

# PaSca:可扩展的图神经 结构搜索系统

---

张文涛

北京大学博士，腾讯Angel Graph团队成员

个人主页: <https://zwt233.github.io/>



# 目录 CONTENT

01 问题

03 方法

02 实验

04 总结

# 01

## 问题

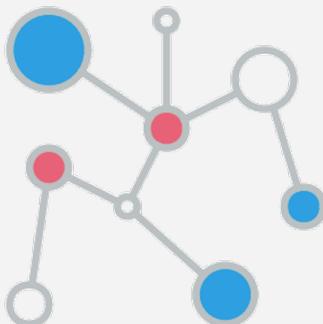


# 图数据

许多数据都是以图的形式存在：



社交网络



知识图谱



药物和新材料

图神经网络被广泛应用于多个场景：

- 推荐系统
- 药物发现
- 异常检测
- 蛋白质结构预测

# 图神经网络

图卷积神经网络(GCN)的表达形式：

第  $l+1$  层的节点表示

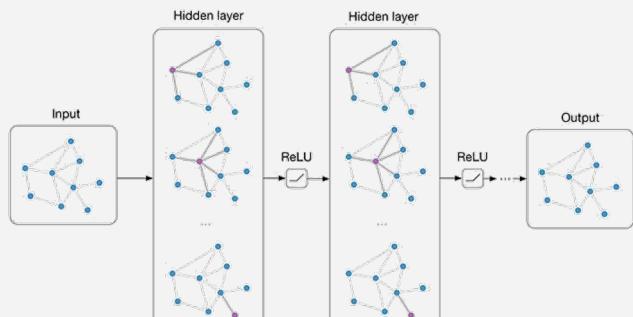
度矩阵

第  $l$  层的节点表示

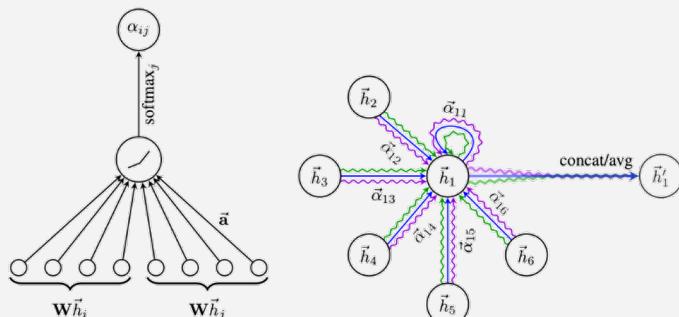
$$\mathbf{X}^{(l+1)} = \delta \left( \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}^{(l)} \mathbf{W}^{(l)} \right)$$

含自环的邻接矩阵

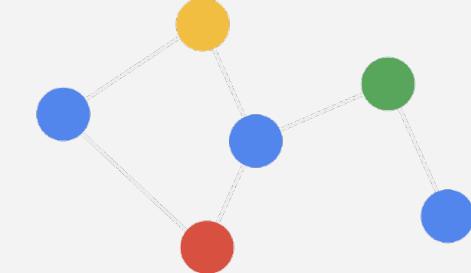
第  $l$  层的模型参数



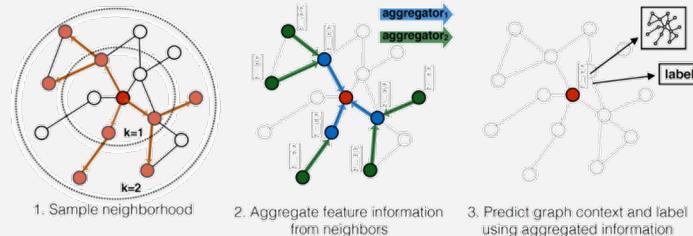
GCN



GAT



- 通过消息传播机制聚合高阶邻居的信息
- 提升自身的表达能力



GraphSAGE

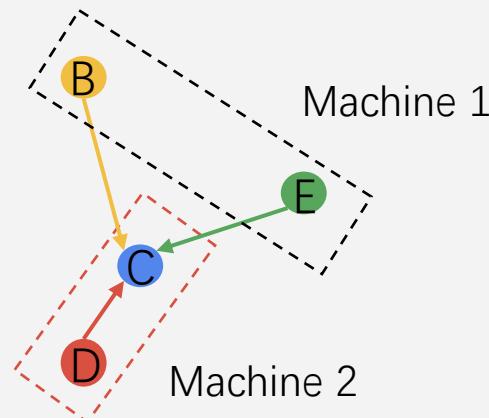
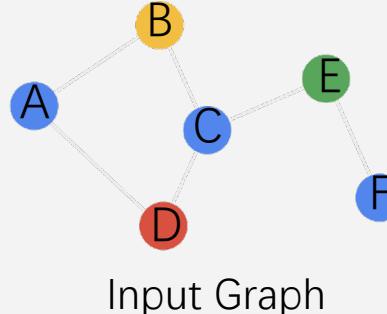
[1] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. ICLR, 2017.

[2] Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks. ICLR. 2018.

[3] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. NeurIPS, 2017.

# Neural Message Passing (消息传递机制)

- 传统的GNN (如GCN[1], GAT[2]) 都遵循 neural message passing (NMP , 消息传递机制) paradigm:
  - Aggregate the neighborhood information (**通信**)
$$\mathbf{m}_v^t \leftarrow \text{aggregate} \left( \{\mathbf{h}_u^{t-1} | u \in \mathcal{N}_v\} \right)$$
  - Update the message via neural networks (**计算**)
$$\mathbf{h}_v^t \leftarrow \text{update}(\mathbf{m}_v^t)$$
- 缺点: **频繁地** 从其他机器上拉取信息 → 大规模图数据上每个epoch都有的**高通信开销**



GIF from [https://blog.csdn.net/DreamHome\\_S/article/details/105619194](https://blog.csdn.net/DreamHome_S/article/details/105619194)

[1] Thomas N Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In ICLR.

[2] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In ICLR.

# GNN 系统

大多数GNN系统使用消息传播机制

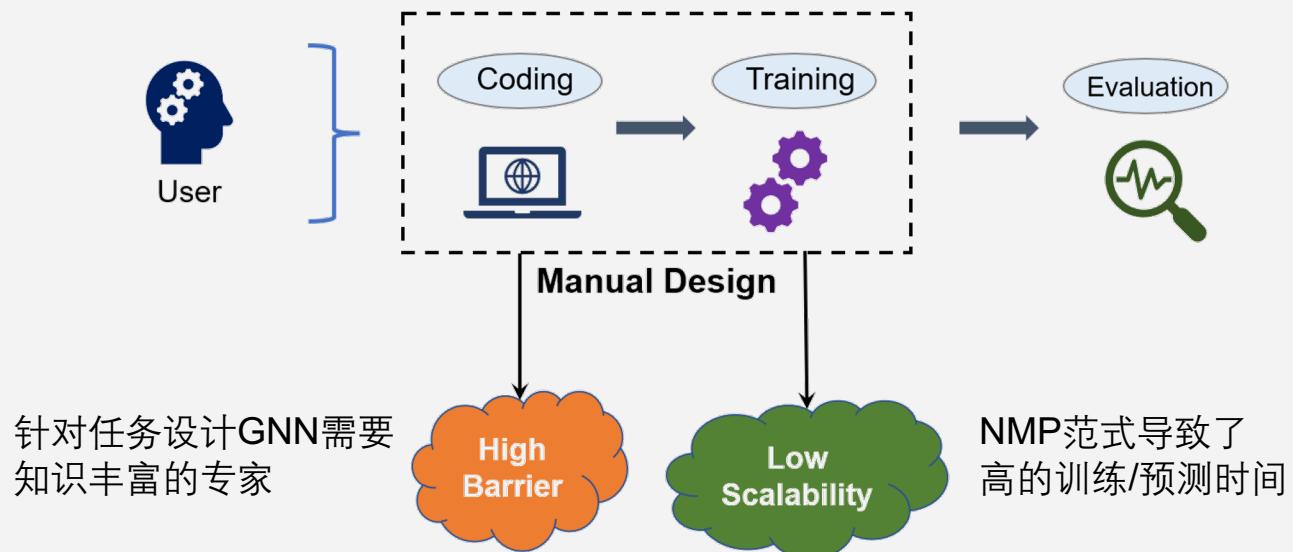


DGL[1]



PyG[2]

## 大规模图数据带来的挑战



[1] <https://github.com/dmlc/dgl>

[2] [https://github.com/pyg-team/pytorch\\_geometric](https://github.com/pyg-team/pytorch_geometric)

# 瓶颈

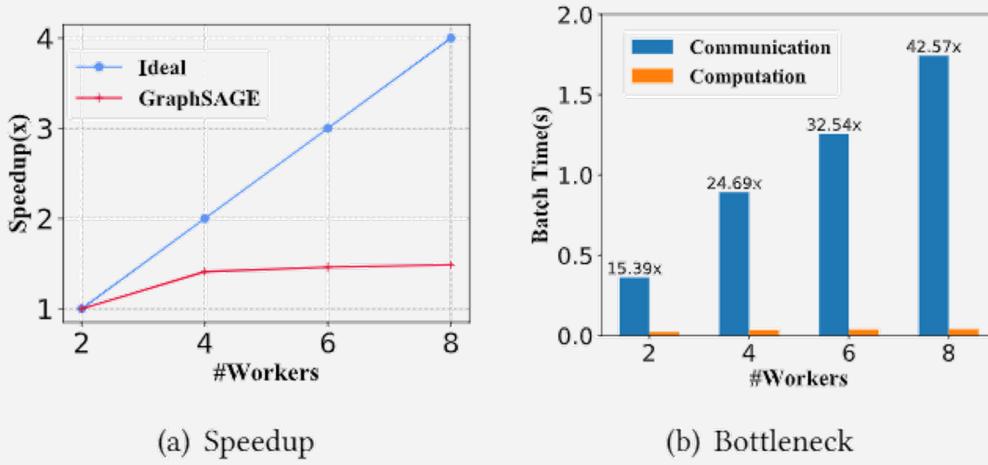


Figure 2: The speedup and bottleneck of a two-layer GraphSAGE along with the increased workers on Reddit dataset.

**可扩展性**：受制于单机的存储开销和分布式通讯开销，现有的消息传递机制不能很好地扩展到大图上。

- 增加更多机器时候，加速比增长不明显
- 通信开销占比过大

**目标**：如何兼顾GNN的**可扩展性**，设计**使用门槛低**的图神经网络系统？

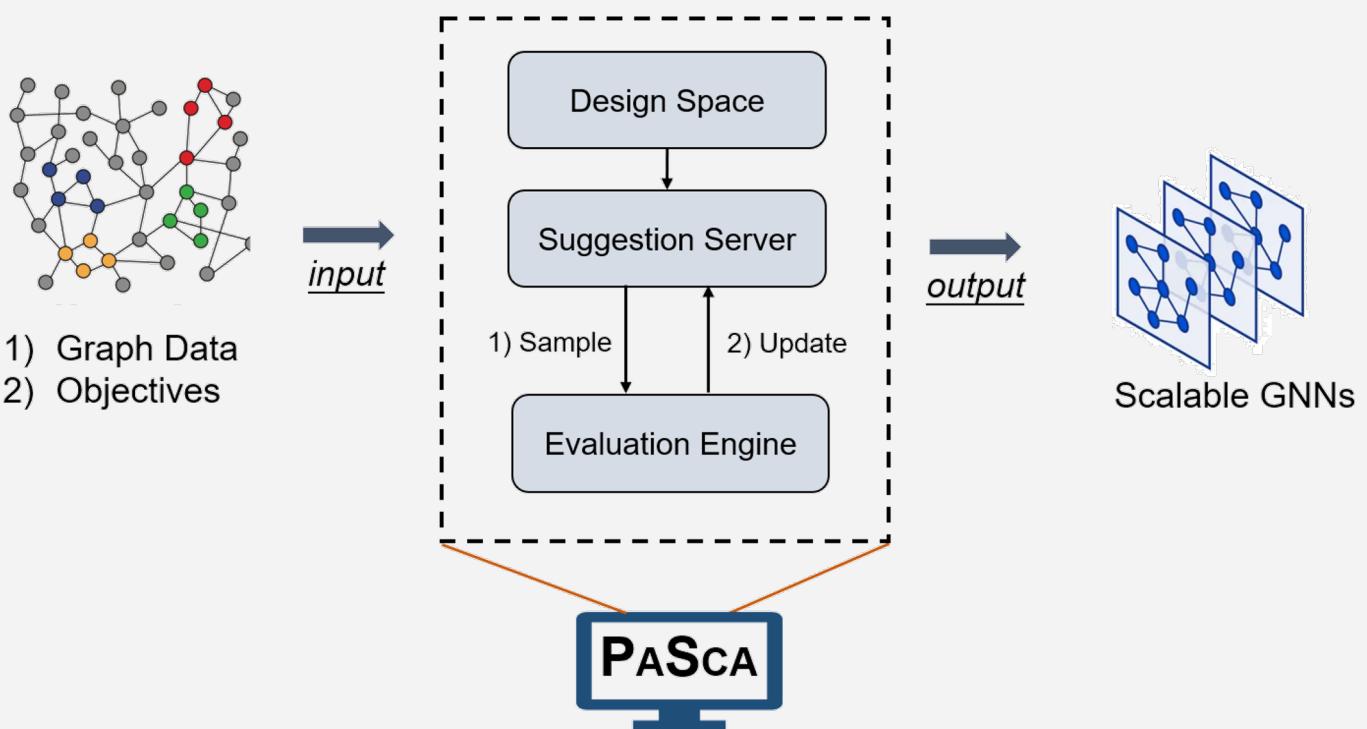
# 02

# 方法



# 系统目标

- 输入: 图数据 + 优化目标
- 输出: 能兼顾多个优化目标的**Scalable GNN**



端到端系统，无需人为定义网络结构和训练流程

# 消息传递 (Message Passing) 范式

$$\mathbf{x}_i^{(k)} = \gamma^{(k)} \left( \mathbf{x}_i^{(k-1)}, \bigcup_{j \in \mathcal{N}(i)} \phi^{(k)} \left( \mathbf{x}_i^{(k-1)}, \mathbf{x}_j^{(k-1)}, \mathbf{e}_{j,i} \right) \right)$$

Aggregate Function

Update Function      Message Function

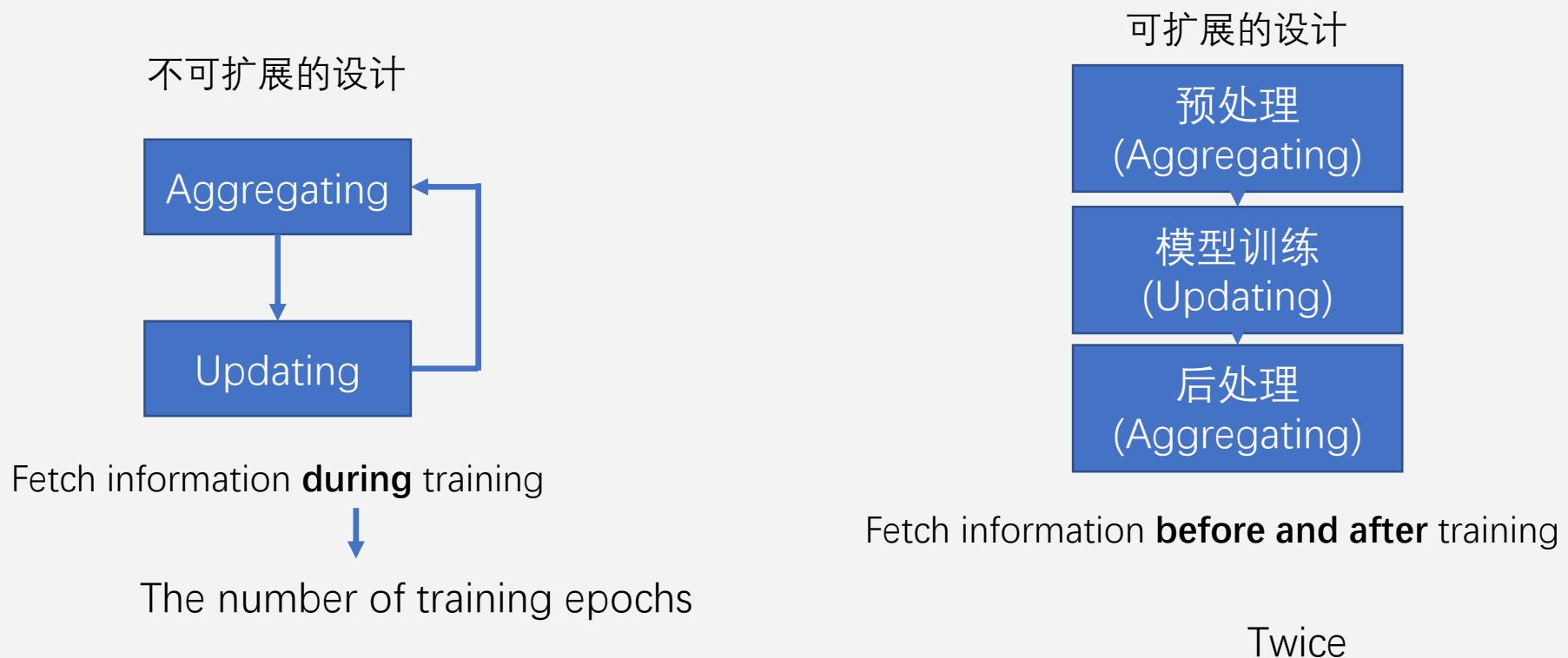
不断迭代的“聚合-更新”流程。

消息传递范式从**节点**层次来刻画数据的流动，主要由三个操作构成：

- 1. Message Function**：定义了从**生成信息**的方式；
- 2. Aggregate Function**：定义了**聚合信息**的方式；
- 3. Update Function**：定义了**更新中心节点特征**的方式。

# 方法概览

- Scalable Graph Neural Architecture Paradigm (SGAP建模范式)
  - 定义可扩展训练流程的抽象
- 自动搜索系统 (PaSca)



$$\mathbf{M} = \text{graph\_propagate}(\mathbf{A}, \mathbf{X}) \quad (1)$$

$$\mathbf{X}' = \text{message\_aggregate}(\mathbf{M}) \quad (2)$$

$$\mathbf{Y} = \text{model\_train}(\mathbf{X}') \quad (3)$$

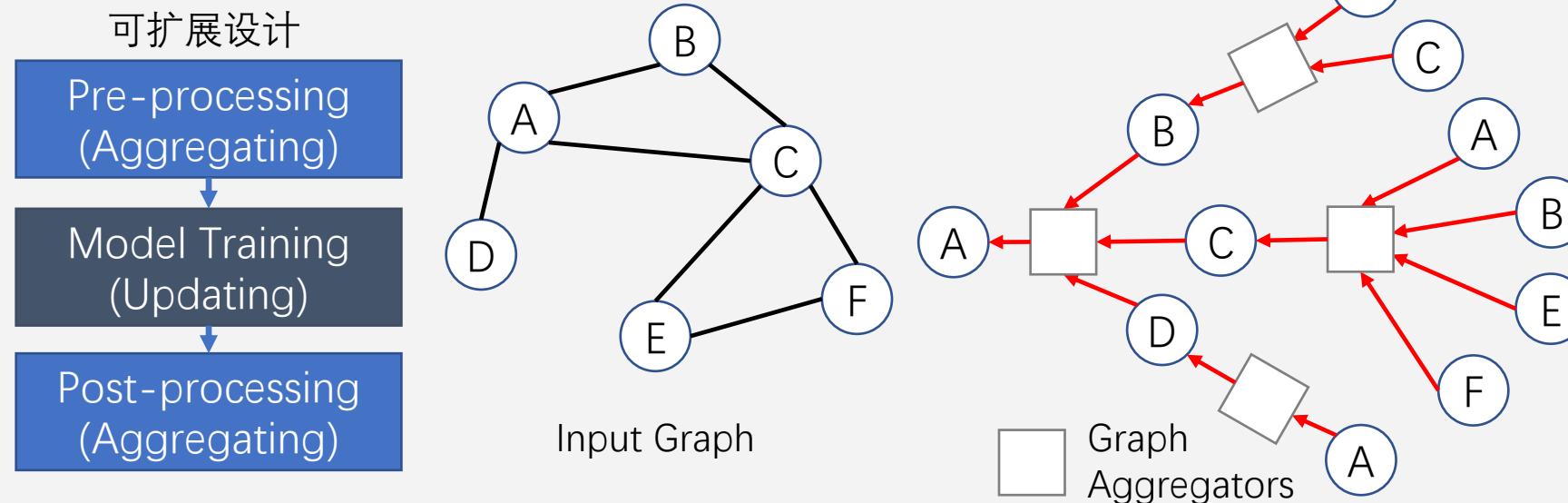
SGAP从图的层次刻画数据的流动，主要由三个操作构成：

- (1) : 在图的层级传播信息，得到不同传播层数的消息， $\mathbf{M}$ ；
- (2) : 聚合不同传播层数的消息，得到新的特征， $\mathbf{X}'$ ；
- (3) : 将 $\mathbf{X}'$ 送入一个机器学习模型（如MLP）进行训练，得到最终输出， $\mathbf{Y}$ 。

# SGAP抽象

- 预处理
  - 从邻居节点聚合消息（**特征**）
- 后处理
  - 从邻居节点聚合消息（**软标签**）

$$\mathbf{m}_v^t \leftarrow \text{graph\_aggregator} \left( \{\mathbf{m}_u^{t-1} | u \in \mathcal{N}_v\} \right)$$



# Graph Aggregator (图聚合器)

- 抽象  $\mathbf{m}_v^t \leftarrow \text{graph\_aggregator} \left( \{\mathbf{m}_u^{t-1} | u \in \mathcal{N}_v\} \right)$
- Augmented normalized adjacency (used in GCN[1])

$$\mathbf{m}_v^t = \sum_{u \in \mathcal{N}_v} \frac{1}{\tilde{d}_u} \mathbf{m}_u^{t-1}$$

- Personalized PageRank (used in APPNP[2])

$$\mathbf{m}_v^t = \alpha \mathbf{m}_v^0 + (1 - \alpha) \sum_{u \in \mathcal{N}_v} \frac{1}{\sqrt{\tilde{d}_v \tilde{d}_u}} \mathbf{m}_u^{t-1}$$

- Triangle-induced adjacency (used MotifNet[3])

$$\mathbf{m}_v^t = \sum_{u \in \mathcal{N}_v} \frac{1}{d_v^{tri}} \mathbf{m}_u^{t-1}$$

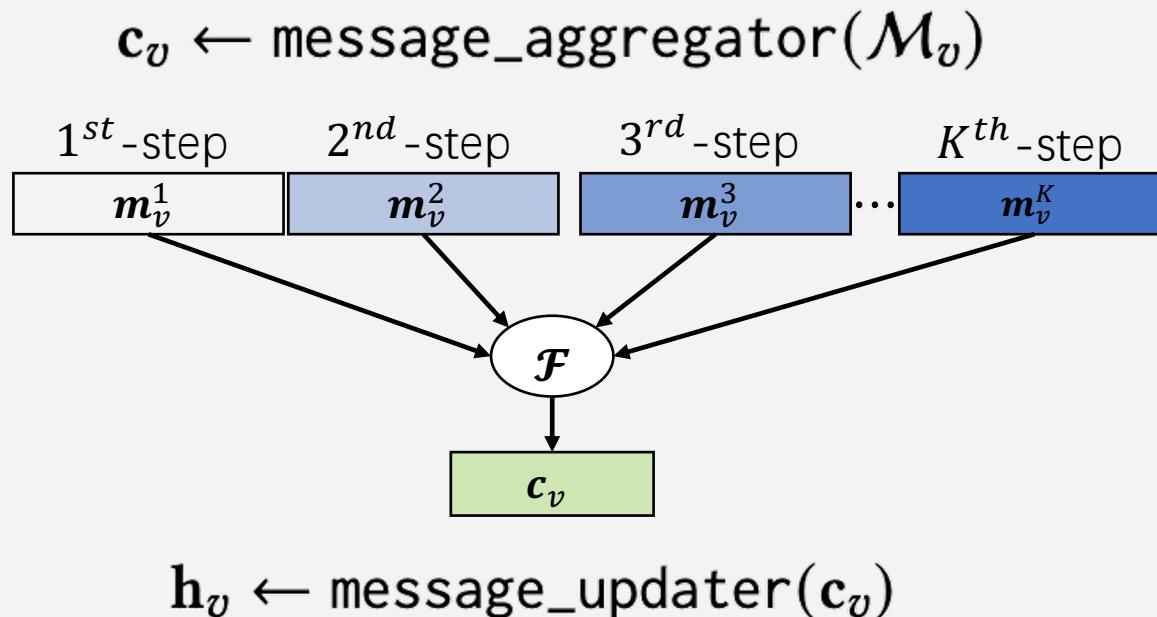
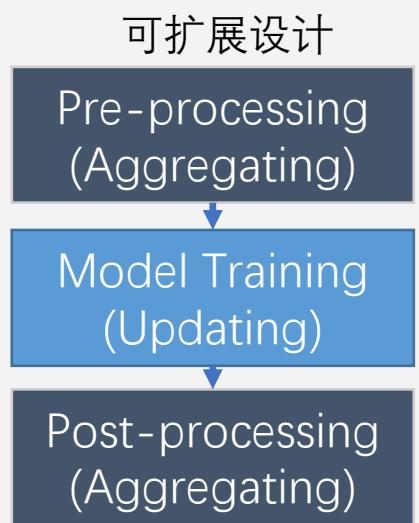
[1] Thomas N Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In ICLR.

[2] Johannes Klicpera, Aleksandar Bojchevski, and Stephan Günnemann. 2019. Predict then Propagate: Graph Neural Networks meet Personalized PageRank. In ICLR.

[3] Federico Monti, Karl Otness, and Michael M Bronstein. 2018. Motifnet: a motif-based graph convolutional network for directed graphs. In 2018 IEEE Data Science Workshop (DSW). IEEE, 225–228.

- 训练

- 聚合来自预处理阶段的消息
- 更新聚合后的消息



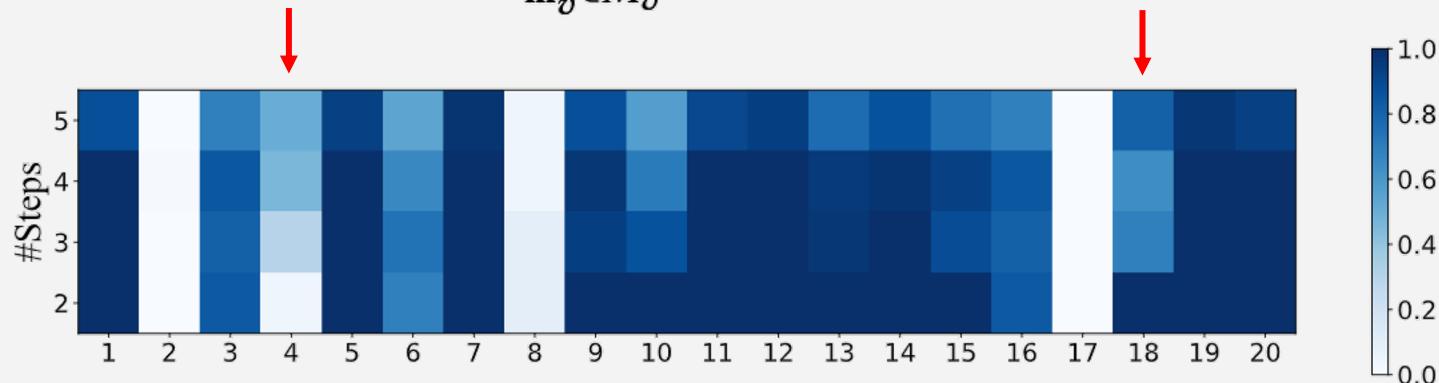
# Message Aggregator (消息聚合器)

- 抽象  $c_v \leftarrow \text{message\_aggregator}(\mathcal{M}_v)$
- 非自适应聚合器 (mean, max)

$$c_{msg} \leftarrow \oplus_{\mathbf{m}_v^i \in M_v} w_i f(\mathbf{m}_v^i)$$

- 自适应聚合器 (gate with trainable parameters)

$$c_{msg} \leftarrow \sum_{\mathbf{m}_v^i \in M_v} w_i \mathbf{m}_v^i, w_i = \sigma(\mathbf{s}\mathbf{m}_v^i)$$



应该给不同节点的不同层表示消息不同的权重!

# SGAP范式

**Algorithm 1: An example of scalable graph neural architectures following SGAP.**

**Input:** Graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , aggregation depth  $K_{pre}, K_{post}$ , features  $\mathbf{x}_v$  for each node  $v \in \mathcal{V}$ .

**Output:** Prediction message  $\mathbf{m}_v^{K_{post}}$  for each node  $v \in \mathcal{V}$ .

1 Initialize message set  $\mathcal{M}_v = \{\mathbf{x}_v\}$  for each node  $v \in \mathcal{V}$ ;

2 **Stage 1: Pre-processing**

3 Initialize feature message  $\mathbf{m}_v^0 = \mathbf{x}_v$  for each node  $v \in \mathcal{V}$ ;

4 **for**  $1 \leq t \leq K_{pre}$  **do**

5   **for**  $v \in \mathcal{V}$  **do**

6      $\mathbf{m}_v^t \leftarrow \text{graph\_aggregator}(\mathbf{m}_{N_v}^{t-1})$ ;

7      $\mathcal{M}_v = \mathcal{M}_v \cup \{\mathbf{m}_v^t\}$ ;

8      $\mathbf{c}_v \leftarrow \text{message\_aggregator}(\mathcal{M}_v)$ ;

9 **Stage 2: Model-training**

10 **for**  $v \in \mathcal{V}$  **do**

11     $\mathbf{h}_v \leftarrow \text{message\_updater}(\mathbf{c}_v)$ ;

12 **Stage 3: Post-processing**

13 Initialize feature message  $\mathbf{m}_v^0 = \mathbf{h}_v$  for each node  $v \in \mathcal{V}$ ;

14 **for**  $1 \leq t \leq K_{post}$  **do**

15   **for**  $v \in \mathcal{V}$  **do**

16      $\mathbf{m}_v^t \leftarrow \text{graph\_aggregator}(\mathbf{m}_{N_v}^{t-1})$ ;

基于**SGAP** 范式来设计GNN：

## 1. 前处理

对每个节点  $v$ , 变换消息聚合次数  $t$  从 1 到  $K_{pre}$  , 并使用 graph\_aggregator 来聚合 $K_{pre}$ 次邻居节点  $N_v$  的特征。

## 2. 训练

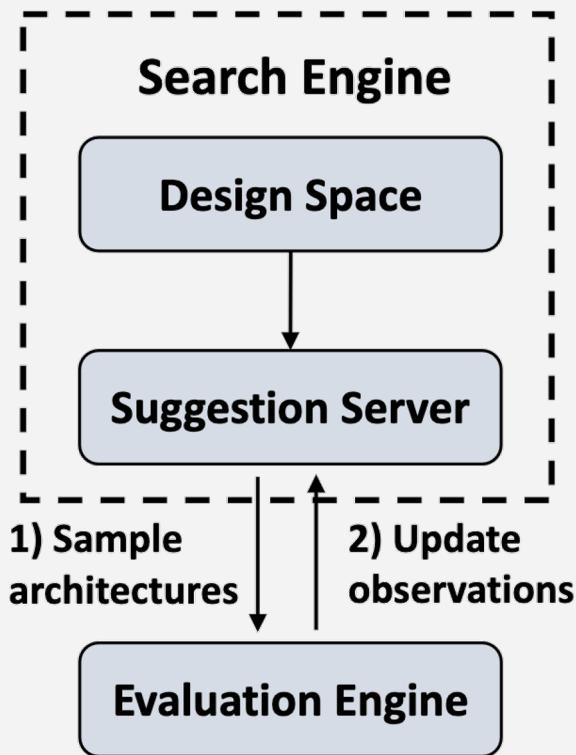
基于聚合后的特征, 利用 message\_updater (如MLP) 来学习节点的软标签类别分布。

## 3. 后处理

讲软标签当做新的特征 , 并使用 graph\_aggregator 来聚合邻居的标签信息  $K_{post}$  次 , 得到最终预测。

# 方法概览

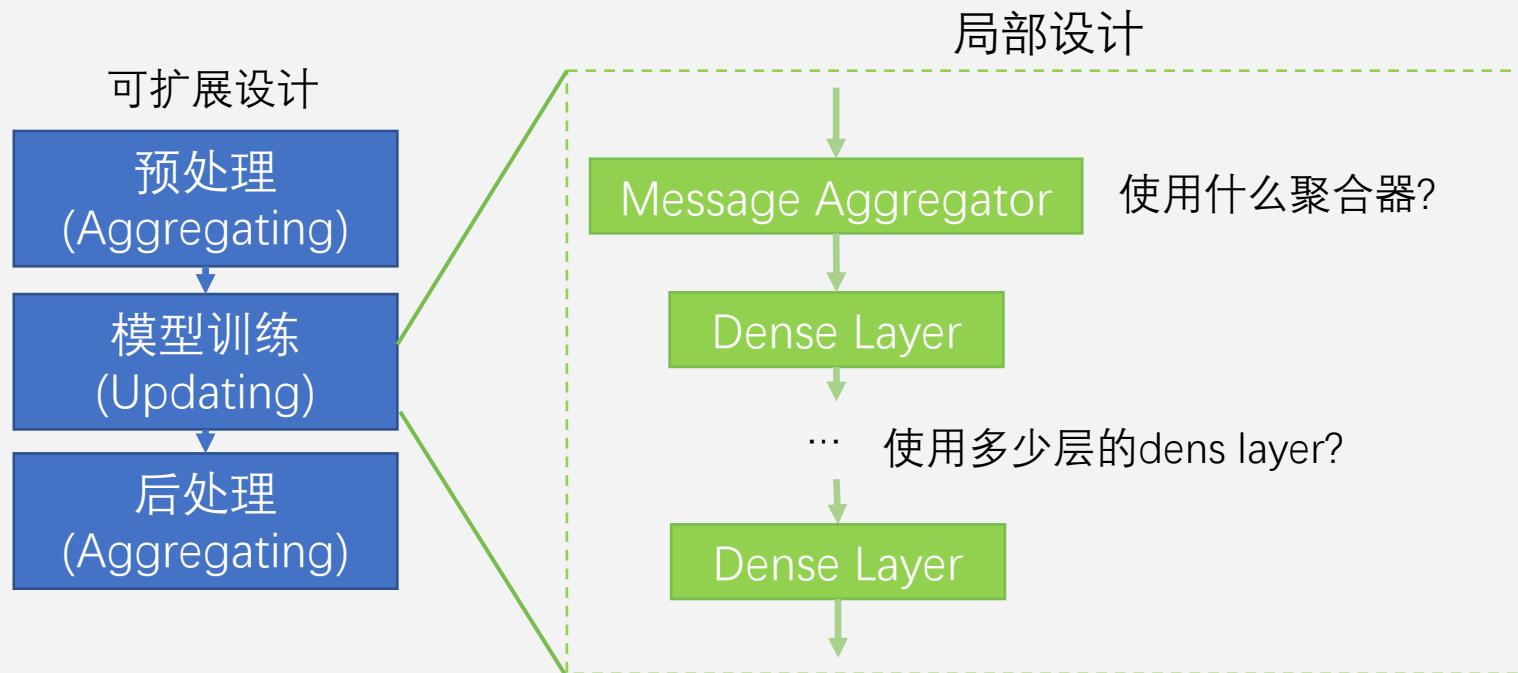
- 可扩展范式 (SGAP)
- 自动化搜索系统 (PaSca)
  - 两个模块
    - (自动化) 搜索引擎
    - (分布式) 评估引擎
  - 搜索引擎推荐一个 configuration instance.
  - 评估引擎 **评估 被推荐的** configuration instance.



**Searching**

# Search Engine ( 搜索引擎 )

- 处理不同优化目标之间的 tradeoff
- 设计空间: 在SGAP 3 个阶段的局部设计 (参数)



# Design Space (设计空间)

- 6个参数可供选择 + 每个阶段2个参数
- 超过 150k 种可能的 configuration instances

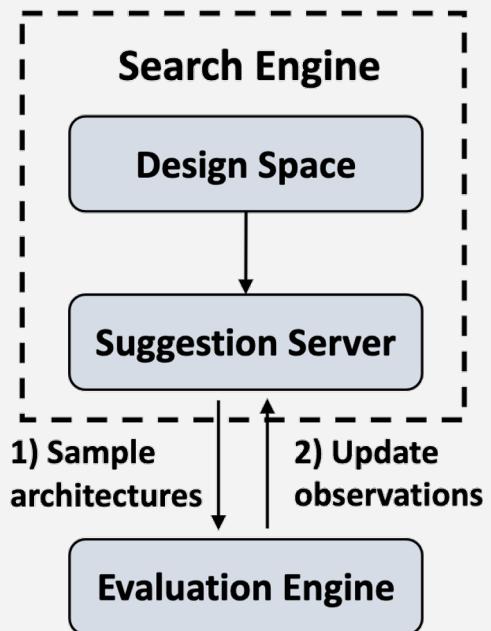
Stages	Name	Range/Choices	Type
Pre-processing	Aggregation steps ( $K_{pre}$ )	[0, 10]	Integer
	Graph aggregators ( $GA_{pre}$ )	{Aug.NA, PPR( $\alpha = 0.1$ ), PPR( $\alpha = 0.2$ ), PPR( $\alpha = 0.3$ ), Triangle. IA}	Categorical
Model training	Message aggregators ( $MA$ )	{None, Mean, Max, Concatenate, Weighted, Adaptive}	Categorical
	Transformation steps ( $K_{trans}$ )	[1, 10]	Integer
Post-processing	Aggregation steps ( $K_{post}$ )	[0, 10]	Integer
	Graph aggregators ( $GA_{post}$ )	{Aug.NA, PPR( $\alpha = 0.1$ ), PPR( $\alpha = 0.2$ ), PPR( $\alpha = 0.3$ ), Triangle. IA}	Categorical

- 现有的Scalable GNN都存在于我们设定的空间中

Models	Pre-processing	Model training		Post-processing
	$GA_{pre}$	$MA$	$K_{trans}$	$GA_{post}$
SGC	Aug.NA	None	1	/
SIGN	Optional	Concatenate	1	/
$S^2GC$	PPR	Mean	1	/
GBP	Aug.NA	Weighted	$\geq 2$	/
PASCA-APPNP	/	/	$\geq 2$	PPR

# Suggestion Server (推荐服务器)

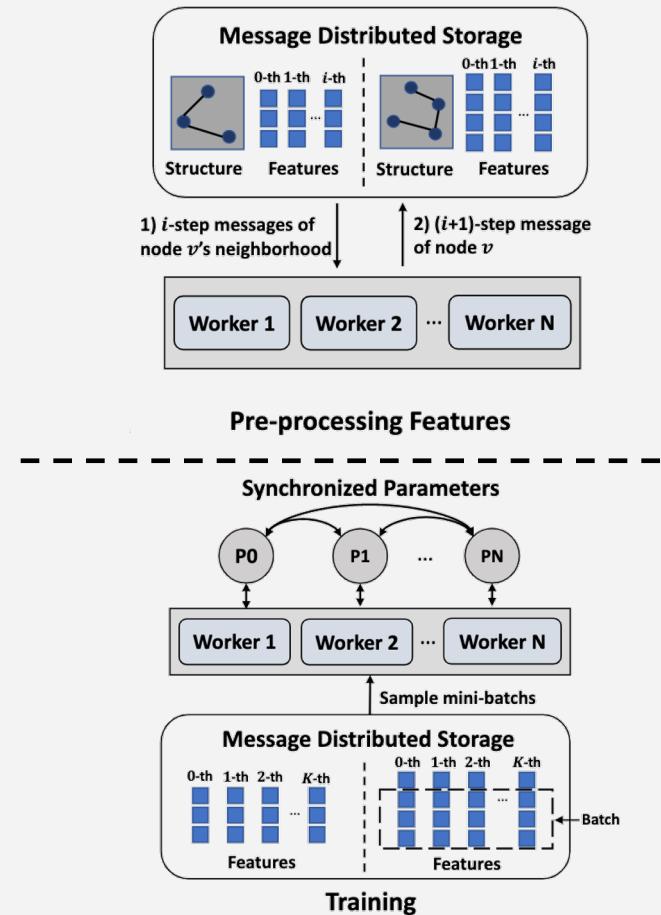
- 建模 配置 和 优化目标之间的关系
- 推荐 能兼顾多个优化目标的配置
- 更新 观测到的历史记录



Searching

# Evaluation Engine (评估引擎)

- Graph data aggregator (图数据聚合器)
  - 切分大图
  - 基于已经计算好的第 (i) 步消息来 计算第(i+1) 步消息
- Neural architecture trainer (网络结构训练器)
  - Mini-batch 训练
  - 基于parameter server的异步网络更新



# 03 实验



# 实验设置

- **数据集**

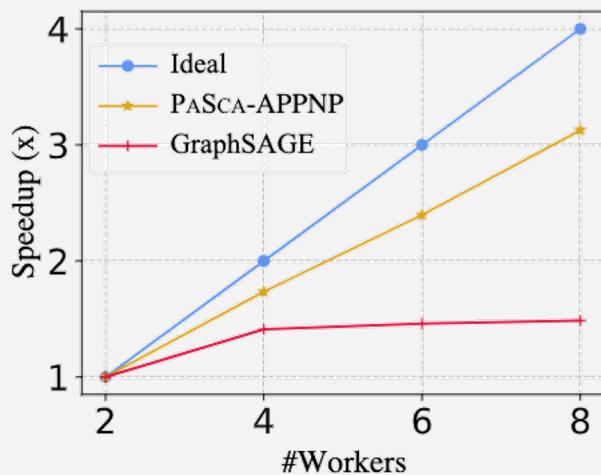
Dataset	#Nodes	#Features	#Edges	#Classes	#Train/Val/Test	Task type	Description
Cora	2,708	1,433	5,429	7	140/500/1000	Transductive	citation network
Citeseer	3,327	3,703	4,732	6	120/500/1000	Transductive	citation network
Pubmed	19,717	500	44,338	3	60/500/1000	Transductive	citation network
Amazon Computer	13,381	767	245,778	10	200/300/12881	Transductive	co-purchase graph
Amazon Photo	7,487	745	119,043	8	160/240/7,087	Transductive	co-purchase graph
ogbn-products	2,449,029	100	61,859,140	47	195922/489811/204126	Transductive	co-purchase network
Coauthor CS	18,333	6,805	81,894	15	300/450/17,583	Transductive	co-authorship graph
Coauthor Physics	34,493	8,415	247,962	5	100/150/34,243	Transductive	co-authorship graph
Flickr	89,250	500	899,756	7	44,625/22,312/22,312	Inductive	image network
Reddit	232,965	602	11,606,919	41	155,310/23,297/54,358	Inductive	social network
Industry	1,000,000	64	1,434,382	253	5,000/10,000/30,000	Transductive	user-video graph

- 验证目标

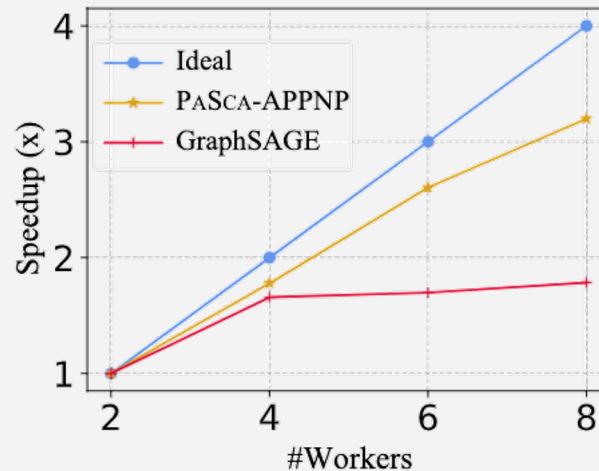
- SGAP 比基于NMP的消息传递机制 **更scalable**,
- PaSca搜索出来的结果能够很好地处理**不同搜索目标之间的tradeoff**,
- 搜索结构能够取得**更高的预测性能**。

# Scalability Analysis (可扩展性分析)

- 对比方法
  - 基于 SGAP的APPNP
  - 基于NMP的GraphSAGE
- 基于 SGAP 的GNN可以取得接近线性的加速比并且更加接近理想的加速比。



Reddit (>230K nodes)



ogbn-product (>2.4M nodes)

# Search Representatives (搜索出来的代表性方法)

- 代表性方法 (在帕累托平面上的)
  - 从SGAP设计空间搜索出来的方法能兼顾多个搜索目标之间的tradeoff。
  - PaSca-V3 取得了最低的预测误差但带来了比PaSca-V2更长的预测时间。
- 我们搜索出来的结果GBP[1], 一个 SOTA 的可扩展网络结构

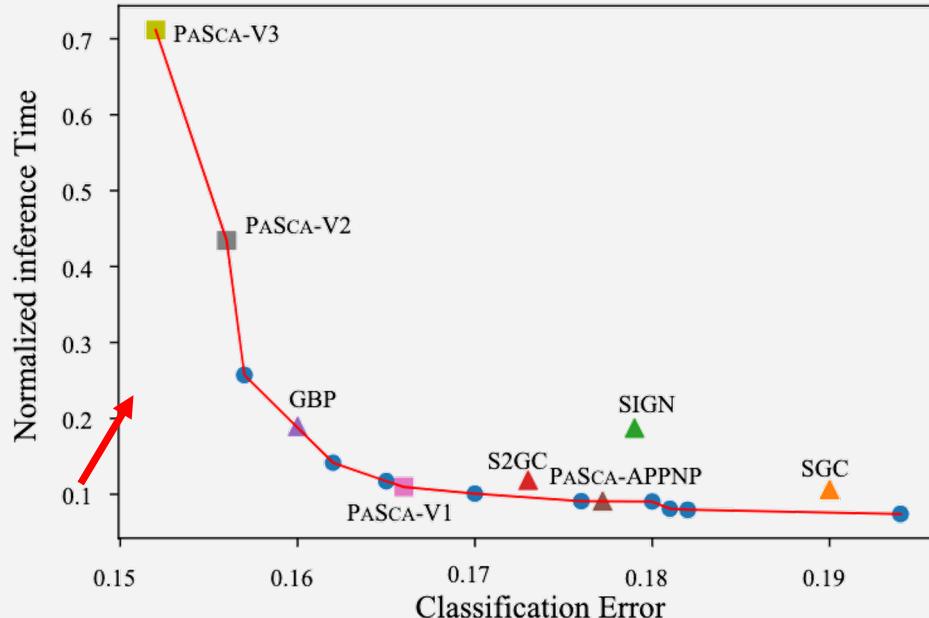


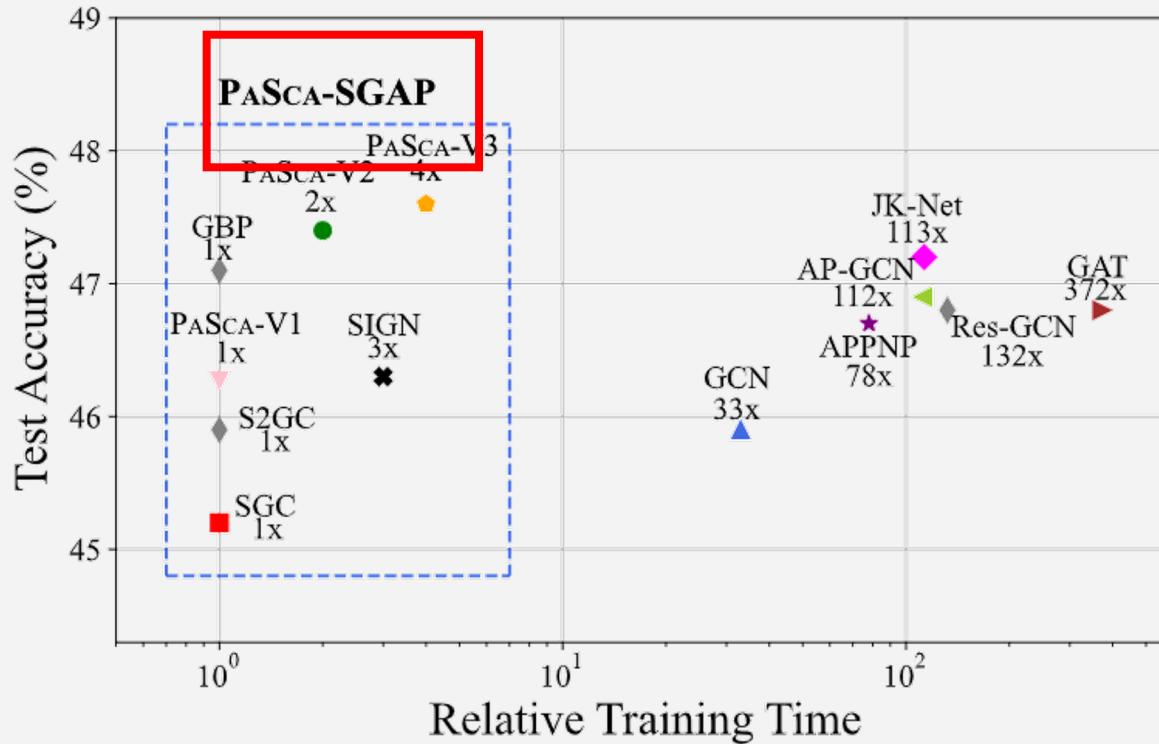
Table 3: Scalable GNNs found by PaSca.

Models	Pre-processing			Model training	Post-processing	
	GA <sub>pre</sub>	MA	K <sub>pre</sub>		GA <sub>post</sub>	K <sub>post</sub>
PASCA-V1	PPR( $\alpha = 0.1$ )	Weighted	3	2	/	/
PASCA-V2	Aug.NA	Adaptive	6	2	/	/
PASCA-V3	Aug.NA	Adaptive	6	3	PPR ( $\alpha = 0.3$ )	4

[1] Chen M, Wei Z, Ding B, et al. 2020. Scalable graph neural networks via bidirectional propagation[J]. In NeurIPS.

## Search Representatives (搜索出来的代表性方法)

- 搜索出来的模型能很好兼顾训练时间与测试准确率。
- PaSca V2 和 V3 都获得了比 JK-Net 更好的准确率，但是只需要明显更少的训练时间。



[1] Xu K, Li C, Tian Y, et al. 2018. Representation learning on graphs with jumping knowledge networks. In ICML.

# 预测性能

- 和其他不scalable的建模范式相比，基于SGAP的网络结构 能取得有竞争力的模型性能。
- PaSca-V3 在不同数据集上都取得了最好的性能。

Type	Models	Cora	Citeseer	PubMed	Amazon Computer	Amazon Photo	Coauthor CS	Coauthor Physics	Industry
NMP	GCN	81.8±0.5	70.8±0.5	79.3±0.7	82.4±0.4	91.2±0.6	90.7±0.2	92.7±1.1	45.9±0.4
	GAT	83.0±0.7	72.5±0.7	79.0±0.3	80.1±0.6	90.8±1.0	87.4±0.2	90.2±1.4	46.8±0.7
	JK-Net	81.8±0.5	70.7±0.7	78.8±0.7	82.0±0.6	91.9±0.7	89.5±0.6	92.5±0.4	47.2±0.3
	ResGCN	82.2±0.6	70.8±0.7	78.3±0.6	81.1±0.7	91.3±0.9	87.9±0.6	92.2±1.5	46.8±0.5
DNMP	APPNP	83.3±0.5	71.8±0.5	80.1±0.2	81.7±0.3	91.4±0.3	92.1±0.4	92.8±0.9	46.7±0.6
	AP-GCN	83.4±0.3	71.3±0.5	79.7±0.3	83.7±0.6	92.1±0.3	91.6±0.7	93.1±0.9	46.9±0.7
SGAP	SGC	81.0±0.2	71.3±0.5	78.9±0.5	82.2±0.9	91.6±0.7	90.3±0.5	91.7±1.1	45.2±0.3
	SIGN	82.1±0.3	72.4±0.8	79.5±0.5	83.1±0.8	91.7±0.7	91.9±0.3	92.8±0.8	46.3±0.5
	S <sup>2</sup> GC	82.7±0.3	73.0±0.2	79.9±0.3	83.1±0.7	91.6±0.6	91.6±0.6	93.1±0.8	45.9±0.4
	GBP	83.9±0.7	72.9±0.5	80.6±0.4	83.5±0.8	92.1±0.8	92.3±0.4	93.3±0.7	47.1±0.6
	PaSca-V1	83.4±0.5	72.2±0.5	80.5±0.4	83.7±0.7	92.1±0.7	91.9±0.3	93.2±0.6	46.3±0.4
	PaSca-V2	84.4±0.3	73.1±0.3	80.7±0.7	84.1±0.7	92.4±0.7	92.6±0.4	93.6±0.8	47.4±0.6
	PaSca-V3	<b>84.6±0.6</b>	<b>73.4±0.5</b>	<b>80.8±0.6</b>	<b>84.8±0.7</b>	<b>92.7±0.8</b>	<b>92.8±0.5</b>	<b>93.8±0.9</b>	<b>47.6±0.3</b>

# 04 总结



# 系统应用

- 实现了能自动化建模10亿节点的**超大规模图神经网络系统**，部署于腾讯太极机器学习平台，并广泛应用于**视频推荐和内容风控等场景**
- 系统部分功能已在Github开源：<https://github.com/PKU-DAIR/SGL>
- 系统论文获得**CCF A类数据挖掘旗舰会议WWW 2022 唯一“最佳学生论文奖”**（中国第2个）
- 系统相关工作刷新了国际图学习榜单**OGB的3项第一**

Leaderboard for [ogbn-mag](#)  
The classification accuracy on the test and validation sets. The higher, the better.  
Package: >=1.2.1

Rank	Method	Test Accuracy	Validation Accuracy	Contact	References	#Params	Hardware	Date
1	NARS-GAMLP+RLU	0.5590 ± 0.0027	0.5702 ± 0.0041	Wentao Zhang (PKU Tencent Joint Lab)	<a href="#">Paper</a> , <a href="#">Code</a>	6,734,882	Tesla V100 (32GB)	Aug 19, 2021

**Best Student Paper Award**  
**PaSca: a Graph Neural Architecture Search System under the Scalable Paradigm**  
*Systems and Infrastructure Track*



Wentao Zhang (Peking University), Yu Shen (Peking University), Zheyu Lin (Peking University), Yang Li (Peking University), Xiaosen Li (Tencent), Wen Ouyang (Tencent), Yangyu Tao (Tencent), Zhi Yang (Peking University), Bin Cui (Peking University)

<https://www2022.thewebconf.org/awards/>

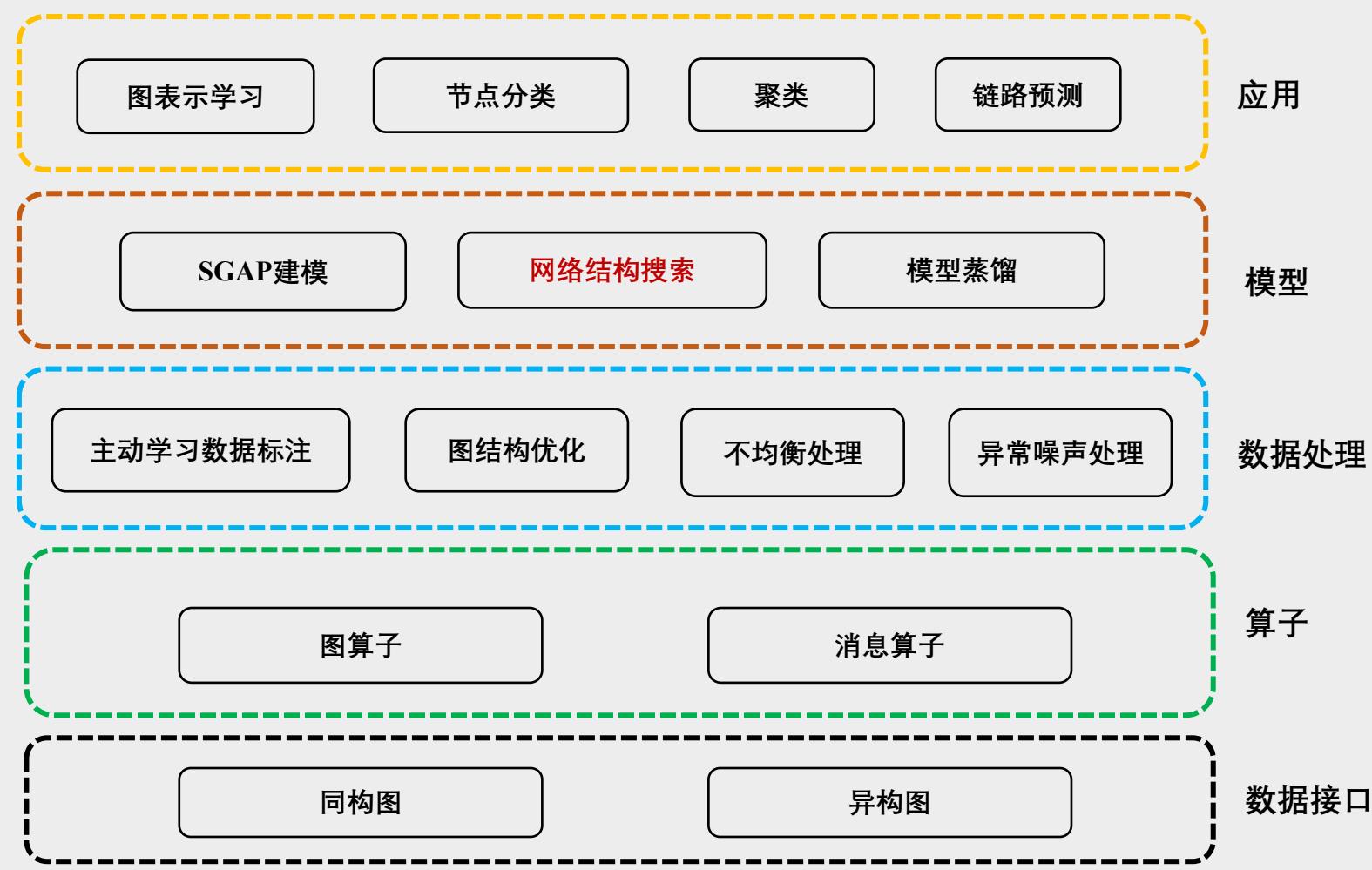
刷新国际图学习榜单  
[https://ogb.stanford.edu/docs/leader\\_nodeprop/](https://ogb.stanford.edu/docs/leader_nodeprop/)

# 总结

- 我们设计了PaSca, 一个新颖的构建和探索可扩展 GNNs的网络结构搜索系统，而不是仅研究单个的网络结构设计。
- PaSca搜索出来的代表性模型能够在预测性能、效率以及可扩展性等多个方面超越现有的SOTA GNN 模型。
- PaSca能够帮助研究者来探索不同的Scalable GNN结构设计，并且理解不同设计的特点和功能。

# 系统开源

<https://github.com/PKU-DAIR/SGL>



## SGL系统设计目标

### 1. 高可扩展性:

基于SGAP, SGL能处理超大规模图数据

### 2. 自动化:

根据指定的多个目标自动化搜索网络结构

### 3. 易用性:

针对多个任务定制的用户友好的接口

### 4. 针对数据的优化

多种数据处理操作

### 5. Bag of Tricks

内置多种有效的提点方法

# 非常感谢您的观看

Tencent  
腾讯

DataFun.

