

FROM STARS TO SUBGRAPHS: UPLIFTING ANY GNN WITH LOCAL STRUCTURE AWARENESS

INTRODUCTION

Message-passing based Graph Neural Networks (GNN) are appealing for being efficient and scalable, however their expressiveness is upperbounded by the 1st-order Weisfeiler-Leman isomorphism test (1-WL). Prior works propose highly expressive models at the cost of scalability and generalization performance.

Our Contribution:

- we introduce a general framework to uplift any message-passing GNN to be more expressive, with limited scalability overhead and greatly improved practical performance
- we show that our framework is strictly more powerful than 1&2-WL, and is not less powerful than 3-WL
- we also design a sampling strategy to greatly reduce memory footprint and improve speed while maintaining performance

BACKGROUND

• Weisfeiler-Lehman Isomorphism Test(1-WL) 1-WL iteratively recolors nodes. At *t*-th iteration every node follows: $c^{(t)}(v) = \text{HASH}\left(c^{(t-1)}(v), \{c^{(t-1)}(u) | u \in \mathcal{N}_v\}\right)$



As visualized above, the recoloring formulation is equivalent to hashing the "star" around a node to the node's new color

Botteckneck of hashing "star"

Example: two non-isomorphic regular graphs





All stars are the same, 1-WL fails to distinguish them

LINGXIAO ZHAO, WEI JIN, LEMAN AKOGLU, NEIL SHAH

FROM STARS TO SUBGRAPHS

• Subgraph is better than star



1-hop egonet based subgraphs in A and B are different. Hashing the subgraph of a node to its color can successfully distinguish A and B.

• Design 1: Subgraph-1-WL



• Design 2: Subgraph-1-WL*



Neural version: GNNAsKernel (GNN-AK)



- Theorems of expressiveness
 - Strictly more powerful than 1&2-WL
 - No less powerful than 3-WL
 - For any k>=3, exist a pair of k-WL-failed graphs that cannot be distinguished by Subgraph-1-WL



SUBGRAPHDROP

To further improve scalability and practicality, we propose a subgraph sampling training method motivated from Dropout.

• Training: at every epoch, randomly drop (a large amount of) subgraphs.

• Testing: use ALL subgraphs.

It speeds up training and reduce memory cost 5 to 10 times without sacrificing any performance.

Method GCN GCN-AK⁺ GIN **GIN-AK GIN-AK**⁻ PNA* $\mathsf{PNA}^*-\mathsf{AK}^+$ PPGN PPGN-AK+ 100^c

EXPERIMENTS ON REAL-WORLD DATASETS

Method	ZINC-12K (MAE)	CIFAR10 (ACC)	PATTERN (ACC)	MolHIV (ROC)	MolPCBA (AP)
GatedGCN HIMP PNA DGN GSN CIN	$\begin{array}{c} 0.363 \pm 0.009 \\ 0.151 \pm 0.006 \\ 0.188 \pm 0.004 \\ 0.168 \pm 0.003 \\ 0.115 \pm 0.012 \\ \textbf{0.079} \pm \textbf{0.006} \end{array}$	$\begin{array}{c} 69.37 \pm 0.48 \\ - \\ 70.47 \pm 0.72 \\ 72.84 \pm 0.42 \\ - \\ - \\ - \end{array}$	$\begin{array}{c} 84.480 \pm 0.122 \\ - \\ 86.567 \pm 0.075 \\ 86.680 \pm 0.034 \\ - \\ - \\ - \end{array}$	$\begin{array}{c} - \\ 0.7880 \pm 0.0082 \\ 0.7905 \pm 0.0132 \\ 0.7970 \pm 0.0097 \\ 0.7799 \pm 0.0100 \\ \textbf{0.8094} \pm \textbf{0.0057} \end{array}$	$\begin{array}{c} - \\ 0.2838 \pm 0.0035 \\ 0.2885 \pm 0.0030 \\ - \\ - \\ - \end{array}$
GCN GCN-AK+	$\begin{array}{c} 0.321 \pm 0.009 \\ 0.127 \pm 0.004 \end{array}$	$\begin{array}{c} 58.39 \pm 0.73 \\ 72.70 \pm 0.29 \end{array}$	$\begin{array}{r} 85.602 \pm 0.046 \\ \textbf{86.887} \pm \textbf{0.009} \end{array}$	$\begin{array}{c} 0.7422 \pm 0.0175 \\ 0.7928 \pm 0.0101 \end{array}$	$\begin{array}{c} 0.2385 \pm 0.0019 \\ 0.2846 \pm 0.0002 \end{array}$
GIN GIN-AK GIN-AK ⁺	$\begin{array}{c} 0.163 \pm 0.004 \\ 0.094 \pm 0.005 \\ \textbf{0.080} \pm \textbf{0.001} \end{array}$	$59.82 \pm 0.33 \\ 67.51 \pm 0.21 \\ 72.19 \pm 0.13$	$\begin{array}{r} 85.732 \pm 0.023 \\ 86.803 \pm 0.044 \\ 86.850 \pm 0.057 \end{array}$	$\begin{array}{c} 0.7881 \pm 0.0119 \\ 0.7829 \pm 0.0121 \\ 0.7961 \pm 0.0119 \end{array}$	$\begin{array}{c} 0.2682 \pm 0.0006 \\ 0.2740 \pm 0.0032 \\ \textbf{0.2930} \pm \textbf{0.0044} \end{array}$
PNA* PNA*-AK+	$\begin{array}{c} 0.140 \pm 0.006 \\ 0.085 \pm 0.003 \end{array}$	73.11 ± 0.11 OOM	85.441 ± 0.009 OOM	$\begin{array}{c} 0.7905 \pm 0.0102 \\ 0.7880 \pm 0.0153 \end{array}$	$\begin{array}{c} 0.2737 \pm 0.0009 \\ 0.2885 \pm 0.0006 \end{array}$
GCN-AK ⁺ -S GIN-AK ⁺ -S PNA*-AK ⁺ -S	$\begin{array}{c} 0.127 \pm 0.001 \\ 0.083 \pm 0.001 \\ 0.082 \pm 0.000 \end{array}$	71.93 ± 0.47 72.39 ± 0.38 74.79 \pm 0.18	86.805 ± 0.046 86.811 ± 0.013 86.676 ± 0.022	$\begin{array}{c} 0.7825 \pm 0.0098 \\ 0.7822 \pm 0.0075 \\ 0.7821 \pm 0.0143 \end{array}$	$\begin{array}{c} 0.2840 \pm 0.0036 \\ 0.2916 \pm 0.0029 \\ 0.2880 \pm 0.0012 \end{array}$



VARIFY EXPRESSIVENESS

EXP contains 1-WL failed graphs; SR25 contains 3-WL failed graphs; Substructure and graph property are performed on random graphs.

P SR25 C) (ACC)	SR25	Counting Substructures (MAE)				Graph Properties $(\log_{10}(MAE))$		
	(ACC)	Triangle	Tailed Tri.	Star	4-Cycle	IsConnected	Diameter	Radius
70 1	6.67%	0.4186	0.3248	0.1798	0.2822	-1.7057	-2.4705	-3.9316
0 77	6.670	0.0137	0.0134	0.0174	0.0185	-2.6705	-3.9102	-5.1150
10 70	6.67%	0.3369 0.0934	0.2373	0.0224 0.0168	0.2185	-1.9239 -1.9934	-3.7573	-4.7584
6	6.67%	0.0123	0.0112	0.0150	0.0126	-2.7513	-3.9687	-5.1846
10 10	6.67% 6.67%	$0.3532 \\ 0.0118$	$0.2648 \\ 0.0138$	0.1278 0.0166	0.2430 0.0132	-1.9395 -2.6189	-3.4382 -3.9011	-4.9470 -5.2026
10 10	6.67% 100%	0.0089 OOM	0.0096 OOM	0.0148 OOM	0.0090 OOM	-1.9804 OOM	-3.6147 OOM	-5.0878 OOM