



HAL
open science

Bridging the Gap of AutoGraph Between Academia and Industry: Analyzing AutoGraph Challenge at KDD Cup 2020

Zhen Xu, Lanning Wei, Huan Zhao, Rex Ying, Quanming Yao, Wei-Wei Tu,
Isabelle Guyon

► **To cite this version:**

Zhen Xu, Lanning Wei, Huan Zhao, Rex Ying, Quanming Yao, et al.. Bridging the Gap of AutoGraph Between Academia and Industry: Analyzing AutoGraph Challenge at KDD Cup 2020. 2022. hal-03629994

HAL Id: hal-03629994

<https://hal.science/hal-03629994>

Preprint submitted on 4 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BRIDGING THE GAP OF AUTOGRAPH BETWEEN ACADEMIA AND INDUSTRY: ANALYSING AUTOGRAPH CHALLENGE AT KDD CUP 2020

A PREPRINT

Zhen Xu
4Paradigm, China
xuzhen@4paradigm.com

Lanning Wei
4Paradigm, China
Institute of Computing Technology,
Chinese Academy of Sciences, China
weilanning@4paradigm.com

Huan Zhao
4Paradigm, China
zhaohuan@4paradigm.com

Rex Ying
Stanford University, USA
rexying@stanford.edu

Quanming Yao
Tsinghua University, China
qyaoaa@tsinghua.edu.cn

Wei-Wei Tu
4Paradigm, China
tuweiwei@4paradigm.com

Isabelle Guyon
ChaLearn, USA
LISN/INRIA/CNRS, University Paris-Saclay, France
guyon@chalearn.org

ABSTRACT

Graph structured data is ubiquitous in daily life and scientific areas and has attracted increasing attention. Graph Neural Networks (GNNs) have been proved to be effective in modeling graph structured data and many variants of GNN architectures have been proposed. However, much human effort is often needed to tune the architecture depending on different datasets. Researchers naturally adopt Automated Machine Learning on Graph Learning, aiming to reduce the human effort and achieve generally top-performing GNNs, but their methods focus more on the architecture search. To understand GNN practitioners' automated solutions, we organized AutoGraph Challenge at KDD Cup 2020, emphasizing on automated graph neural networks for node classification. We received top solutions especially from industrial tech companies like Meituan, Alibaba and Twitter, which are already open sourced on Github. After detailed comparisons with solutions from academia, we quantify the gaps between academia and industry on modeling scope, effectiveness and efficiency, and show that (1) academia AutoML for Graph solutions focus on GNN architecture search while industrial solutions, especially the winning ones in the KDD Cup, tend to obtain an overall solution (2) by neural architecture search only, academia solutions achieve on average 97.3% accuracy of industrial solutions (3) academia solutions are cheap to obtain with several GPU hours while industrial solutions take a few months' labors. Academic solutions also contain much fewer parameters.

Keywords Graph Neural Networks · Automated Machine Learning · Data Challenge · Node Classification

1 Introduction

Graph structured data has been prominent in our life and various tasks are studied based upon, including recommendation on Social Networks [Fan et al., 2019], traffic forecasting on road networks [Li et al., 2018], drug discovery on molecule graph [Tornig and Altman, 2019], link prediction on knowledge graph [Zhang et al., 2020], etc. Graph Neural Networks (GNN) [Kipf and Welling, 2017] have been proved to be effective in modeling graph data and tremendous GNN architectures are proposed every year [Hamilton et al., 2017, Xu et al., 2019, Wu et al., 2019, Veličković et al., 2018].

When applying GNN on graph structured data, expertise and domain knowledge is often required and numerous human effort is required to adapt to new datasets. Automated Machine Learning (AutoML) [Hutter et al., 2019, Yao et al., 2018] aims to reduce human efforts in deploying on various applications. AutoML, especially Neural Architecture Search (NAS), has been successfully explored on tremendous applications, including Image Classification [Tan and Le, 2019], Object Detection [Tan et al., 2020], Semantic Segmentation [Nekrasov et al., 2019], Language Modeling [Jiang et al., 2019], Time Series Forecasting [Chen et al., 2021], etc. As a result, researchers start to explore Automated Graph Neural Networks (AutoGraph). AutoGraph researchers focus mainly on automatically designing GNN architectures by NAS. The majority of these methods focus on designing the aggregation functions/layers in GNNs with different search algorithms [Gao et al., 2020, Zhou et al., 2019, Yoon et al., 2020, Li et al., 2021, Peng et al., 2020]. Other works, SANE [Zhao et al., 2021] and AutoGraph [Li and King, 2020], provide the extra layer-wise skip connections design dimension; GNAS [Cai et al., 2021], DeepGNAS [Feng et al., 2021] and Policy-GNN [Lai et al., 2020] learn to design the depth of GNNs. DiffMG [Ding et al., 2021] proposed to use NAS to search data-specific meta-graphs in heterogeneous graph, and PAS [Wei et al., 2021] is proposed to search data-specific pooling architectures for graph classification. The recently proposed F²GNN [Wei et al., 2022] method decouples the design of aggregation operations with architecture topology, which is not considered before.

Despite the rich literature from academia, we ask the question of how AutoGraph is used in industrial practitioners. Towards this end, we organize the first AutoGraph challenge at KDD Cup 2020, collaborated with 4Paradigm, ChaLearn and Stanford University. This challenge asks participants to provide AutoGraph solutions for node classification task. The code is executed by the platform on various graph datasets without any human intervention. Through the AutoGraph challenge, we wish to push forward the limit of AutoGraph as well as to understand the gap between industrial solutions and academia ones. In this paper, we first introduce the AutoGraph challenge setting. Then, we present the winning solutions which are open sourced for everyone to use. At last, we experiment further and compare with NAS methods for GNN methods and quantify empirically the gap with respect to top solutions. **We conclude three gaps of AutoGraph between academia and industry: Modeling scope, Effectiveness and Efficiency.**

2 Challenge background

2.1 General statistics

The AutoGraph challenge lasted for two months. We received over 2200 submissions and more than 140 teams from both high-tech companies (Ant Financial, Bytedance, Criteo, Meituan Dianping, Twitter, NTT DOCOMO, etc.) and universities (MIT, UCLA, Tsinghua University, Peking University, Nanyang Technological University, National University of Singapore, IIT Kanpur, George Washington University, etc.), coming from various countries. The top three teams are: *aister*, *PASA_NJU*, *qquerret*. Top 10 winners’ information are shown in Table 1. The 1st winner *aister* comes from Meituan Dianping, a company on location-based shopping and retailing service. This makes the challenge particularly valuable since we can compare academic solutions with industrial best AutoGraph practices.

Table 1: General information about winning teams. Two teams tie on the 6th and 10th place. We list them both.

| Place | Team Name | Institute | Place | Team Name | Institute |
|-------|-----------------|---------------------|-------|--|---|
| 1st | <i>aister</i> | Meituan Dianping | | | |
| 2nd | <i>PASA_NJU</i> | Nanjing University | 6th | <i>SmartMN-THU</i> <i>JunweiSun</i> | Tsinghua University Beijing University of Posts and Telecommunications |
| 3rd | <i>qquerret</i> | Ant Financial | 8th | <i>u1234x1234</i> | self-employed |
| 4th | <i>common</i> | Alibaba Inc. | 9th | <i>AML</i> | Ant Financial |
| 5th | <i>PostDawn</i> | Zhejiang University | 10th | <i>supergx</i> <i>daydayup</i> | Nanyang Tech. University Hikvision Inc. |

2.2 Problem formulation

The task of AutoGraph challenge is node classification under the transductive setting. Formally speaking, consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, \dots, v_N\}$ is the set of nodes, i.e. $|\mathcal{V}| = N$ and \mathcal{E} is the set of edges, which is usually encoded by an adjacency matrix $A \in [0, 1]^{N \times N}$. A_{ij} is positive if there is an edge connecting from node v_i to node v_j . Additionally, a feature matrix $X \in \mathbb{R}^{N \times D}$ gives features of each node. Each node v_i has a class label $y_i \in \mathcal{L} = \{1, \dots, c\}$, resulting in the label vector $Y \in \mathcal{L}^N$. In the transductive semi-supervised node classification

task, part of labels are available during training and the goal is to learn a mapping $\mathcal{F} : \mathcal{V} \rightarrow \mathcal{L}$ and predict classes of unlabeled nodes.

2.3 Protocol

The protocol of AutoGraph challenge is straightforward. Participants should submit a python file containing a `Model` class with required `fit` and `predict` method. We prepare an ingestion program reading dataset and instantiate the class and call `fit` and `predict` method until prediction finishes or the running time has reached the budget limit. Ingestion program outputs model’s prediction on test data and save to a shared space. Then, another scoring program reads the prediction and ground truth and outputs evaluation scores. The execution of the program is totally on the challenge platform. When developing locally, we provide script to call `model.py` file methods directly.

2.4 Metric

We use Accuracy (Acc) and Balanced Accuracy (BalAcc) as evaluation metrics, defined as

$$\text{Acc} = \frac{1}{|\Omega|} \sum_{i \in \Omega} \mathbb{1}_{\hat{y}_i = y_i}, \quad \text{BalAcc} = \frac{1}{|C|} \sum_{i \in C} \text{Recall}_i,$$

where Ω is the set of test nodes indexes, y_i is the ground truth label for node v_i and \hat{y}_i is the predicted label, C is the set of classes and Recall_i is the recall score for class i . Accuracy (Acc) is used in the challenge to rank participants and Balanced Accuracy (BalAcc) is applied for additional analyses since it takes into account the imbalanced label distribution of datasets.

2.5 Datasets

Fifteen graph datasets were used during the competition: 5 public datasets were directly downloadable by the participants so they could develop their solutions offline. Five feedback datasets were made available on the platform during the feedback phase to evaluate AutoGraph algorithms on the public leaderboard. Finally, the AutoGraph algorithms were evaluated with 5 private datasets, without human intervention. These dataset are quite diverse in domains, shapes, density and other graph properties because we expect AutoGraph solutions to have good generalization ability. On the other hands, we intentionally keep the characteristics of 5 feedback datasets and 5 private datasets similar to enable transferability. We summarize dataset statistics in Table 2. The licenses and original sources of these datasets are also provided¹.

Table 2: **Statistics of all datasets.** “Avg Deg” is the average number of edges per node. “Directed” and “Weighted” indicate the two properties of a graph. “Skewness” here is calculated by number of nodes in the largest class divided by number of nodes in the smallest class.

| Dataset | Phase | Domain | #Node | #Edge | #Feature | #Class | Avg Deg | Directed? | Weighted? | Skewness |
|---------|----------|----------|-------|--------|----------|--------|---------|-----------|-----------|----------|
| a | Public | Citation | 2.7K | 5.3K | 1.4K | 7 | 1.9 | F | F | 5 |
| b | Public | Citation | 3.3K | 4.6K | 3.7K | 6 | 1.4 | F | F | 3 |
| c | Public | Social | 10K | 733K | 0.6K | 41 | 73.3 | F | F | 81 |
| d | Public | News | 10K | 2,917K | 0.3K | 20 | 291.7 | T | T | 467 |
| e | Public | Finance | 7.5K | 7.8K | 0 | 3 | 1.0 | F | F | 111 |
| f | Feedback | Sales | 10K | 194K | 0.7K | 10 | 19.4 | F | F | 18 |
| g | Feedback | Citation | 10K | 41K | 8K | 5 | 4.1 | F | F | 6 |
| h | Feedback | Medicine | 10K | 2,461K | 0.3K | 23 | 246.1 | T | T | 1,773 |
| i | Feedback | Finance | 15K | 16K | 0 | 3 | 1.1 | F | F | 213 |
| j | Feedback | Medicine | 11K | 22K | 0 | 9 | 2.0 | F | F | 227 |
| k | Private | Sales | 8K | 119K | 0.7K | 8 | 14.9 | F | F | 6 |
| l | Private | Citation | 10K | 40K | 7K | 15 | 4 | F | F | 34 |
| m | Private | News | 10K | 1,425K | 0.3K | 8 | 142.5 | T | T | 360 |
| n | Private | Finance | 14K | 22K | 0 | 10 | 1.6 | F | F | 61 |
| o | Private | Social | 12K | 19K | 0 | 19 | 1.6 | F | F | 62 |

¹<https://github.com/AutoML-Research/AutoGraph-KDDCup2020>

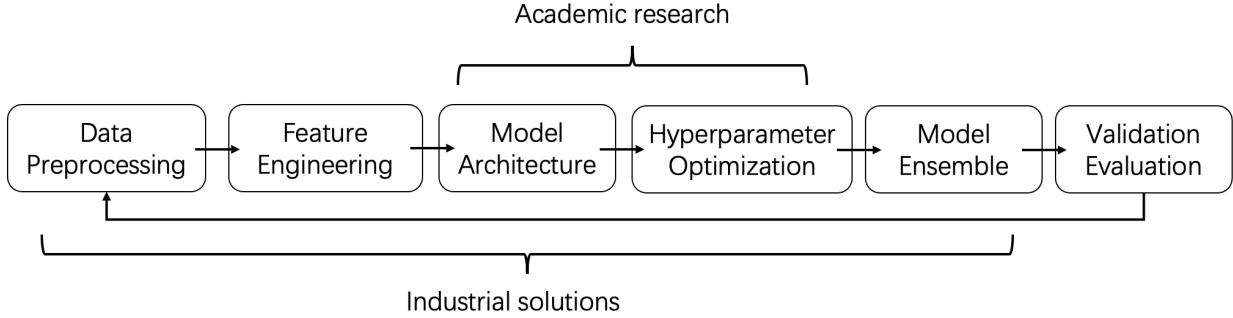


Figure 1: Illustration of AutoGraph scope. Industrial people provide a full pipeline solution that covers from data preprocessing to evaluation. Academic researchers focus mainly on model architecture and hyperparameter optimization.

3 Solutions

In this part, we introduce various methods suitable for the AutoGraph challenge, including the provided challenge baseline and solutions from top-3 winners. We conclude the first gap at the end.

Baselines (GCN(L2)). In the provided baselines, there is no feature engineering except for using the raw node features. For graph without node features, e.g. dataset i,j, one hot encoding is used to unroll the node lists to a dummy feature table. During model training, a MLP is first used to map node features to the same embedding dimension. Then a two layer vanilla GCN is applied for learning node embeddings. Another MLP with softmax outputs the final classification. Dropout is used. All the hyperparameters are fixed by experience. No time management since the model is simple and one full training will not cost more than the allowed time budget.

1st placed winner. The 1st winner is from team aister. Their code is open source here². The authors use four GNN models, two spatial ones: GraphSage [Hamilton et al., 2017] and GAT [Veličković et al., 2018], two spectral ones: GCN [Kipf and Welling, 2017] and TAGConv [Du et al., 2017] to process node features collectively. For each GNN model, a heavy search is applied offline to determining the important hyperparameters as well as the boundaries. In the online stage, they use a smaller search space to determine the hyperparameters. In order to accelerate the search, they do not fully train each configuration but instead early stop in 16 epochs if the validation loss is not satisfactory. Additional features are used: node degrees, distribution of 1-hop and 2-hop neighbor nodes’ features, etc.

2nd place winner. The 2nd winner is from team PASA_NJU. Their code is open source here³. They also split the solution in two stages: offline stage and online stage. In the offline stage, the authors train a decision tree based on public data and other self collected datasets to classify graph type into one of three classes. Then they use GraphNAS [Gao et al., 2020] to search massively optimal GNN architectures including aggregation function, activation, number of heads in attention, hidden units, etc. In the online stage, the authors rapidly classify the dataset and fine tune the offline searched model.

3rd place winner. The 3rd winner is from team qqerret. Their code is open source here⁴. The core model is a variant of spatial based GNN, which aggregates 2-hop neighbors of a node with additional linear parts for the node itself. Basically, the new embedding of node i is $\hat{h}(i) = \sum_{j \in N_2(i)} a_j h(j) + \alpha(wh(i) + b)$. Additionally, in the GNN output layer, a few features per node are concatenated for final fully connected layer, including number of edges, whether this node connects to a central node who has a lot of edges, label distribution of 1-hop neighbor nodes, and label distribution of 2-hop neighboring nodes.

²https://github.com/aister2020/KDDCUP_2020_AutoGraph_1st_Place

³<https://github.com/Unkrible/AutoGraph2020>

⁴https://github.com/white-bird/kdd2020_GCIN

4 Results

We conduct additional experiments after the AutoGraph challenge to further analyze the results. We first reproduce winning solutions and then we compare with academia solutions. Three gaps are concluded. The first gap is presented as follows and two other gaps are concluded in Sec 4.2.

Gap #1: Modeling scope is the first gap of AutoGraph between academia and industry. In academia, researchers focus mainly on Neural Architecture Search methods to find better GNN architectures. Their contributions differ in their search space, search strategy and evaluation methods. However, industrial solutions, e.g. 1st solution, focus more on the feature engineering and model ensemble. For GNN architectures, they prefer existing found ones with little modification. In other words, industrial people provide a full pipeline solution including data preprocessing, feature engineering, model architecture, hyperparameter optimization, model ensemble, while academia researchers focus on model architecture part only. The gap is also illustrated in Figure 1. It might be an interesting direction for both groups to merge, i.e. AutoGraph researchers could explore the automated feature engineering, automated ensemble and AutoGraph practitioners could adopt NAS methods for GNN.

4.1 Reproducing winning solutions

We reproduce all winning methods on all the datasets and include their results in Table 3. We observe that all three winning solutions are close in performance and all significantly beat the GCN baselines. On the other hand, in the AutoGraph challenge, due to the nature of the competition, we rank methods based on accuracy, we state that this is not sufficient to evaluate comprehensively solutions from the scientific perspective. We add balanced accuracy here just to show that for some methods that show close performance in accuracy, they could diverge a lot in balanced accuracy. Regarding both accuracy and balanced accuracy, we conclude that 1st solution which comes from Meituan Dianping Company, is indeed best among top winners. Thus, we will later use their solutions for comparing with academia solutions. These winning solutions are already open sourced, which are reproducible and lower the barriers of using AutoGraph.

Table 3: Accuracy and Balanced accuracy of top methods on all datasets (%). Baseline is a two layer GCN.

| Dataset | Phase | Baseline (GCN(L2)) | | 1st solution | | 2nd solution | | 3rd solution | |
|---------|----------|--------------------|--------|--------------|-------------|--------------|-------------|--------------|-------------|
| | | Acc | BalAcc | Acc | BalAcc | Acc | BalAcc | Acc | BalAcc |
| a | Public | 85.7 | 84.9 | 88.5 | 87.8 | 88.2 | 87.2 | 87.2 | 85.5 |
| b | Public | 71.4 | 67.8 | 75.2 | 71.2 | 75.8 | 71.2 | 75.6 | 69.0 |
| c | Public | 86.5 | 72.0 | 94.3 | 87.5 | 94.2 | 90.9 | 95.4 | 91.3 |
| d | Public | 93.7 | 6.1 | 96.5 | 48.7 | 95.1 | 28.8 | 94.6 | 21.0 |
| e | Public | 59.6 | 38.8 | 88.7 | 92.8 | 88.5 | 90.7 | 88.8 | 92.8 |
| f | Feedback | 86.6 | 78.2 | 92.8 | 92.1 | 92.3 | 92.1 | 92.4 | 91.4 |
| g | Feedback | 94.7 | 92.8 | 95.3 | 93.5 | 95.6 | 93.8 | 95.8 | 94.2 |
| h | Feedback | 90.4 | 8.8 | 93.5 | 26.3 | 92.2 | 17.6 | 92.1 | 16.6 |
| i | Feedback | 88.2 | 59.2 | 88.4 | 87.5 | 88.4 | 92.6 | 88.5 | 91.1 |
| j | Feedback | 90.7 | 68.1 | 95.9 | 89.0 | 96.1 | 93.7 | 96.6 | 93.3 |
| k | Private | 93.5 | 92.2 | 95.5 | 94.4 | 95.5 | 94.4 | 94.8 | 93.1 |
| l | Private | 90.9 | 84.5 | 94.9 | 92.6 | 94.7 | 91.8 | 94.5 | 92.6 |
| m | Private | 85.5 | 24.5 | 98.1 | 79.7 | 95.7 | 69.0 | 98.0 | 79.4 |
| n | Private | 85.6 | 47.3 | 99.0 | 97.3 | 99.0 | 98.4 | 98.9 | 97.0 |
| o | Private | 49.6 | 15.6 | 91.0 | 84.6 | 91.3 | 90.6 | 91.4 | 88.5 |

4.2 Neural Architecture Search for GNN

We further adopt NAS methods for GNN and compare with the baseline and 1st solution coming from industry. We choose the recent F²GNN [Wei et al., 2022], which searches for data-specific GNN topology, in our experiment. To compare fairly with GCN baselines, we fix the aggregation to GCN and search only the GNN topology, which we call F²GCN. Since F²GCN requires at least 4 layers, we also run a 4 layer GCN baseline for better comparison. The results are given in Table 4.

| Dataset | GCN(L2) | GCN(L4) | F ² GCN(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, F²GCN 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.

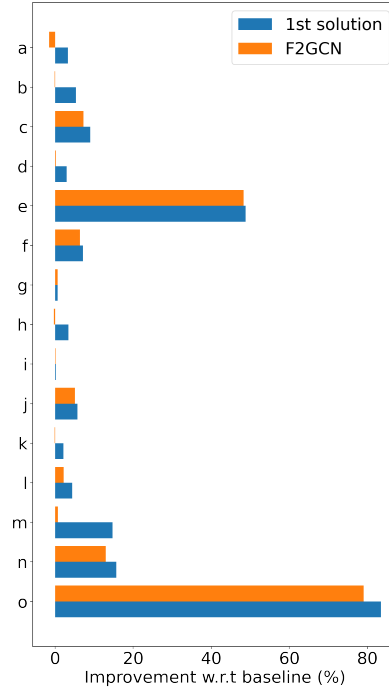


Figure 2: Accuracy improvement with respect to baseline.

Gap #2: Effectiveness is the second gap of AutoGraph between academia and industry. We observe from Table 4 and Figure 2 that all baselines and F²GCN methods are not as good as 1st winning solution. However, for many datasets, e.g. e, f, g, i, j, F²GCN is very close to the best industrial solution. On average, F²GCN which focuses only on architecture search, reaches 97.3% of best solution. Note that the 1st solution constructs additional node features and uses multiple GNN architectures for ensemble while F²GCN does not use any feature engineering or model ensemble. This shows the effectiveness of winner’s engineering practices as well as F²GCN’s adaptive NAS search. Winning teams also have access to public datasets and public leaderboard to iteratively fine tune their methods. F²GCN does not assume any prior knowledge of the datasets, which shows further its effectiveness.

To better understand the solutions, we calculate the number of parameters of baseline, F²GCN, and 1st solution, as shown in Table 5.

Gap #3: Efficiency is the third gap of AutoGraph between academia and industry. From Table 5 and Figure 3, F²GCN uses significantly fewer parameters than the best industrial solution on most datasets (13 out of 15). On average, F²GCN consumes 45.1% of the 1st solution in terms of parameter size, which is quite resource efficient. Note that feature engineering and ensemble do not contain additional parameters and basically, F²GCN searches one GNN model to compete with ensemble of 4 types of GNN models in 1st solution. As for time devotion, winning solutions come from a team’s months of work, which consists of 5 or more members. F²GCN only runs for a few GPU hours per dataset, demonstrating its time efficiency compared to industrial solutions.

| Dataset | GCN(L2) | F ² GCN(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 F²GCN (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.

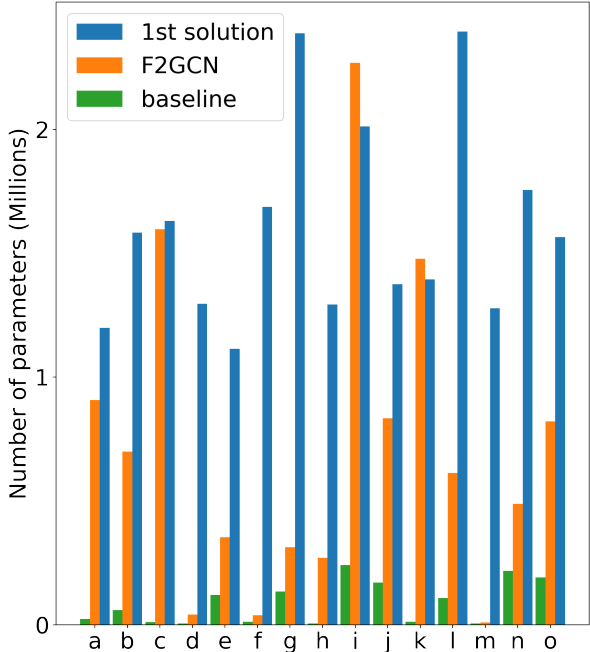


Figure 3: Comparison on number of parameters of baseline, 1st solution and F²GCN.

5 Conclusion

We organized the first Automated Graph Learning (AutoGraph) Challenge at KDD Cup 2020. We presented in this paper its settings, dataset, and solutions, which are all open sourced. We ran additional post-challenge experiments to compare the baseline (Graph Convolution Network (GCN)), the winning solution (feature engineering-based ensemble of various Graph Neural Networks), and a recent and efficient Neural Architecture Search (NAS) for GNN method called F²GCN. This paper provides results that could bridge the gap between academic research and industry practices, by correcting bias favoring certain approaches. This gap is currently at 3 aspects: **Gap #1 modeling scope.** (academia focuses more on model-centric approaches, emphasizing NAS; industry emphasizes data centric approaches and feature engineering); **Gap #2 effectiveness.** (academic solutions are perceived by industry to be less effective than their industry counterpart); **Gap #3 efficiency.** (academic solutions are perceived to be parsimonious or slower than industry solutions). Our results indicate that the “academic” NAS-based approach that we applied attains performances closely matching those of the winning industry solution, while being both faster and more parsimonious in number of parameters, therefore closing Gap #2 and #3. Moreover, we hope that these results will help reducing Gap #1, by encouraging industry practitioners to apply NAS methods (and particularly F²GCN), eventually combining the best of both approaches. We believe the results we obtained are significant, since they involve a benchmark on 15 datasets.

Acknowledgments

Funding and support have been received by several research grants, including 4Paradigm, Big Data Chair of Excellence FDS Paris-Saclay, Paris Région Ile-de-France, and ANR Chair of Artificial Intelligence HUMANIA ANR-19-CHIA-0022, ChaLearn, Microsoft, Google. We acknowledge the following people for helping organize AutoGraph challenge: Xiawei Guo, Shouxiang Liu. We also appreciate the following people and institutes for open sourcing datasets which are used in our use cases: Andrew McCallum, C. Lee Giles, Ken Lang, Tom Mitchell, William L. Hamilton, Maximilian

Mumme, Oleksandr Shchur, David D. Lewis, William Hersh, Just Research and Carnegie Mellon University, NEC Research Institute, Carnegie Mellon University, Stanford University, Technical University of Munich, AT&T Labs, Oregon Health Sciences University.

References

- Wenqi Fan, Yao Ma, Qing Li, Yuan He, Yihong Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation. In *WWW*, 2019.
- Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *ICLR*, 2018.
- Wen Torng and Russ B. Altman. Graph convolutional neural networks for predicting drug-target interactions. *Journal of Chemical Information and Modeling*, 2019.
- Zhao Zhang, Fuzhen Zhuang, Hengshu Zhu, Zhi-Ping Shi, Hui Xiong, and Qing He. Relational graph neural network with hierarchical attention for knowledge graph completion. In *AAAI*, 2020.
- Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.
- Will Hamilton, Zitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *NIPS*, 2017.
- Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *ICLR*, 2019.
- Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *ICML*, 2019.
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.
- Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.
- Quanming Yao, Mengshuo Wang, Hugo Jair Escalante, Isabelle Guyon, Yi-Qi Hu, Yu-Feng Li, Wei-Wei Tu, Qiang Yang, and Yang Yu. Taking human out of learning applications: A survey on automated machine learning. 2018.
- Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *ICML*, 2019.
- Mingxing Tan, Ruoming Pang, and Quoc V. Le. Efficientdet: Scalable and efficient object detection. In *CVPR*, 2020.
- Vladimir Nekrasov, Hao Chen, Chunhua Shen, and Ian D. Reid. Fast neural architecture search of compact semantic segmentation models via auxiliary cells. In *CVPR*, 2019.
- Yufan Jiang, Chi Hu, Tong Xiao, Chunliang Zhang, and Jingbo Zhu. Improved differentiable architecture search for language modeling and named entity recognition. In *EMNLP-IJCNLP*, 2019.
- Donghui Chen, Ling Chen, Zongjiang Shang, Youdong Zhang, Bo Wen, and Chenghu Yang. Scale-aware neural architecture search for multivariate time series forecasting. *CoRR*, 2021.
- Yang Gao, Hong Yang, Peng Zhang, Chuan Zhou, and Yue Hu. Graph neural architecture search. In *IJCAI*, 2020.
- Kaixiong Zhou, Qingquan Song, Xiao Huang, and Xia Hu. Auto-gnn: Neural architecture search of graph neural networks. *arXiv preprint arXiv:1909.03184*, 2019.
- Minji Yoon, Théophile Gervet, Bryan Hooi, and Christos Faloutsos. Autonomous graph mining algorithm search with best speed/accuracy trade-off. *ICDM*, 2020.
- Yanxi Li, Zean Wen, Yunhe Wang, and Chang Xu. One-shot graph neural architecture search with dynamic search space. 2021.
- Wei Peng, Xiaopeng Hong, Haoyu Chen, and Guoying Zhao. Learning graph convolutional network for skeleton-based human action recognition by neural searching. In *AAAI*, 2020.
- Huan Zhao, Quanming Yao, and Weiwei Tu. Search to aggregate neighborhood for graph neural network. In *ICDE*, 2021.
- Yaoman Li and Irwin King. Autograph: Automated graph neural network. In *ICONIP*, 2020.
- Shaofei Cai, Liang Li, Jincan Deng, Beichen Zhang, Zheng-Jun Zha, Li Su, and Qingming Huang. Rethinking graph neural network search from message-passing. *CVPR*, 2021.
- Guosheng Feng, Chunnan Wang, and Hongzhi Wang. Search for deep graph neural networks. *arXiv preprint arXiv:2109.10047*, 2021.

-
- Kwei-Herng Lai, Daochen Zha, Kaixiong Zhou, and Xia Hu. Policy-gnn: Aggregation optimization for graph neural networks. In *KDD*, 2020.
- Yuhui Ding, Quanming Yao, Huan Zhao, and Tong Zhang. Diffmg: Differentiable meta graph search for heterogeneous graph neural networks. In *KDD*, 2021.
- Lanning Wei, Huan Zhao, Quanming Yao, and Zhiqiang He. Pooling architecture search for graph classification. In *CIKM*, 2021.
- Lanning Wei, Huan Zhao, and Zhiqiang He. Designing the topology of graph neural networks: A novel feature fusion perspective. 2022.
- Jian Du, Shanghang Zhang, Guanhang Wu, José M. F. Moura, and Soumya Kar. Topology adaptive graph convolutional networks. 2017.