

HubGT: Fast Graph Transformer with Decoupled Hierarchy Labeling

Ningyi Liao*, Zihao Yu*, Siqiang Luo✉, Gao Cong

Nanyang Technological University, Singapore



MOTIVATION

Positional Encoding

$$\tilde{H}_i = \text{softmax} \left(\frac{Q_i K_i^T}{\sqrt{F_K}} + P \right) V_i$$

Positional Encoding

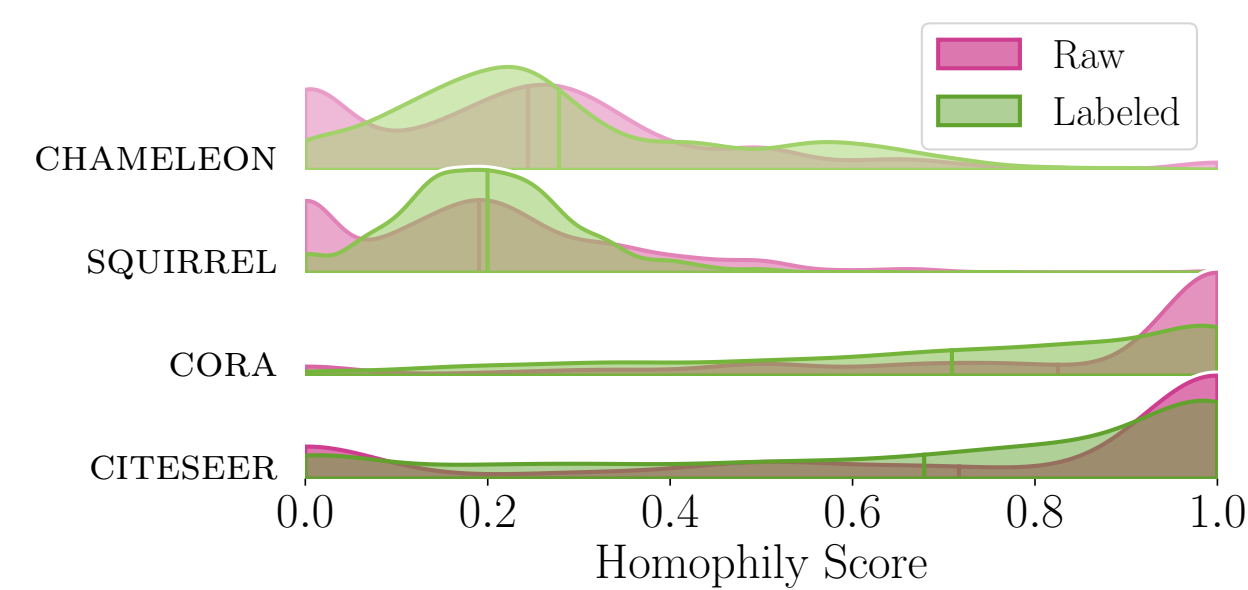
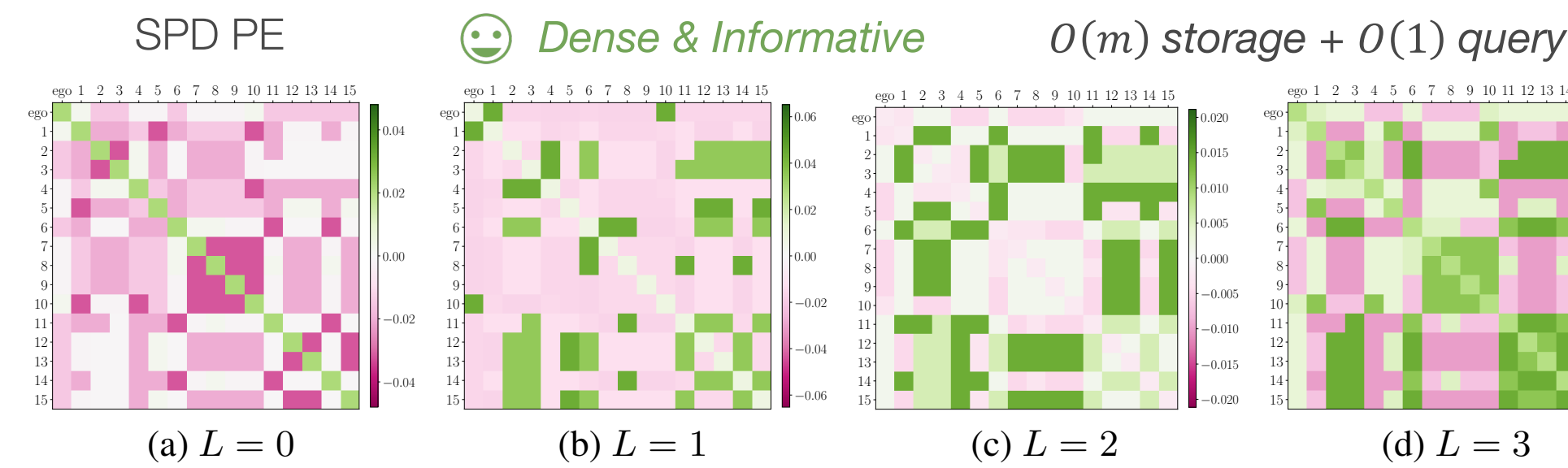
Node Feature

Transformer Attention

Adjacency PE

Only locality

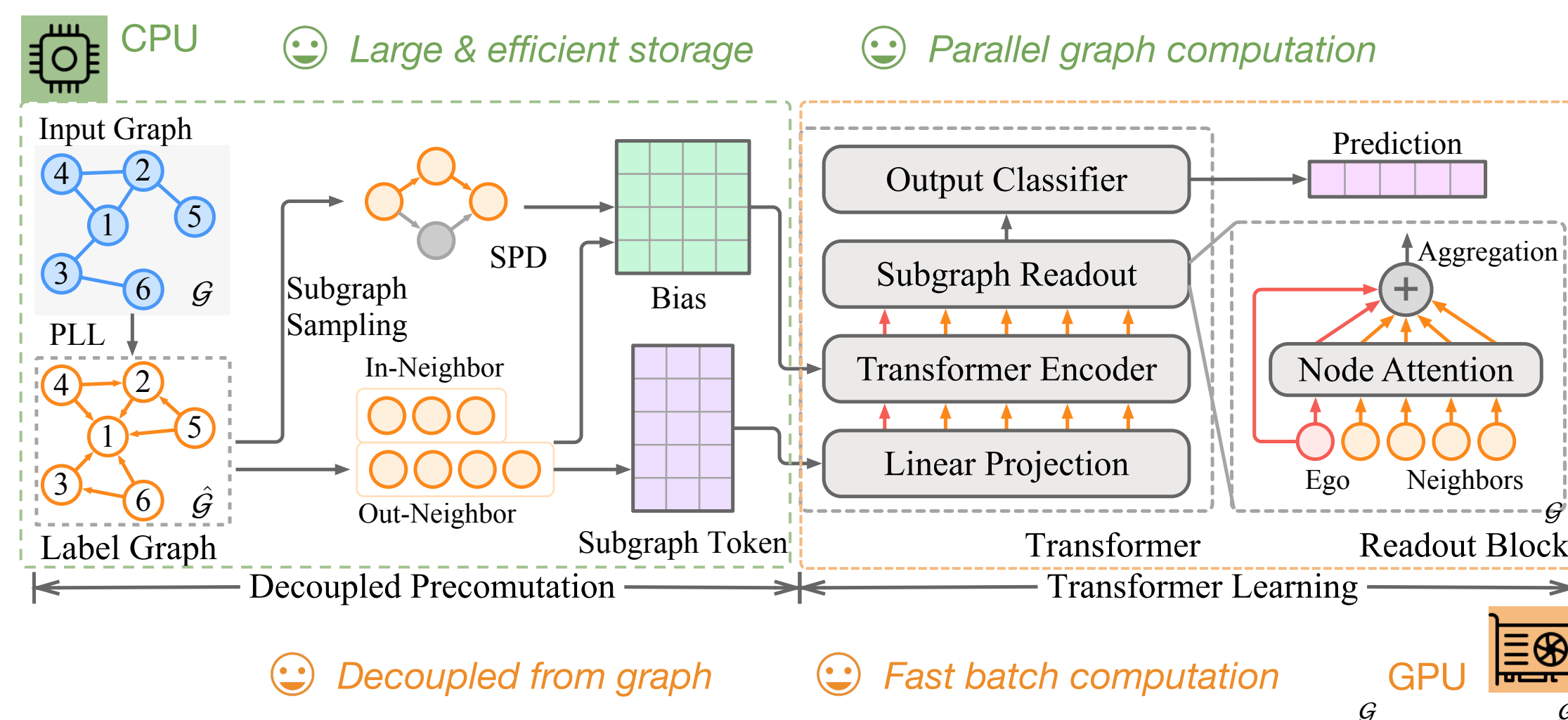
$O(n^2)$ storage



Graph Hierarchy

Gap between different graph adjacency

FRAMEWORK



METHOD

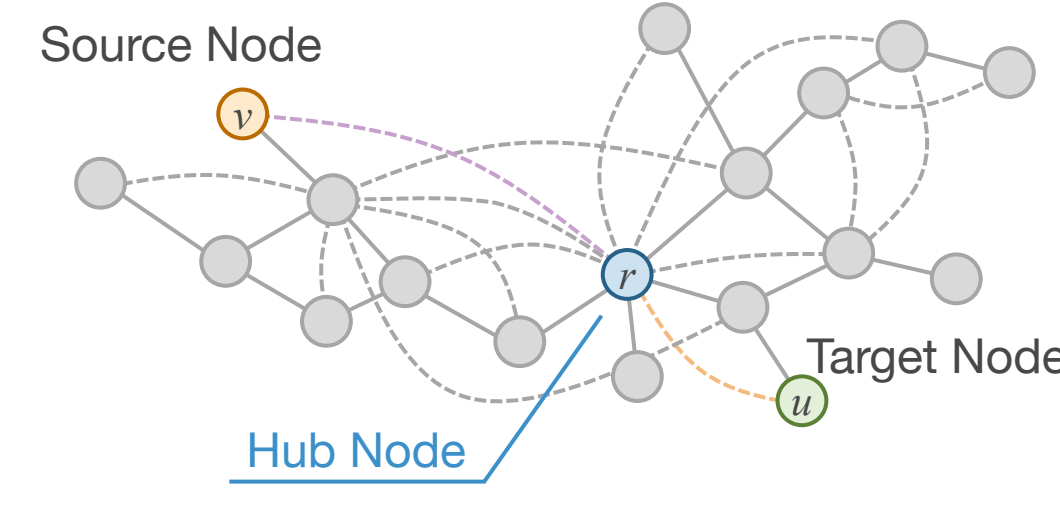
Hub Labeling

Single pair, Full graph

$$L(v) = (u, b(u, v))$$

Node Label

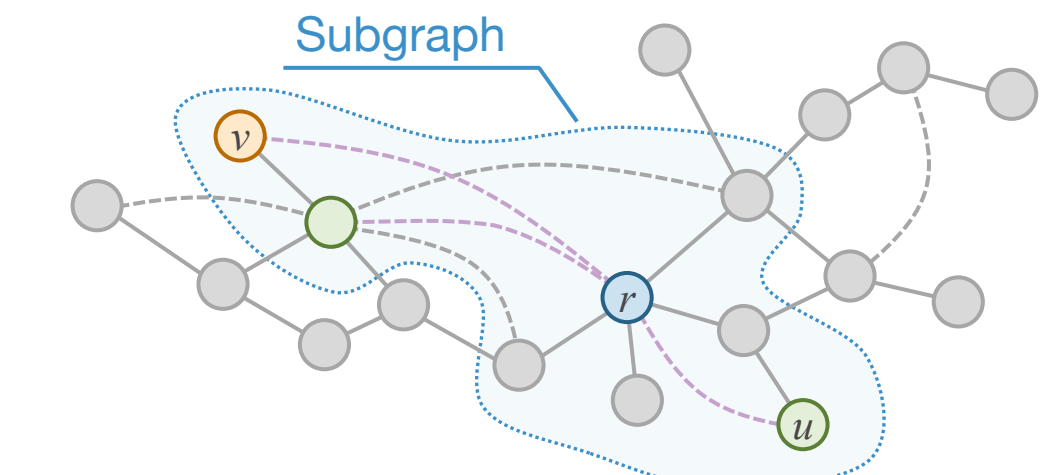
Shortest Distance



All pair, Subgraph

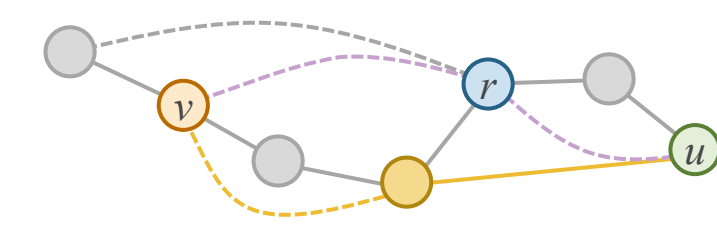
$$L(v) = (u, b(u, v), M_v^1(u), M_v^0(u))$$

Neighbor Lookup



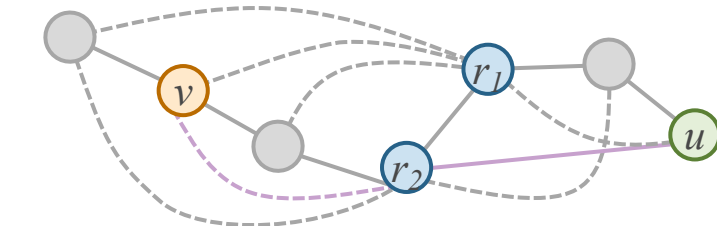
H-1: Local

$$b_r^{(1)}(u, v) = b(u, r) + b(r, v) - \delta_r(u, v)$$



H-2: Global

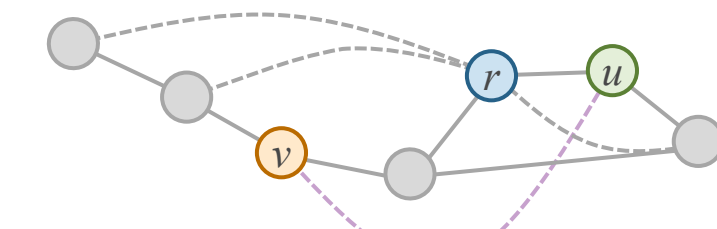
$$b_r^{(2)}(u, v) = \min_r b(u, r) + b(r, v) - \delta_r(u, v)$$



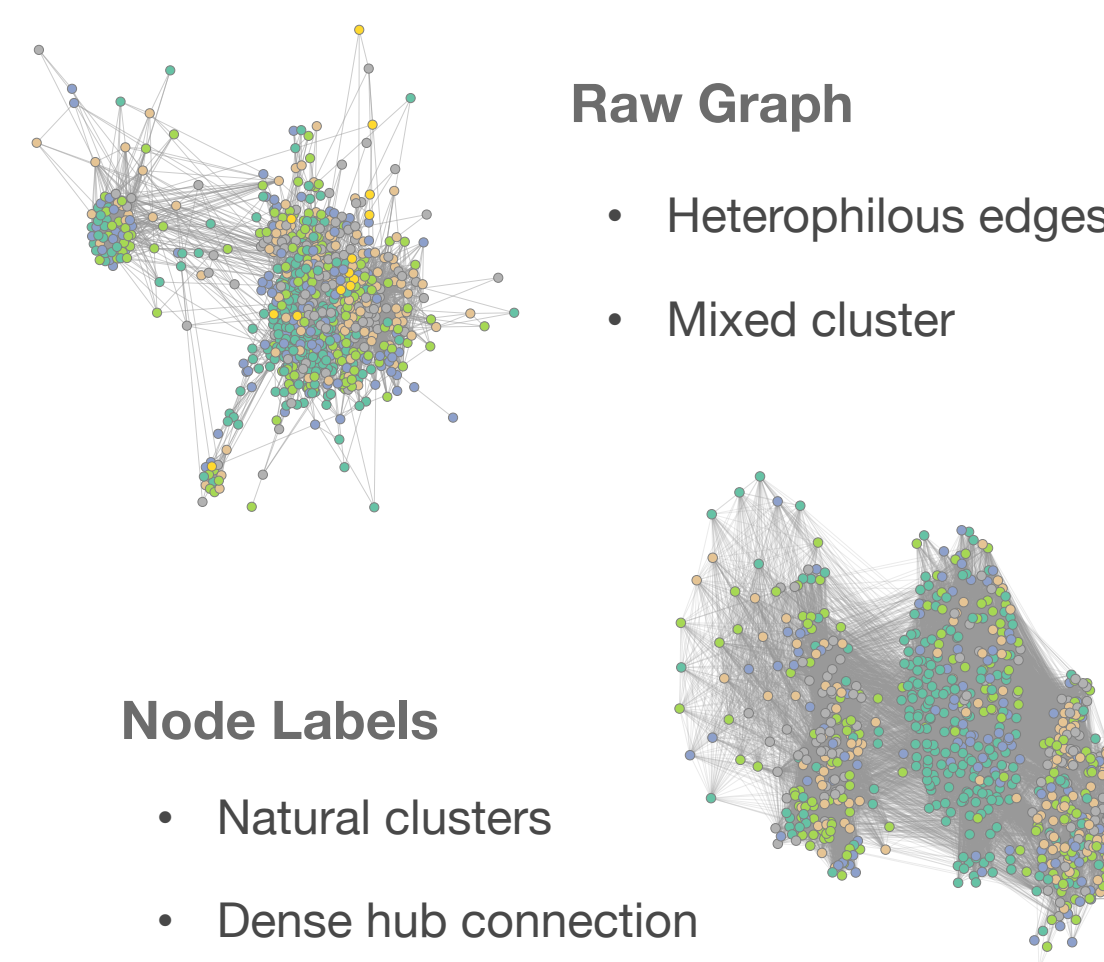
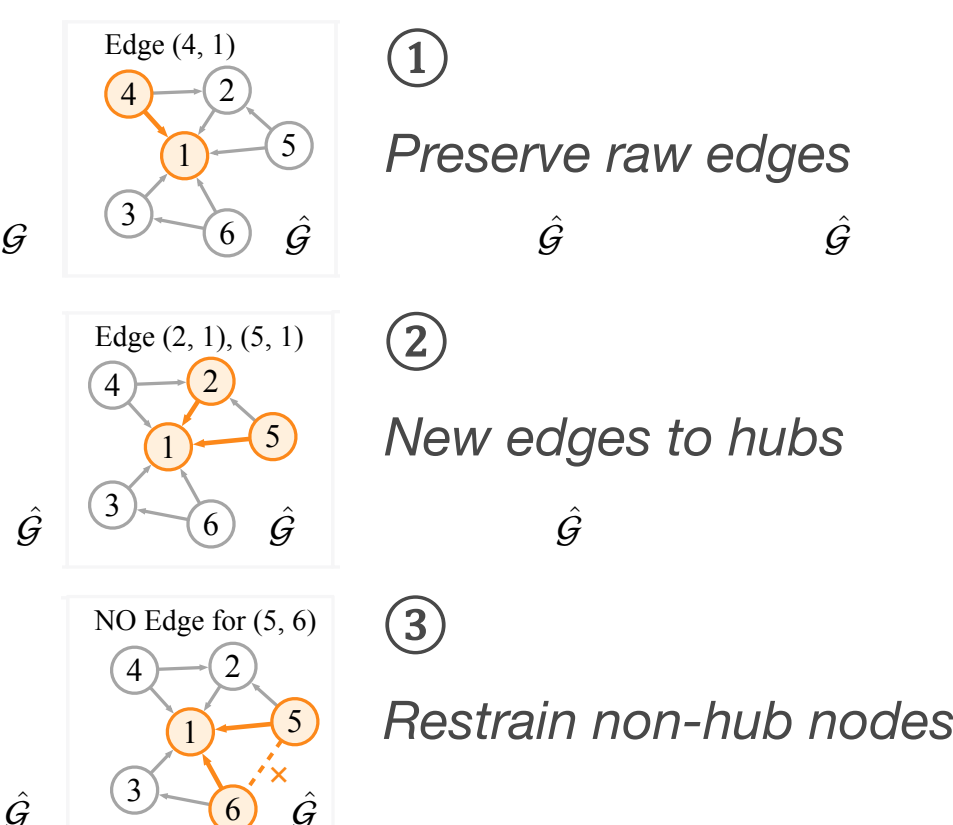
H-0: Cache

$$b^{(0)}(u, v) = I_v[u]$$

$$= \min_{w \in \mathcal{L}(u) \cap \mathcal{L}(v)} \{b(u, w) + b(w, v)\}$$



Subgraph Hierarchy



EXPERIMENTS

Complexity Analysis

Not scalable to large m and n

Insufficient memory

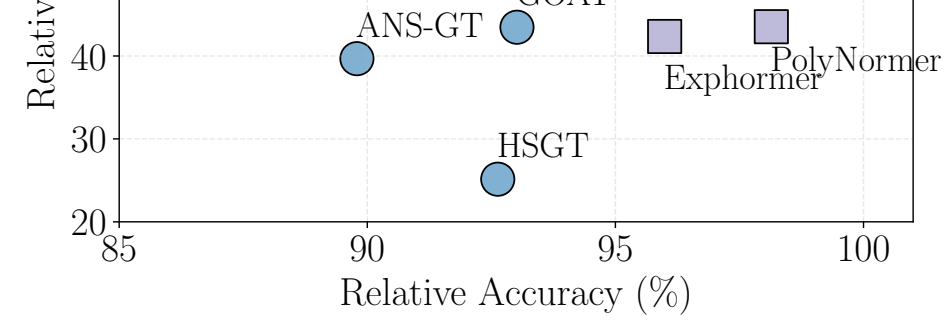
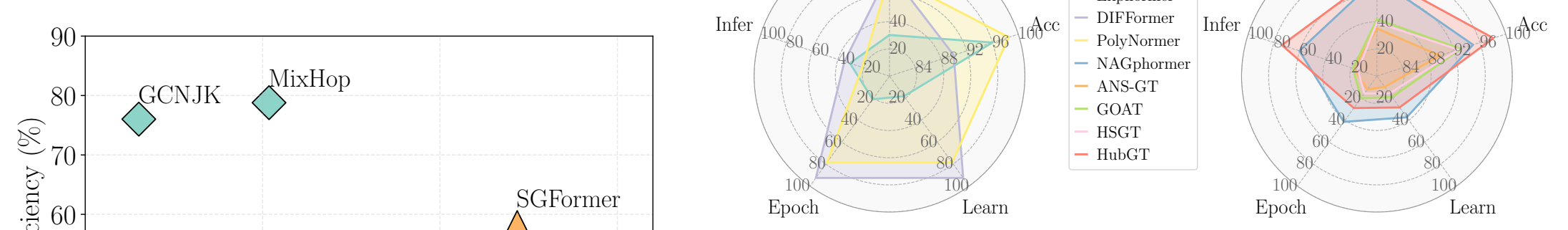
Taxonomy	Model	Precompute Time	Train Time	RAM Mem.	GPU Mem.	Scale
Vanilla (FB)	Graphormer [2]	$O(n^3)$	$O(Ln^2F)$	$O(n^2)$	$O(Ln^2F)$	0.3K/0.6K
	GRPE [18]	$O(n^2)$	$O(Ln^2F)$	$O(n^2)$	$O(Ln^2F)$	0.3K/0.6K
Kernel-based (NS)	GraphGPS [7]	$O(n^3)$	$O(LnF^2 + LmF)$	$O(nF + n^2)$	$O(Ln_bF + m)$	1.0K/3.0K
	Expformer [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
Hierarchical (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]	$O(ns^2 + md^L)$	$O(LnF^2 + Lns^2F + Lm)$	$O(nF + ns^2 + md^L)$	$O(Ln_bF + n_b s^2)$	20K/0.2M
	GOAT [10]	$O(nF)$	$O(LnF^2 + LmF)$	$O(nF + m)$	$O(Ln_b^2F + Lm)$	2.9M/0.1B
	HSGT [14]	$O(n + md^L)$	$O(LnF^2 + LmF)$	$O(nF + md^L)$	$O(Ln_b^2F + Lm)$	2.4M/0.1B
HubGT (ours)		$O(ns^3 + ms)$	$O(LnF^2 + Lns^2F)$	$O(nF + ns^2)$	$O(Ln_bF + n_b s^2)$	1.6M/0.1B

Empirical Comparison

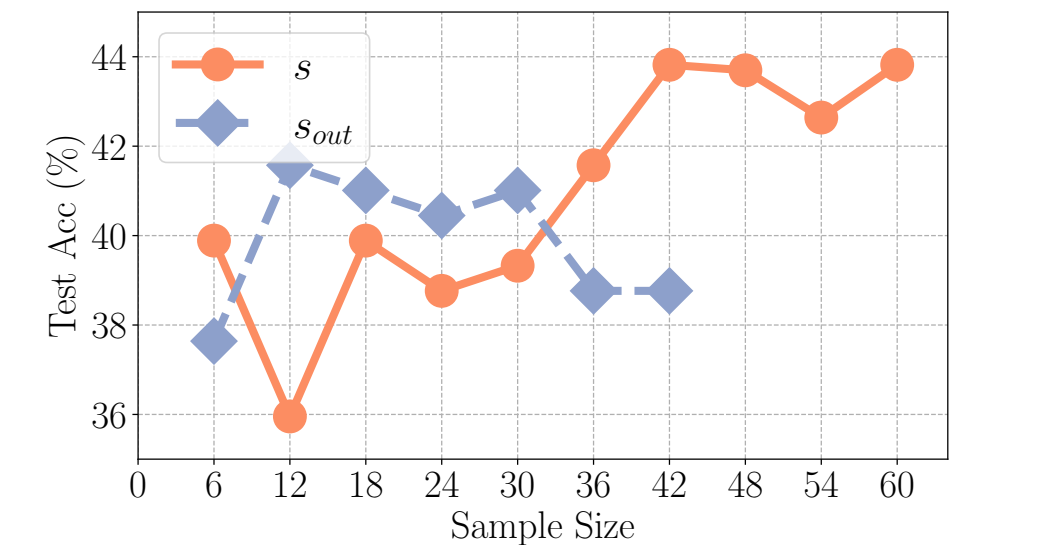
Heterophilous Scale	PENN94				GENIUS				TWITCH-GAMER				POKEC			
	Pre.	Epoch	Infer	Acc	Pre.	Epoch	Infer	Acc	Pre.	Epoch	Infer	Acc	Pre.	Epoch	Infer	Acc
GCNJK	-	0.35	0.04	65.91±0.16	-	0.06	0.03	80.65±0.07	-	0.15	0.03	59.91±0.42	-	0.06	0.09	59.38±0.21
MixHop	-	0.31	0.04	75.00±0.37	-	0.08	0.01	80.63±0.04	-	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
Expformer	(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±0.52	-	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

* Inference of these models is performed on the CPU in a full-batch manner due to their requirement of the whole graph.

Efficiency Evaluation



Parameter Study



[Paper]

neurips.cc/virtual/2025/poster/118302



[Code]

github.com/gdmnl/HubGT



[Author]

nyliao.github.io

