Automatic Relation-aware Graph Network Proliferation

Shaofei Cai[1,2], Liang Li[1], Xinzhe Han[1,2], Jiebo Luo[3], Zhengjun Zha[4], Qingming Huang[1,2,5]

1. Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS, Beijing, China

2. University of Chinese Academy of Sciences, Beijing, China, 3. University of Rochester

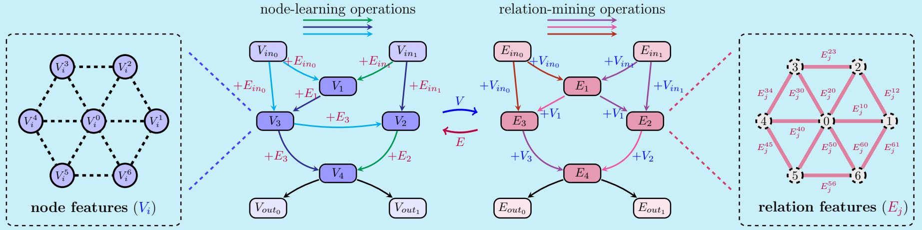
4. University of Science and Technology of China, China, 5. Peng Cheng Laboratory, Shenzhen, China

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

1. Abstract

We propose Automatic Relation-aware Graph Network Proliferation (ARGNP) for efficiently searching GNNs with a relation-guided message passing mechanism. The experiments on six datasets for four graph learning tasks demonstrate that GNNs produced by our method are superior to the current state-of-the-art hand crafted and search-based GNNs.

2. Relation-aware GNN Search Space



A. Node Search Space   B. Relation Search Space

The proposed dual relation-aware graph search space comprises:

1. **Node-learning operations** implement the anisotropic message aggregation under the guidance of relation features.

2. **Relation-mining operations** extract relational information hidden in each pair of edge-connected nodes.

<u>This realizes the organic combination of node learning and relation mining in message passing mechanism.</u>

3. Ablation Study on Search Space

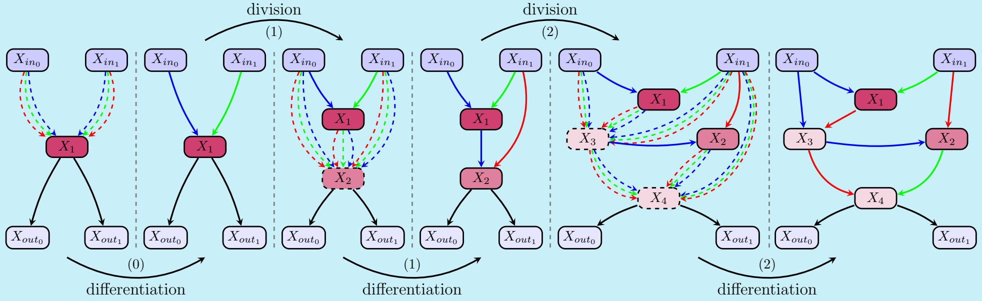
Architecture	E ☑	Node Level: CLUSTER Metric (AA %) ↑	Params (M)	Search (day)	Graph Level: ZINC Metric (MAE) ↓	Params (M)	Search (day)	Graph Level: CIFAR10 Metric (OA %) ↑	Params (M)	Search (day)	Edge Level: TSP Metric (F1) ↑	Params (M)	Search (day)
GCN [29]	×	$68.50_{\pm 0.98}$	0.50	ⓜ	$0.367_{\pm 0.011}$	0.50	ⓜ	$56.34_{\pm 0.38}$	0.10	ⓜ	$0.630_{\pm 0.001}$	0.10	ⓜ
GIN [59]	×	$64.72_{\pm 1.55}$	0.52	ⓜ	$0.526_{\pm 0.051}$	0.51	ⓜ	$55.26_{\pm 1.53}$	0.10	ⓜ	$0.656_{\pm 0.003}$	0.10	ⓜ
GraphSage [21]	×	$63.84_{\pm 0.11}$	0.50	ⓜ	$0.398_{\pm 0.002}$	0.51	ⓜ	$65.77_{\pm 0.31}$	0.10	ⓜ	$0.665_{\pm 0.003}$	0.10	ⓜ
GAT [53]	×	$70.59_{\pm 0.45}$	0.53	ⓜ	$0.384_{\pm 0.007}$	0.53	ⓜ	$64.22_{\pm 0.46}$	0.11	ⓜ	$0.671_{\pm 0.002}$	0.10	ⓜ
GatedGCN [9]	✓	$76.08_{\pm 0.34}$	0.50	ⓜ	$0.214_{\pm 0.013}$	0.51	ⓜ	$67.31_{\pm 0.31}$	0.10	ⓜ	$0.838_{\pm 0.002}$	0.53	ⓜ
PNA [15]	×	N/A	N/A	N/A	$0.320_{\pm 0.032}$	0.39	ⓜ	$70.46_{\pm 0.44}$	0.11	ⓜ	N/A	N/A	N/A
PNA [15]	✓	N/A	N/A	N/A	$0.188_{\pm 0.004}$	0.39	ⓜ	$70.47_{\pm 0.72}$	0.11	ⓜ	N/A	N/A	N/A
DGN [5]	×	N/A	N/A	N/A	$0.219_{\pm 0.010}$	0.39	ⓜ	$72.70_{\pm 0.54}$	0.11	ⓜ	N/A	N/A	N/A
DGN [5]	✓	N/A	N/A	N/A	$0.168_{\pm 0.003}$	0.39	ⓜ	$72.84_{\pm 0.42}$	0.11	ⓜ	N/A	N/A	N/A
GNAS-MP [12]	×	$74.77_{\pm 0.15}$	1.61	0.80	$0.242_{\pm 0.005}$	1.20	0.40	$70.10_{\pm 0.44}$	0.43	3.20	$0.742_{\pm 0.002}$	1.20	2.10
ARGNP (2)	×	$61.61_{\pm 0.27}$	0.07	0.04	$0.430_{\pm 0.003}$	0.09	0.01	$66.55_{\pm 0.13}$	0.10	0.11	$0.655_{\pm 0.003}$	0.09	0.05
ARGNP (4)	×	$64.06_{\pm 0.45}$	0.14	0.07	$0.303_{\pm 0.013}$	0.14	0.01	$66.65_{\pm 0.39}$	0.18	0.14	$0.668_{\pm 0.003}$	0.17	0.06
ARGNP (8)	×	$68.73_{\pm 0.12}$	0.25	0.20	$0.239_{\pm 0.009}$	0.27	0.02	$67.37_{\pm 0.32}$	0.33	0.48	$0.674_{\pm 0.002}$	0.29	0.21
ARGNP (16)	×	$71.92_{\pm 0.29}$	0.53	0.71	$0.221_{\pm 0.004}$	0.51	0.06	$67.10_{\pm 0.51}$	0.58	1.77	$0.684_{\pm 0.002}$	0.56	0.76
ARGNP (2)	✓	$64.99_{\pm 0.31}$	0.08	0.06	$0.318_{\pm 0.009}$	0.08	0.01	$69.14_{\pm 0.30}$	0.10	0.17	$0.773_{\pm 0.001}$	0.08	0.08
ARGNP (4)	✓	$74.75_{\pm 0.25}$	0.15	0.09	$0.197_{\pm 0.006}$	0.15	0.01	$71.83_{\pm 0.32}$	0.17	0.23	$0.821_{\pm 0.001}$	0.14	0.10
ARGNP (8)	✓	$76.32_{\pm 0.03}$	0.29	0.31	$0.155_{\pm 0.003}$	0.28	0.04	$73.72_{\pm 0.32}$	0.33	0.84	$0.841_{\pm 0.001}$	0.30	0.39
ARGNP (16)	✓	**$77.35_{\pm 0.05}$**	0.52	1.10	**$0.136_{\pm 0.002}$**	0.52	0.15	**$73.90_{\pm 0.15}$**	0.64	2.95	**$0.855_{\pm 0.001}$**	0.62	1.23

1. Mining relational information can significantly improve the GNN's reasoning ability.

2. Relation-aware GNN search space achieves higher score with fewer parameters.

CLUSTER: **77.35%**; CIFAR10: **73.90%**; ZINC-100k: **0.136**; TSP: **0.855**

4. Network Proliferation Search Paradigm



The network proliferation is an iterative process, which consists of:

1. **Network division** divides each feature vertex into two parts and constructs a series of local super-networks.

2. **Network differentiation** aims to differentiate each local super-network into a specific sub-network.

<u>The spatial-temporal complexity is reduced from $O(n^2)$ to $O(n)$.</u>

5. Ablation Study on Search Paradigm

#	Method	L (#)	ZINC: Search Strategy	Cell ☑	NPSP ☑	Metric (MAE) ↓	Params (M)	Search (Day)
1	R-space	8	Random	×	×	$0.303_{\pm 0.058}$	0.27	0.
2	R-space	8	DARTS	✓	×	$0.160_{\pm 0.005}$	0.28	0.17
3	R-space	8	DARTS	×	×	$0.157_{\pm 0.008}$	0.28	0.30
4	R-space	8	DARTS	×	✓	**$0.150_{\pm 0.006}$**	0.29	**0.08**
5	R-space	8	SGAS	✓	×	$0.165_{\pm 0.008}$	0.30	0.13
6	R-space	8	SGAS	×	×	$0.161_{\pm 0.008}$	0.30	0.25
7	R-space	8	SGAS	×	✓	**$0.155_{\pm 0.003}$**	0.28	**0.06**
8	R-space	16	Random	×	×	$0.185_{\pm 0.024}$	0.51	0.
9	R-space	16	DARTS	✓	×	$0.144_{\pm 0.004}$	0.57	0.38
10	R-space	16	DARTS	×	×	N/A	N/A	OOM
11	R-space	16	DARTS	×	✓	**$0.139_{\pm 0.005}$**	0.56	**0.24**
12	R-space	16	SGAS	✓	×	$0.140_{\pm 0.003}$	0.60	0.32
13	R-space	16	SGAS	×	×	N/A	N/A	OOM
14	R-space	16	SGAS	×	✓	**$0.136_{\pm 0.002}$**	0.52	**0.21**

a. L denotes the size of the searched network.
b. Cell indicates whether to use cell-sharing trick.
c. NPSP indicates whether to use the network proliferation search paradigm.

1. The cell-sharing trick improves the search efficiency but seriously narrows the original search space and limits the final searched GNN's capability.

2. Our network proliferation search paradigm can both improves the search effect and search efficiency.

3. The proposed search paradigm works well with different search strategy (such as DARTS and SGAS).

6. Task-based Layer

Different from traditional GNNs whose global graph representation is only constructed on the readout of node features. Our method explicitly models relational information, so it naturally constructs global graph representation with both node and relation features.

The Global Node Feature:

$$V_g = \sigma(BN([V_1 \parallel \cdots \parallel V_L]))$$

The Global Relation Feature:

$$E_g = \sigma(BN([E_1 \parallel \cdots \parallel E_L]))$$

The Global Graph Representation:

$$G_g = \left[\frac{1}{|V_g|} \sum_{i \in V_g} V_g^i \parallel \frac{1}{|E_g|} \sum_{j \in E_g} E_g^j \right]$$