The Web Conference 2022

Designing the Topology of Graph Neural Networks: A Novel Feature Fusion Perspective

Lanning Wei[1,2] weilanning18z@ict.ac.cn
Huan Zhao[2] zhaohuan@4paradigm.com
Zhiqiang He[1,3] hezq@lenovo.com

Institute of Computing Technology, Chinese Academy of Sciences[1]
4Paradigm[2] Lenovo[3]

Outline

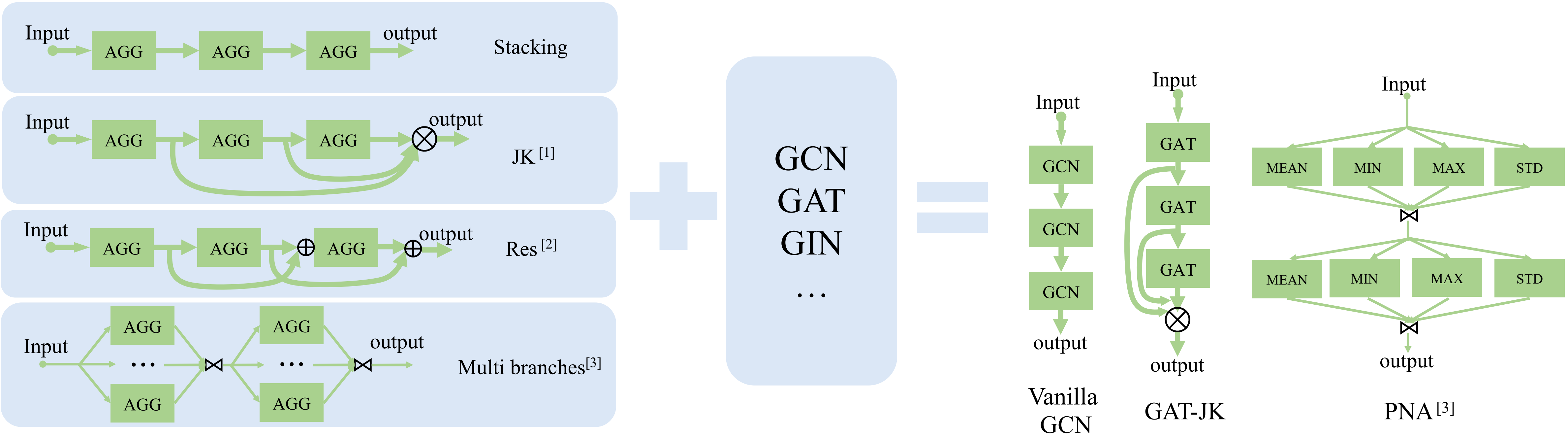
- Introduction
- Method
- Experiments
- Conclusions and Future work

Outline

- Introduction
    - GNN topology
    - Feature fusion perspective of GNN topology
- Method
- Experiments
- Conclusions and Future work

GNN Topology

Different GNNs can be built by designing the aggregation operations and the topology.

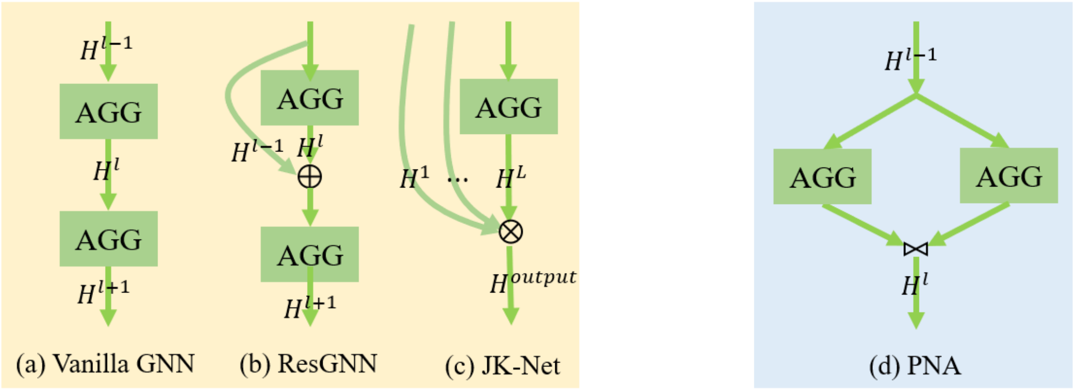


⊗ maxpool, ⊕ summation, ⋈ concatenation

[1] Representation Learning on Graphs with Jumping Knowledge Networks. ICML 2018
[2] Deepgcns: Can gcns go as deep as cnns? ICCV 2019
[3] Principal Neighbourhood Aggregation for Graph Nets. NeurIPS 2020

GNN Topology

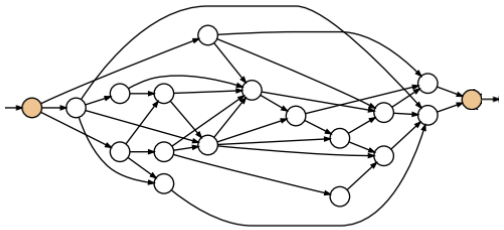
- Two mainstream GNN topology design manners
  - Stacking AGG (Left): extract higher-level feature / Over-smoothing.
  - Using multiple AGG (Right): adequate and independent feature extraction / costly to obtain the higher-level information.



- Design Target: *Can we enjoy the benefits while alleviate the corresponding deficiencies on top of these two topology design manners?*

Feature fusion perspective

- There lacks a systematic approach for the GNN topology design.
- The topology of a neural network can be represented by its "computational graph".
- In GNNs, designing the links is equivalent to selecting the features of different levels.
- Fusion strategy is also indispensable in improving the GNN.



Computational Graph: nodes represent the operations and the directed edges link operations in different layers.

Unify the GNN topology designs with feature selection and fusion strategies.

Exploring Randomly Wired Neural Networks for Image Recognition. ICCV 2019

F2GNN Contributions

- Feature fusion framework to unify the GNN topologies.
- Adaptive GNN design to achieve the design target
  - Borrow the power of neural architecture search (NAS)
  - A novel search space for GNN topology design.
  - An improved search algorithm to address the obvious optimization gap induced by the search space.
- Experiments
  - Conduct extensive experiments on eight real-world datasets, including homophily and heterophily.
  - F2GNN can improve the performance while alleviating the deficiencies, especially alleviating the over-smoothing problem.

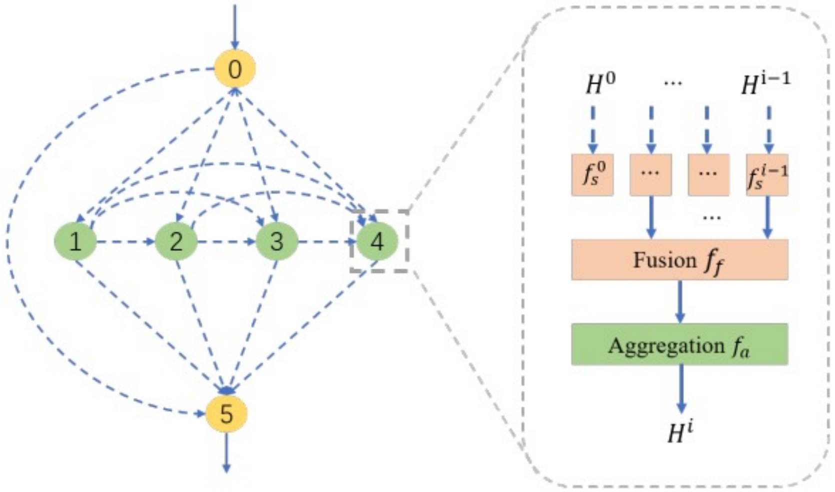
Outline

- Introduction
- Method
  - Feature fusion framework
  - Design adaptive GNNs with this framework
    - The design of the search space
    - The improved differentiable search algorithm
- Experiments
- Conclusions and Future work

Feature fusion framework

- The feature selection and fusion strategies lead to the key difference of topology designs in GNNs.

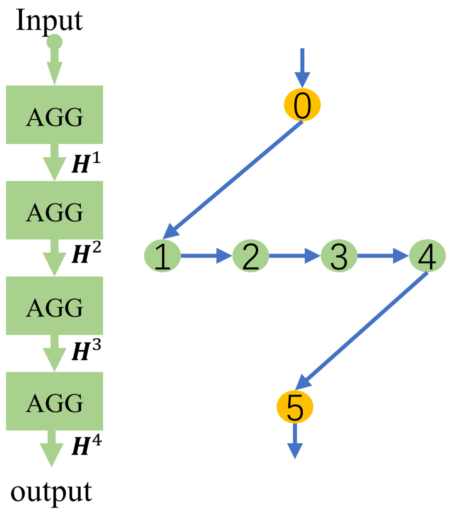
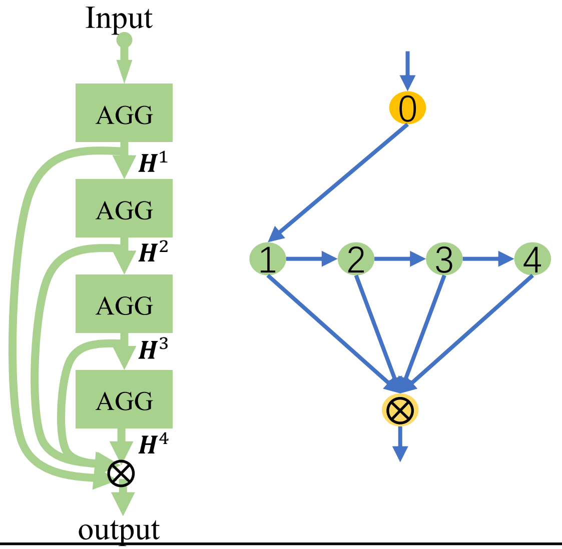
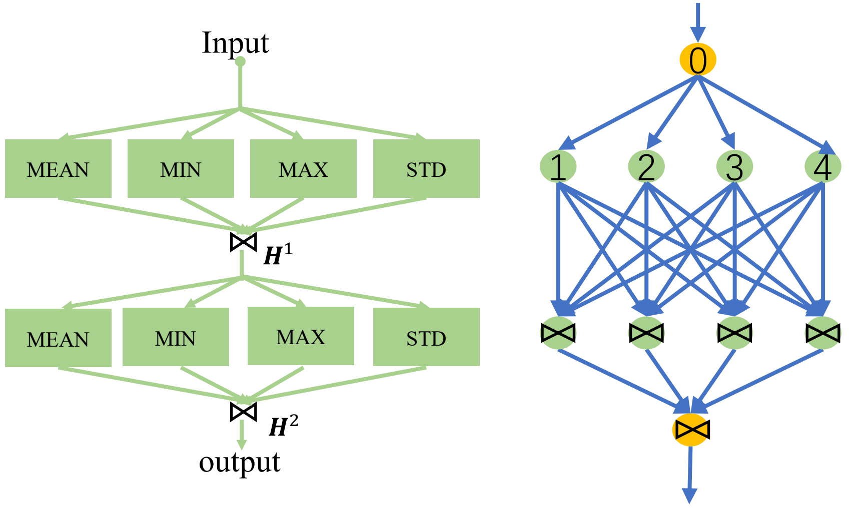
- Feature fusion framework
  - 0 Pre-processing: 2-layer MLPs
  - SFA Block: Selection + Fusion + Aggregation
  - 5 Post-processing: Selection + Fusion + 2-layer MLPs



(a) The unified framework (b) SFA Block i

Feature fusion framework

Translating the framework into diverse GNNs.

	Vanilla GNNs	JK-Net (maxpool)	2-layer PNA
Topology formulation	$\boldsymbol{H}^{l+1} = f(\boldsymbol{H}^{l})$	$\boldsymbol{H}^{output} = max(\boldsymbol{H}^{1}, \cdots \boldsymbol{H}^{L})$	$\boldsymbol{H}^{l+1} = \Vert_{i \in M} f_i(\boldsymbol{H}^{l})$
GNNs and their topology illustrations			

Design adaptive GNNs — Search space

- Design GNNs = Topology + Aggregation operations
- Topology: Selection $\mathcal{O}_s$ + Fusion $\mathcal{O}_f$
- Aggregation operations
  - Predefined aggregation operations: GraphSAGE (F2SAGE) / GAT(F2GAT)
  - Learnable aggregation operations $\mathcal{O}_a$ (F2GNN)

Table 1: The operations used in our search space.

	Operations
Selection $\mathcal{O}_s$	ZERO, IDENTITY
Fusion $\mathcal{O}_f$	SUM, MEAN, MAX, CONCAT, LSTM, ATT
Aggregation $\mathcal{O}_a$	GCN, GAT, SAGE, GIN

Design adaptive GNNs — Search Algorithm

- Differentiable NAS method is adopted considering the efficiency.
- The optimization gap in feature fusion
  - Mixed selection results

$$\bar{o}^{ij}(\mathbf{x}_i) = \sum_{k=1}^{|O_s|} c_k^{ij} o_k^{ij}(\mathbf{x}_i) = c_1^{ij}\mathbf{0} + c_2^{ij}\mathbf{x}_i = c_2^{ij}\mathbf{x}_i.$$

  - ChildNet results:
    - ZERO: $c_1^{ij}\mathbf{0}$
    - IDENTITY: $c_2^{ij}\mathbf{x}_i$

The IDENTITY operation has a large influence when ZERO is selected, and the influence will accumulate along with the feature selection operation in the framework.

- Improved search with the usage of temperature

Add a small temperature λ in the softmax.

$$c_k = \frac{\exp(\alpha_k/\lambda)}{\sum_{i=1}^{|O|}\exp(\alpha_i/\lambda)}$$

Outline

- Introduction
- Feature fusion Graph Neural Networks
- Experiments
    - Experimental settings
    - Performance comparisons
    - Advantages of the adaptive topology design
    - Ablation study
- Conclusions and Future work

Experimental settings

- Datasets
  - Five homophily + three Heterophily
- Baselines
  - Stacking baselines
  - Various identity skip-connections
  - Multiple aggregations
  - Graph NAS baselines

Datasets	#Nodes	#Edges	#Features	#Classes	h
Cora [37]	2,708	5,278	1,433	7	0.81
Computers [32]	13,381	245,778	767	10	0.78
DBLP [2]	17,716	105,734	1,639	4	0.83
PubMed [37]	19,717	44,324	500	3	0.80
Physics [38]	34,493	495,924	8,415	5	0.93
Wisconsin [34]	251	466	1,703	5	0.21
Actor [34]	7,600	30,019	932	5	0.22
Flickr [52]	89,250	899,756	500	7	0.32

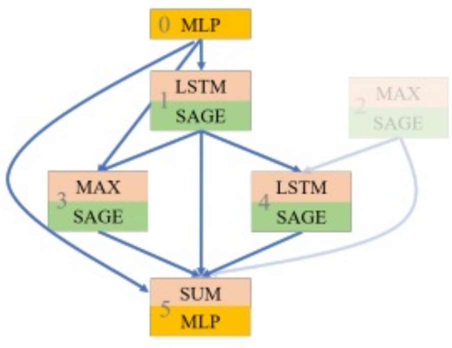
$$h = \frac{|\{(u,v):(u,v)\in\mathcal{E}\wedge y_u = y_v\}|}{|\mathcal{E}|}$$

Performance comparisons

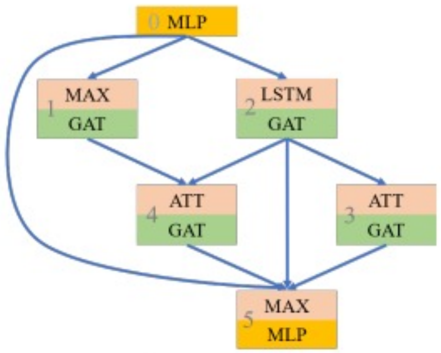
- Adaptive GNN topologies (ours+Random) have better performance than the human-designed topologies.
- Our methods rank 1^{st} compared with all the human-designed topologies and the Random Baselines.

Aggregation	Topology	Cora	DBLP	PubMed	Computers	Physics	Actor	Wisconsin	Flickr	Avg. Rank (Group)	Avg. Rank (All)
SAGE	Stacking (L2)	86.09(0.50)	83.58(0.33)	88.96(0.29)	91.14(0.30)	96.42(0.11)	34.78(1.10)	79.61(5.56)	51.21(0.71)	6.63	15.00
	Stacking (L4)	85.68(0.61)	83.83(0.32)	88.23(0.28)	90.52(0.42)	95.97(0.14)	34.61(1.08)	60.39(10.77)	53.07(0.50)	8.25	17.00
	RES (L4)	85.66(0.52)	83.39(0.30)	88.99(0.25)	91.51(0.18)	96.31(0.17)	35.16(0.94)	76.47(5.26)	53.72(0.27)	5.25	13.13
	DENSE (L4)	86.68(0.59)	83.30(0.73)	89.42(0.27)	90.74(0.51)	96.48(0.14)	34.78(0.60)	77.06(6.01)	53.17(0.19)	4.50	12.75
	JK (L4)	86.47(0.60)	83.94(0.62)	89.21(0.29)	91.21(0.30)	96.56(0.05)	36.53(0.92)	81.96(4.71)	52.41(0.33)	4.75	10.38
	GNNII (L4)	85.83(0.42)	84.46(0.45)	89.21(0.24)	91.38(0.27)	96.45(0.15)	35.70(1.11)	81.57(4.13)	52.24(0.29)	4.50	11.50
	PNA (L2)	84.29(0.67)	82.76(0.42)	89.25(0.26)	90.67(0.42)	96.32(0.10)	33.89(2.68)	75.29(6.46)	52.09(0.73)	8.88	17.75
	MixHop (L2)	84.81(0.95)	82.65(0.65)	89.25(0.28)	88.56(1.61)	96.11(0.17)	35.19(0.62)	81.57(2.51)	51.75(0.59)	6.75	17.75
	Random	86.75(0.29)	83.60(0.29)	89.21(0.04)	91.30(0.19)	96.46(0.03)	36.30(0.58)	85.10(5.63)	54.10(0.15)	3.50	8.75
	F^2SAGE	87.72(0.26)	84.81(0.06)	89.73(0.26)	91.81(0.26)	96.72(0.01)	36.61(1.00)	85.88(1.92)	53.66(0.16)	2.00	4.38
GAT	Stacking (L2)	85.92(0.72)	84.34(0.26)	87.56(0.23)	91.49(0.21)	95.76(0.16)	29.28(1.02)	53.73(7.24)	53.83(0.28)	5.25	14.25
	Stacking (L4)	86.16(0.55)	84.29(0.41)	85.73(0.34)	89.08(0.43)	93.47(3.93)	26.45(1.00)	45.29(5.65)	50.34(2.68)	8.25	19.88
	RES (L4)	84.66(0.92)	84.11(0.34)	87.56(0.44)	90.84(0.49)	95.67(0.28)	28.98(0.36)	48.82(3.77)	53.63(0.24)	7.50	18.50
	DENSE (L4)	85.31(0.86)	83.43(0.37)	88.67(0.19)	91.30(0.37)	96.16(0.06)	31.78(1.03)	53.33(7.73)	53.61(0.26)	6.25	16.38
	JK (L4)	86.55(0.46)	83.73(0.35)	89.71(0.16)	91.80(0.23)	96.80(0.09)	35.43(0.88)	84.51(5.58)	53.02(0.29)	3.88	8.75
	GNNII (L4)	85.40(1.06)	83.83(0.33)	88.44(0.25)	91.91(0.11)	96.14(0.15)	30.29(0.78)	55.29(6.25)	53.03(0.29)	5.38	15.00
	PNA (L2)	85.06(0.72)	83.46(0.47)	87.18(0.30)	90.84(0.24)	95.85(0.18)	28.56(0.82)	49.22(5.91)	54.02(0.33)	7.38	18.25
	MixHop (L2)	85.38(1.04)	82.50(0.34)	88.91(0.19)	91.27(0.37)	96.46(0.21)	35.70(0.90)	81.57(4.40)	53.67(0.30)	5.13	13.25
	Random	85.73(0.06)	83.60(0.19)	88.86(0.18)	91.76(0.14)	96.84(0.09)	36.07(0.83)	86.08(4.15)	52.43(0.29)	4.38	10.38
	F^2GAT	88.31(0.12)	84.76(0.04)	90.38(0.14)	92.04(0.17)	97.10(0.03)	36.65(1.13)	87.06(4.13)	53.45(0.19)	1.63	3.13
Learnable	SNAG (L4)	84.99(1.04)	84.29(0.15)	87.93(0.16)	85.98(0.72)	96.18(0.11)	28.13(0.74)	43.92(4.65)	53.50(0.31)	4.00	18.63
	SANE (L4)	86.40(0.38)	84.58(0.13)	89.34(0.31)	91.02(0.21)	96.80(0.06)	36.77(1.15)	86.47(3.09)	53.92(0.14)	2.63	6.38
	Random	86.99(0.60)	84.62(0.15)	89.37(0.26)	91.03(0.20)	96.72(0.04)	36.29(1.52)	85.49(4.31)	54.33(0.11)	2.25	6.13
	F^2GNN	87.42(0.42)	84.95(0.15)	89.79(0.20)	91.42(0.26)	96.92(0.06)	37.08(1.00)	88.24(3.72)	53.96(0.20)	1.13	2.75

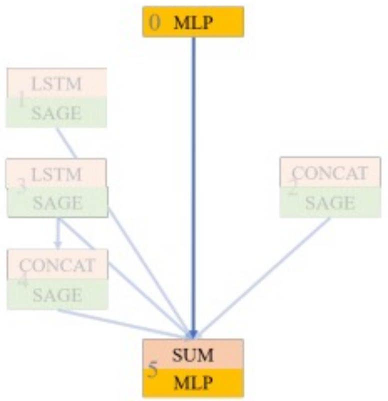
Searched Architectures



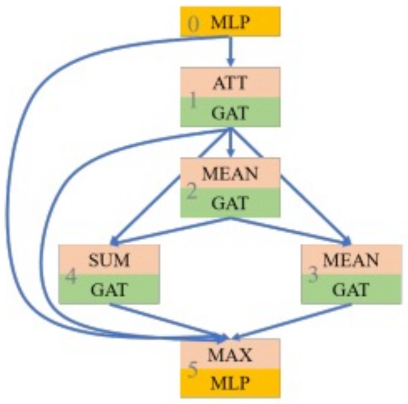
(a) F²SAGE on Cora



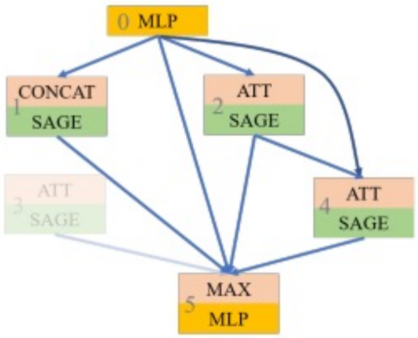
(b) F²GAT on Cora



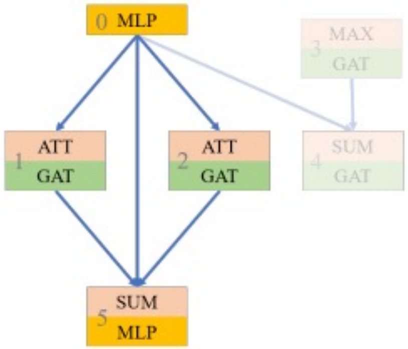
(e) F²SAGE on Actor



(f) F²GAT on Actor



(c) F²SAGE on Physics



(d) F²GAT on Physics

- Data-specific GNN topologies are obtained.
- The initial feature is selected in most GNNs.
- We can benefit from the multiple aggregation design manner.
- On the heterophily dataset Actor, we obtained an MLP network, which shows that the graph structure is not always useful for the final performance.

NAS provides a unifying solution of GNN design on both homophily and heterophily graphs.

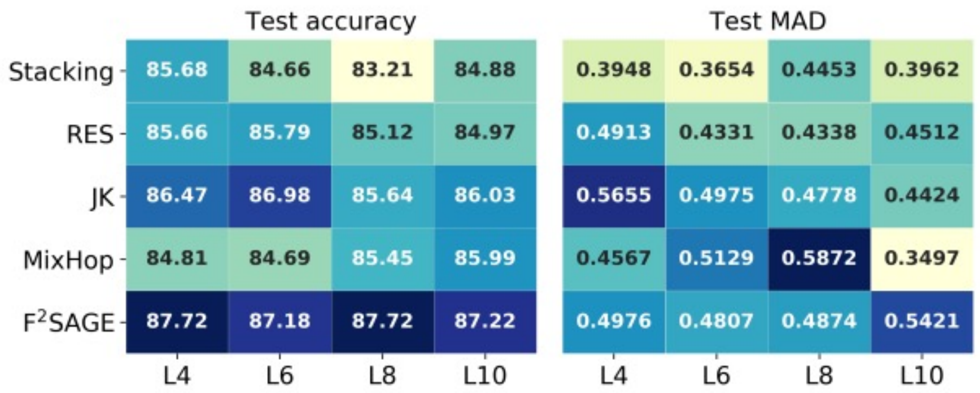
Advantages of the adaptive topology design

Alleviating the over-smoothing problem

- Over-smoothing: In deep GNNs, the node representations become indistinguishable and easily get performance drop.

- Using different levels of features can alleviate this problem.

- F2SAGE achieves the SOTA performance and higher MAD values by utilizing features in each block adaptively.

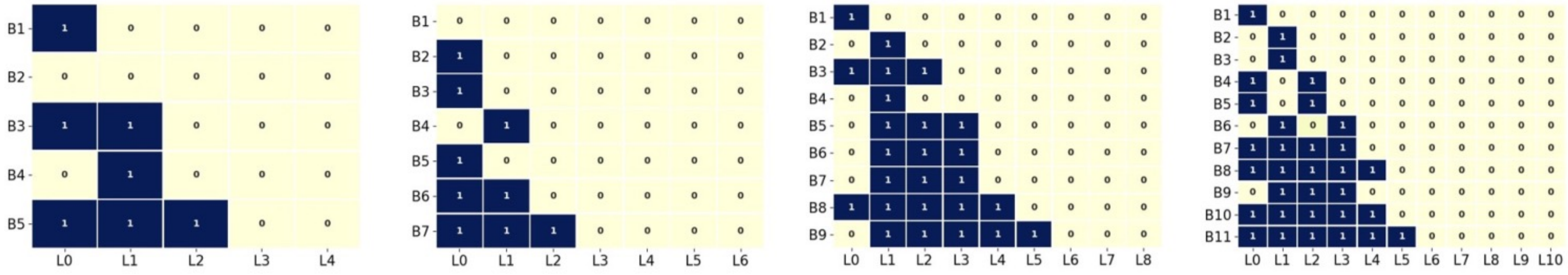


Advantages of the adaptive topology design

Flexibility in obtaining the higher-level features.

- With 4(8) aggregations, PNA obtains the features in level 1(2), while F2SAGE obtains the features in level 2 (5).

- Our method achieves higher performance than PNA with 35% and and 15% fewer parameters on the GraphSAGE and GAT, respectively.



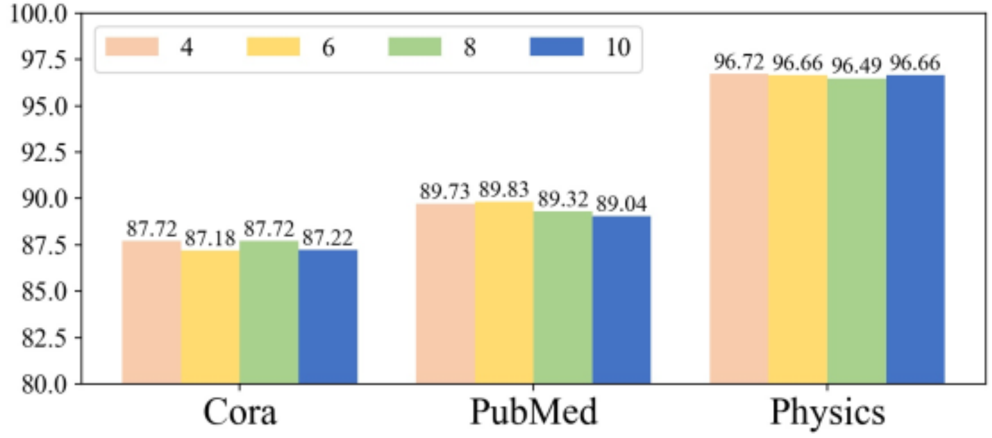
(a) F^2SAGE (N=4) on Cora (b) F^2SAGE (N=6) on Cora (c) F^2SAGE (N=8) on Cora (d) F^2SAGE (N=10) on Cora

Ablation study

- Designing the fusion strategy is significant.
- The optimization gap has a large influence on the feature selection, and it can be addressed with λ .
- The increasing number of SFA blocks do not bring about the performance drop due to the adaptive utilization of different levels of features .

Method	Cora	PubMed	Physics
F^2SAGE-SUM	84.73(0.63)	89.39(0.21)	96.44(0.01)
F^2SAGE-MEAN	84.30(0.61)	89.58(0.22)	96.42(0.03)
F^2SAGE-CONCAT	86.07(0.45)	89.31(0.19)	96.69(0.01)
F^2SAGE	87.72(0.26)	89.73(0.26)	96.72(0.01)

Temperature λ	F^2SAGE		F^2AGG	
	Supernet	Childnet	Supernet	Childnet
1	80.33	6.68	86.83	85.71
0.1	73.65	10.96	84.23	83.86
0.01	70.13	70.13	84.60	84.60
0.001	80.15	80.15	86.83	86.83



Revisiting AutoGraph challenge at KDD Cup 2020

- 15 node classification datasets in diverse domains.
- F2GCN reaches 97.3% performance with 45.1% parameter size compared with the best solution.

Dataset	Phase	Domain	#Node	#Edge	#Feature	#Class
a	Public	Citation	2.7K	5.3K	1.4K	7
b	Public	Citation	3.3K	4.6K	3.7K	6
c	Public	Social	10K	733K	0.6K	41
d	Public	News	10K	2,917K	0.3K	20
e	Public	Finance	7.5K	7.8K	0	3
f	Feedback	Sales	10K	194K	0.7K	10
g	Feedback	Citation	10K	41K	8K	5
h	Feedback	Medicine	10K	2,461K	0.3K	23
i	Feedback	Finance	15K	16K	0	3
j	Feedback	Medicine	11K	22K	0	9
k	Private	Sales	8K	119K	0.7K	8
l	Private	Citation	10K	40K	7K	15
m	Private	News	10K	1,425K	0.3K	8
n	Private	Finance	14K	22K	0	10
o	Private	Social	12K	19K	0	19

Dataset	GCN(L2)	GCN(L4)	F^2GCN(L4)	1st solution
a	85.7	84.4	84.4 (95.4)	88.5 (100)
b	71.4	70.5	71.3 (94.8)	75.2 (100)
c	86.5	82.3	92.8 (98.4)	94.3 (100)
d	93.7	93.6	93.9 (97.3)	96.5 (100)
e	59.6	87.5	88.4 (99.7)	88.7 (100)
f	86.6	87.6	92.1 (99.2)	92.8 (100)
g	94.7	93.4	95.3 (100)	95.3 (100)
h	90.4	90.3	90.1 (96.4)	93.5 (100)
i	88.2	87.6	88.3 (99.9)	88.4 (100)
j	90.7	83.6	95.3 (99.4)	95.9 (100)
k	93.5	93.2	93.4 (97.9)	95.5 (100)
l	90.9	89.1	92.9 (97.9)	94.9 (100)
m	85.5	86.1	86.1 (87.8)	98.1 (100)
n	85.6	95.2	96.7 (97.7)	99.0 (100)
o	49.6	71.8	88.8 (97.6)	91.0 (100)
Avg			- (97.3)	- (100)

Table 4: Accuracy comparison of GCN baselines, F^2GCN and industrial best solution (%). L2, L4 means 2 and 4 layers for the GNN architecture. Numbers in parentheses are relative accuracy w.r.t 1st solution. We regard 1st solution as 100%. Last line is the average percentage.

Dataset	GCN(L2)	F^2GCN(L4)	1st solution
a	0.023	0.908 (75.7)	1.199 (100)
b	0.059	0.700 (44.2)	1.583 (100)
c	0.011	1.598 (98.0)	1.631 (100)
d	0.006	0.042 (3.20)	1.296 (100)
e	0.121	0.354 (31.8)	1.114 (100)
f	0.013	0.039 (2.30)	1.688 (100)
g	0.134	0.313 (13.1)	2.389 (100)
h	0.006	0.271 (20.9)	1.294 (100)
i	0.241	2.269 (113.0)	2.013 (100)
j	0.171	0.834 (60.6)	1.376 (100)
k	0.012	1.478 (108.0)	1.395 (100)
l	0.108	0.614 (25.6)	2.395 (100)
m	0.005	0.010 (0.80)	1.278 (100)
n	0.218	0.488 (27.8)	1.756 (100)
o	0.192	0.822 (52.5)	1.565 (100)
Avg		- (45.1)	- (100)

Table 5: Number of parameters of baseline, 1st solution and F^2GCN (Unit: Millions). Numbers in parentheses are relative # parameters w.r.t 1st solution. We regard 1st solution as 100%. Last line is the average percentage.

(a) Dataset statistics (b) Performance comparisons. (c) Parameter size comparisons.

Bridging the Gap of AutoGraph between Academia and Industry: Analysing AutoGraph Challenge at KDD Cup 2020. arxiv 2022

Outline

- Introduction
- Method
- Experiments
- Conclusions and Future work

Conclusion and Future work

- We provide a novel feature fusion perspective in designing the GNN topology.
- A novel framework is designed to unify the existing topology designs with feature selection and fusion strategies, and a NAS method is developed to obtained the adaptive topology design.
- The experimental results demonstrate the effectiveness and versatility of the proposed F2GNN.
- Future work: we will investigate the influence of different candidate operations and algorithms, and explore F2GNN in the OGB datasets.

Code

- Code: https://github.com/AutoML-Research/F2GNN
- More related methods: https://github.com/AutoML-Research
  - Search to aggregate neighborhood for graph neural network (ICDE 2021)
  - Pooling Architecture Search for Graph Classification. (CIKM 2021)
  - Bridging the Gap of AutoGraph between Academia and Industry: Analysing AutoGraph Challenge at KDD Cup 2020.

Code link | Paper link | AutoML Research Group

Thank you !

Q&A

Contact: Lanning Wei, weilanning18z@ict.ac.cn