

# Capacity-Agnostic Parameter Isolation for Continual Graph Learning

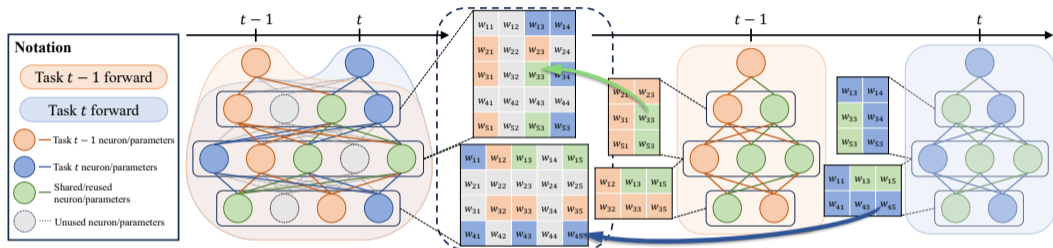
Ye Xiao<sup>1</sup>   Ruikun Li<sup>2</sup>   Zhenyu Yang<sup>3</sup>   Andrey Vasnev<sup>1</sup>   Junbin Gao<sup>1</sup>

<sup>1</sup>Discipline of Business Analytics  
The University of Sydney, Sydney, Australia

<sup>2</sup>Financial Technology Thrust  
The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

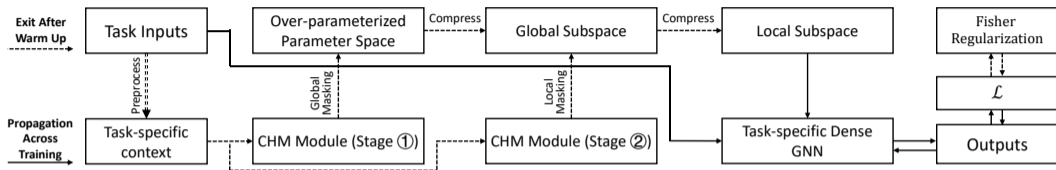
<sup>3</sup>School of Computing  
Macquarie University, Sydney, Australia

ICML 2026, Seoul



## Key Problems

- ▶ Assign each task a subnetwork to avoid forgetting, but most methods still run **full-network propagation**.
- ▶ Masked parameters are **unused** logically, but still costly computationally.
- ▶ These computational overheads will be amplified under **large model capacity**.



## Targets

- ▶ Train and infer only on decoupled task-specific subnetworks.
- ▶ Keep fine-grained parameter use for knowledge transfer and greater scalability.

### 1. Context as prior

Graph context guides subnetwork construction for each task.

### 2. Brief warm-up

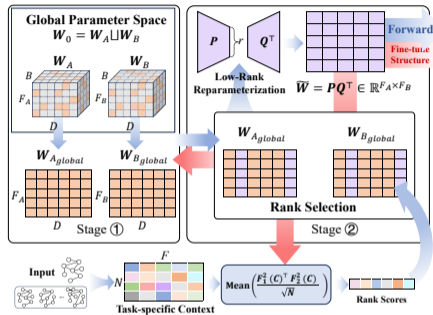
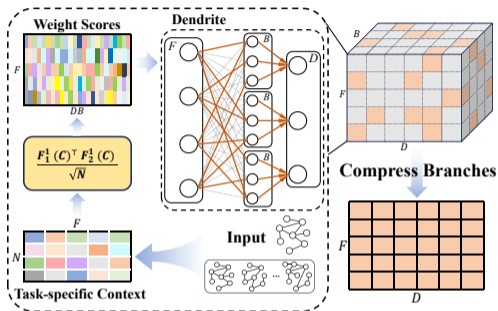
CHM searches a structured subnetwork, refined by Fisher regularization.

### 3. Decoupled training

After warm-up, CAGNN trains and infers as a compact dense GNN.



# Context-aware Hierarchical Masking (CHM)



## Stage 1: dendritic masking

Select one branch for each input-output pair using task-specific context.

## Stage 2: low-rank masking

Select rank factors  $P, Q$  so the final weight is compact:  $\tilde{W} = PQ^T$ .

**Fisher regularization** encourages reuse of less sensitive old parameters, reducing interference while preserving knowledge transfer.

DATASET (# Task)	Node Classification										Graph Classification					
	CORA (3)		CITSEER (3)		CORAFULL (35)		COMPUTERS (5)		OGB-ARXIV (8)		MNIST (5)		CIFAR10 (5)		AROMATICITY (10)	
	AP	AF	AP	AF	AP	AF	AP	AF	AP	AF	AP	AF	AP	AF	AP	AF
Bare	90.28 ± 2.2	2.1 ± 1.7	80.43 ± 3.6	5.3 ± 1.9	79.79 ± 3.5	19.8 ± 4.2	96.18 ± 1.9	4.2 ± 1.1	70.62 ± 6.7	21.7 ± 6.6	56.24 ± 12.8	51.1 ± 9.7	55.30 ± 3.7	35.5 ± 6.2	43.26 ± 3.1	23.6 ± 2.7
LwF	92.17 ± 1.1	1.5 ± 1.0	86.63 ± 0.5	1.8 ± 0.7	OOM	OOM	95.92 ± 1.7	6.9 ± 2.0	80.21 ± 4.1	12.9 ± 4.8	72.25 ± 7.7	31.5 ± 8.2	72.18 ± 4.2	6.7 ± 3.9	49.33 ± 2.1	16.5 ± 2.0
EWC	90.39 ± 1.9	2.0 ± 1.2	85.43 ± 1.1	2.7 ± 1.2	90.68 ± 2.9	5.7 ± 2.1	97.22 ± 0.8	3.6 ± 1.2	88.52 ± 2.0	3.6 ± 1.7	72.30 ± 5.4	23.9 ± 5.1	71.02 ± 2.9	6.0 ± 2.7	55.47 ± 1.4	6.4 ± 1.7
TWP	90.52 ± 1.3	3.3 ± 2.1	86.07 ± 0.7	2.2 ± 0.9	89.35 ± 3.3	8.9 ± 3.5	97.90 ± 0.6	2.0 ± 0.8	87.45 ± 2.4	3.8 ± 2.1	76.20 ± 7.8	18.2 ± 7.0	72.18 ± 3.7	6.7 ± 3.3	57.25 ± 2.0	6.4 ± 2.2
ERGNN	86.92 ± 4.2	3.7 ± 3.1	86.52 ± 0.7	3.6 ± 1.1	94.39 ± 2.3	3.8 ± 2.7	97.83 ± 0.6	0.5 ± 0.2	89.96 ± 1.3	2.6 ± 1.5	N/A	N/A	N/A	N/A	N/A	N/A
PackNet	93.79 ± 0.7	0.0 ± 0.0	85.66 ± 1.8	0.0 ± 0.0	97.14 ± 0.7	0.0 ± 0.0	98.17 ± 0.2	0.0 ± 0.0	90.93 ± 1.1	0.0 ± 0.0	96.05 ± 1.1	0.0 ± 0.0	80.17 ± 0.7	0.0 ± 0.0	41.08 ± 3.3	0.0 ± 0.0
Piggyback	93.87 ± 0.9	0.0 ± 0.0	87.33 ± 0.4	0.0 ± 0.0	96.39 ± 0.7	0.0 ± 0.0	98.23 ± 0.3	0.0 ± 0.0	92.07 ± 0.5	0.0 ± 0.0	83.12 ± 3.0	0.0 ± 0.0	76.17 ± 1.9	0.0 ± 0.0	41.64 ± 2.7	0.0 ± 0.0
HAT	94.01 ± 1.2	2.4 ± 1.5	85.58 ± 1.2	0.6 ± 0.3	91.79 ± 2.1	7.0 ± 1.4	97.92 ± 0.5	0.4 ± 0.0	90.07 ± 0.4	1.2 ± 0.5	59.91 ± 6.9	46.4 ± 6.8	54.87 ± 4.6	36.3 ± 4.4	54.28 ± 1.9	11.9 ± 2.5
TAAM	94.20 ± 1.1	0.0 ± 0.0	87.99 ± 0.4	0.0 ± 0.0	97.21 ± 1.9	0.0 ± 0.0	98.64 ± 1.1	0.0 ± 0.0	86.27 ± 1.2	0.0 ± 0.0	95.26 ± 1.4	0.0 ± 0.0	70.99 ± 4.2	0.0 ± 0.0	53.51 ± 1.9	0.0 ± 0.0
WSN	94.06 ± 1.1	0.0 ± 0.0	87.38 ± 0.3	0.0 ± 0.0	97.28 ± 0.8	0.0 ± 0.0	98.24 ± 0.2	0.0 ± 0.0	90.28 ± 0.9	0.0 ± 0.0	95.92 ± 1.6	0.0 ± 0.0	80.02 ± 1.7	0.0 ± 0.0	47.15 ± 2.0	0.0 ± 0.0
CAGCN	94.68 ± 0.7	0.0 ± 0.0	87.63 ± 0.3	0.0 ± 0.0	97.32 ± 1.2	0.0 ± 0.0	98.01 ± 0.6	0.0 ± 0.0	91.55 ± 0.7	0.0 ± 0.0	96.50 ± 1.6	0.0 ± 0.0	81.13 ± 1.5	0.0 ± 0.0	49.70 ± 1.1	0.0 ± 0.0
CAGAT	95.28 ± 1.6	0.0 ± 0.0	88.28 ± 0.7	0.0 ± 0.0	97.80 ± 0.9	0.0 ± 0.0	98.32 ± 0.2	0.0 ± 0.0	87.91 ± 1.7	0.0 ± 0.0	96.41 ± 1.1	0.0 ± 0.0	79.81 ± 1.8	0.0 ± 0.0	49.96 ± 1.5	0.0 ± 0.0
Joint	96.02 ± 1.6	0.0 ± 0.0	89.06 ± 2.6	0.0 ± 0.0	98.45 ± 2.3	0.0 ± 0.0	99.09 ± 0.4	0.0 ± 0.0	91.62 ± 2.6	0.0 ± 0.0	97.58 ± 0.7	0.0 ± 0.0	81.65 ± 1.5	0.0 ± 0.0	63.49 ± 3.4	0.0 ± 0.0

<sup>1</sup> OOM indicates out-of-memory during training.

<sup>2</sup> N/A indicates not applicable.

## Average runtime per epoch

MODEL	OGB-ARXIV		MNIST		CIFAR10	
	Time (ms)	Ratio	Time (ms)	Ratio	Time (ms)	Ratio
WSN	787	1.00×	9,611	1.00×	10,325	1.00×
PackNet	335	0.43×	6,419	0.67×	6,922	0.67×
Piggyback	299	0.38×	6,831	0.71×	7,114	0.69×
HAT	343	0.44×	6,590	0.68×	7,068	0.68×
CAGNN (0.95)	<b>215</b>	<b>0.27×</b>	<b>3,345</b>	<b>0.35×</b>	<b>2,961</b>	<b>0.29×</b>
CAGNN (0.99)	<b>197</b>	<b>0.25×</b>	<b>3,101</b>	<b>0.32×</b>	<b>2,783</b>	<b>0.27×</b>

## Memory usage across tasks (MB)

MODEL	$\mathcal{T}_0$	$\mathcal{T}_{35}$	$\mathcal{T}_{70}$	$\mathcal{T}_{105}$	$\mathcal{T}_{140}$
WSN	1918.5	3744.9	5508.2	7270.7	8980.4
CAGNN	3121.5	3966.6	4750.2	5558.7	<b>6296.2</b>

## Key Contributions

- ▶ Identify **capacity-driven efficiency degradation** in continual graph learning.
- ▶ Construct **context-aware structured subnetworks** via CHM.
- ▶ Enable **decoupled training and inference** in a dense subnetwork after a short warm-up.
- ▶ Preserve **fine-grained** cross-tasks knowledge transfer (Parameter reuse)

# Thank you for Listening!

Email: [ye.xiao@sydney.edu.au](mailto:ye.xiao@sydney.edu.au)