	DIFFormer [8]	–	$O(LnF^2 + LmF)$	$O(nF + m)$	$O(Ln_bF + m)$	1.6M/40M
	PolyNormer [12]	–	$O(LnF^2 + LmF)$	$O(nF + m)$	$O(Ln_bF + m)$	2.4M/0.1B
Hierar-chical (RS)	NAGphormer [9]	$O(LmF_0)$	$O(LnF^2)$	$O(LnF)$	$O(Ln_bF^2)$	2.4M/60M
	PolyFormer [11]	$O(LmF_0)$	$O(LnF^2)$	$O(LnF)$	$O(Ln_bF^2)$	0.2M/30M
	ANS-GT [13]	$O(ns^2 + md^L)$	$O(LnF^2 + Lns^2F + Lm)$	$O(nF + ns^2 + md^L)$	$O(Ln_bF + n_bs^2)$	20K/0.2M
	GOAT [10]	$O(nF)$	$O(LnF^2 + LmF)$	$O(nF + m)$	$O(Ln_b^2 + Ln_bF + m)$	2.9M/0.1B
	HSGT [14]	$O(n + md^L)$	$O(LnF^2 + LmF)$	$O(nF + md^L)$	$O(Ln_b^2 + Ln_bF + Lm)$	2.4M/0.1B
	HubGT (ours)	$O(ns^3 + ms)$	$O(LnF^2 + Lns^2F)$	$O(nF + ns^2)$	$O(Ln_bF + n_bs^2)$	1.6M/0.1B

☺ *Decoupled mini-batch learning*

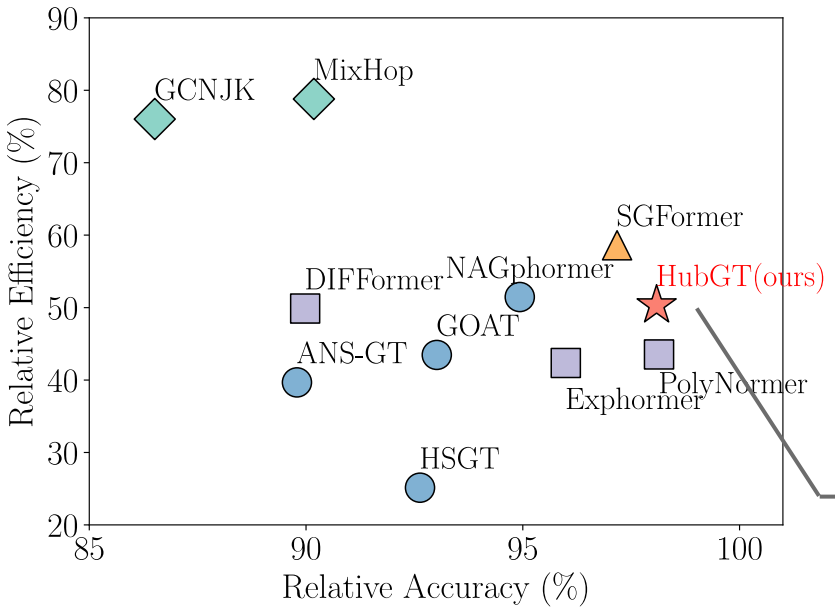
Results: Experimental Evaluation

- Effectiveness: Expressive under heterophily
- Efficiency: Faster precomputation and inference

Heterophilous	PENN94				GENIUS				TWITCH-GAMER				POKEC			
Scale	41,554 / 2,724,458				421,961 / 1,845,736				168,114 / 13,595,114				1,632,803 / 44,603,928			
Metrics	**Pre.**	**Epoch**	**Infer**	**Acc**	**Pre.**	**Epoch**	**Infer**	**Acc**	**Pre.**	**Epoch**	**Infer**	**Acc**	**Pre.**	**Epoch**	**Infer**	**Acc**
GCNJK	-	0.35	0.04	65.91±0.16	-	0.06	0.03	80.65±0.07	-	0.15	0.03	59.91±0.42	-	0.06	0.09	59.38±0.21
MixHop	-	0.31	0.04	75.00±0.37	-	0.08	0.01	80.63±0.04	-	0.11	0.01	61.80±0.01	-	0.15	0.03	64.02±0.02
SGFormer*	-	0.95	0.55	77.52±0.56	-	0.72	2.4	85.01±0.25	-	0.38	3.3	65.93±0.15	-	4.4	29	73.13±0.16
Exphormer	(OOM)				(OOM)				42	1.3	0.41	64.24±0.35	(OOM)			
DIFFormer*	-	0.53	0.65	61.77±3.41	-	0.77	5.5	84.52±0.36	-	0.61	5.1	60.81±0.44	-	4.6	15	73.89±0.35
PolyNormer*	-	0.58	18.4	**79.87**±0.06	-	0.77	28	85.64±0.52	-	1.5	89	64.72±0.65	-	4.2	67	**81.03**±0.08
NAGphormer	237	6.1	2.1	74.45±0.60	38	5.4	1.0	83.88±0.13	16	1.9	2.4	61.92±0.19	70	16.1	3.1	73.06±0.05
ANS-GT	3889	42	4.9	67.76±1.32	34092	37	5.0	67.76±1.32	12924	19	6.7	61.55±0.45	(OOM)			
GOAT	1332	33	18	71.42±0.44	2664	28	39	80.12±2.32	3348	37	63	61.38±0.83	3855	760	804	(TLE)
HSGT*	12	115	110	67.77±0.27	21	98	114	84.03±0.24	68	235	253	61.60±0.09	551	1420	1557	(TLE)
HubGT (ours)	7.3	3.4	0.29	78.13±0.43	125	23	2.5	**89.68**±0.48	49	12	1.2	**67.03**±2.17	5422	99	13	76.96±0.44

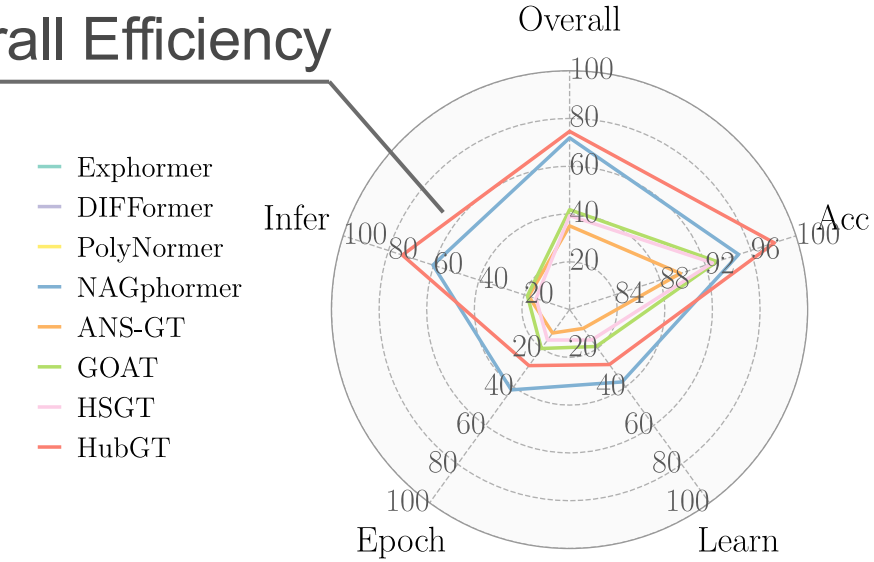
Results: Experimental Evaluation

- Effectiveness: Expressive under heterophily
- Efficiency: Faster precomputation and inference



Fast + top-tier Accuracy

Overall Efficiency



THANK YOU

Acknowledgments

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG3-RP-2024-034), Singapore MOE AcRF Tier-2 funding (MOE-T2EP20122-0003), and NTU startup grant (020948-00001). Ningyi Liao is supported by the Joint NTU-WeBank Research Centre on FinTech, Nanyang Technological University, Singapore. Zihao Yu is supported by A*STAR RIE2025 Manufacturing, Trade and Connectivity (MTC) Programmatic Fund (M24N6b0043) administered by A*STAR.