GRIFFITH UNIVERSITY

arc training centre for **information resilience**

Information Resilience PhD School 2024

Automated Graph Machine Learning Operations (MLOps) Workflow:

A Data-Centric Perspective

Dr. Xin Zheng

Lecturer

School of ICT, Griffith University

Oct, 2024

Contents

- **Introduction & Overview**
  - *Automated Graph MLOps in Data-Centric AI*
- **Graph Data-Centric Exploitation**
  - *Graph Type*
  - *Graph Scale*
- **Graph Data-Centric Model Deployment**
  - *GNN Model Evaluation*
- **Future Opportunities**

Contents

- **Introduction & Overview**
  - *Automated Graph MLOps in Data-Centric AI*
- **Graph Data-Centric Exploitation**
  - *Graph Scale*
  - *Graph Type*
- **Graph Data-Centric Model Deployment**
  - *GNN Model Evaluation*
- **Future Opportunities**

GRIFFITH UNIVERSITY

AI System

AI system = Code + Data

(model/algorithm)

What is Data-Centric AI?

“ Data-centric AI (DCAI) is the discipline of systematically engineering the data used to build an AI system.” – –Andrew Ng

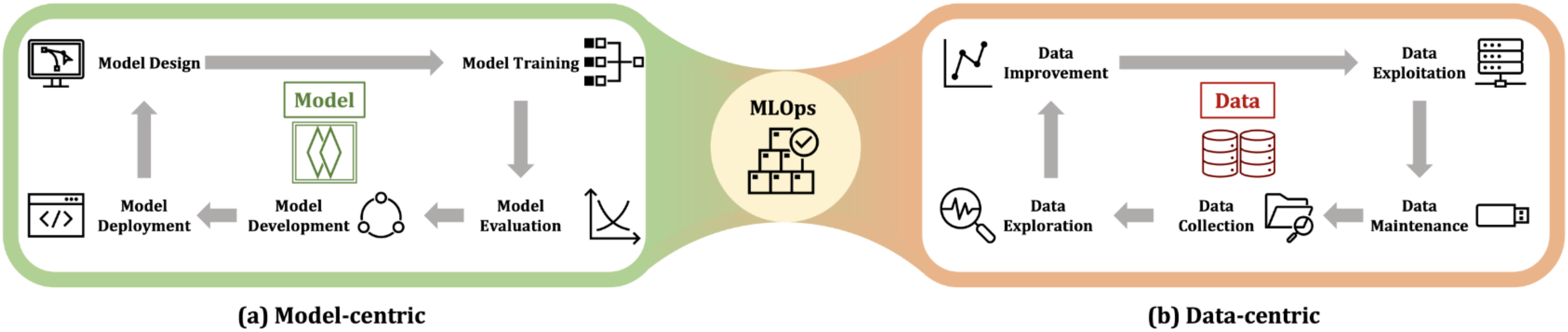
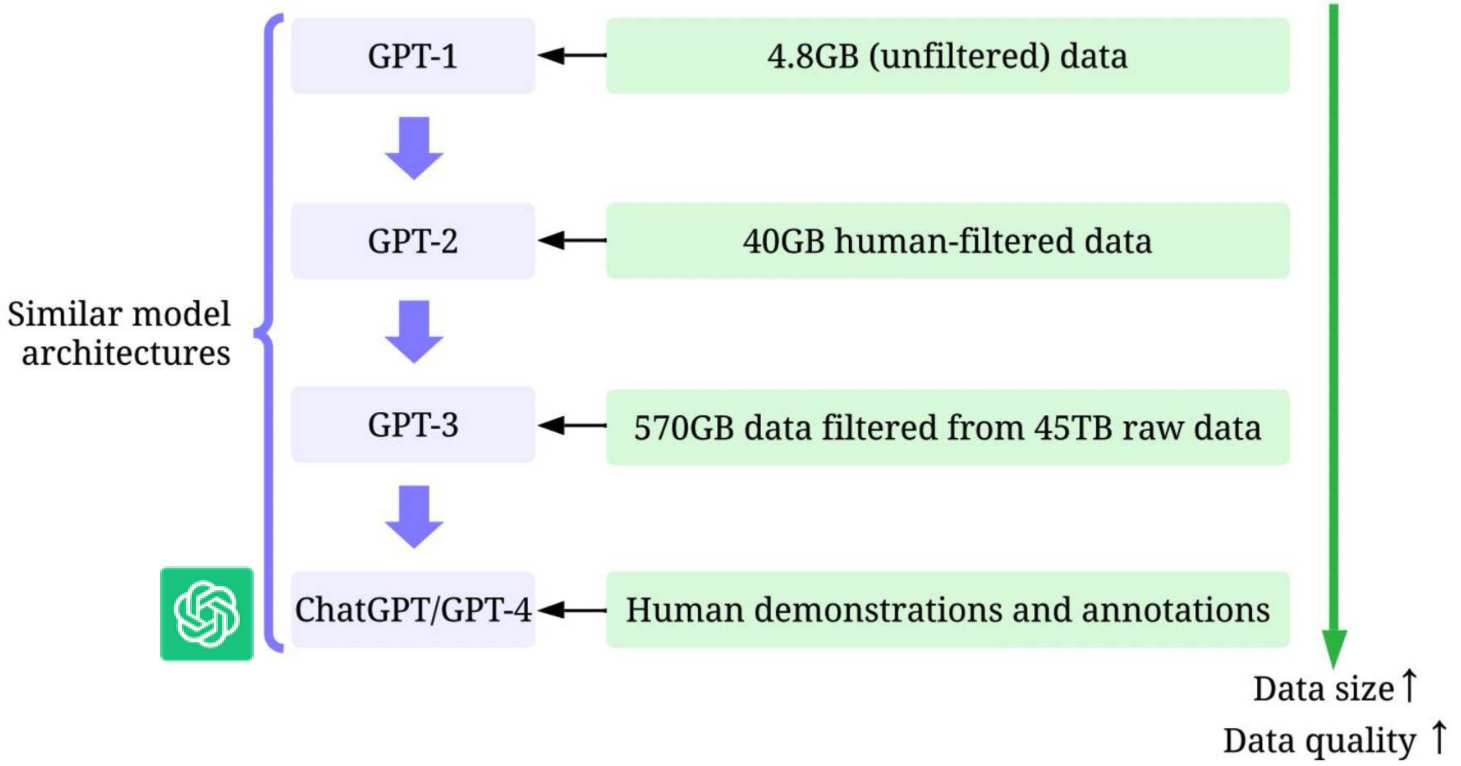


Fig. 1. General comparison between (a) model-centric AI and (b) data-centric AI.

Why Data-Centric AI matters

When model design becomes mature, the significance of both the size and quality of the data increases.



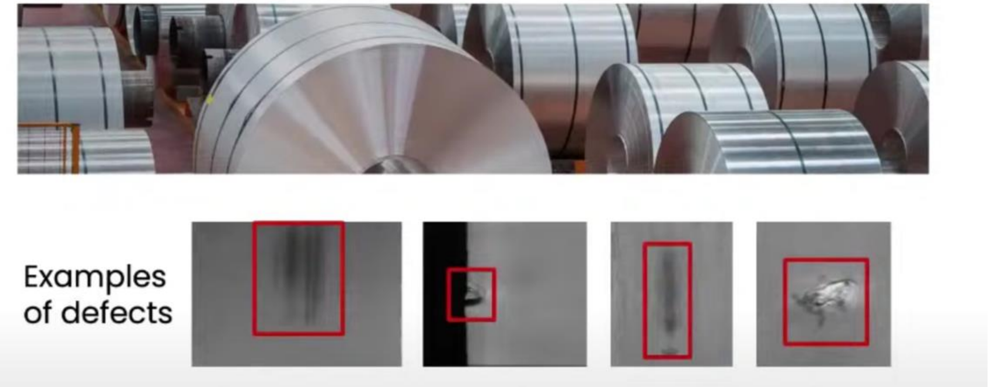
❖ **Core idea:** Engineering data to enable great **"availability and quality"** for serving model-related ML tasks.

[1] Zha, Daochen, et al. Data-centric Artificial Intelligence: A Survey. arXiv, 2023.

Why Data-Centric AI matters

An example:

Inspecting steel sheets for defects

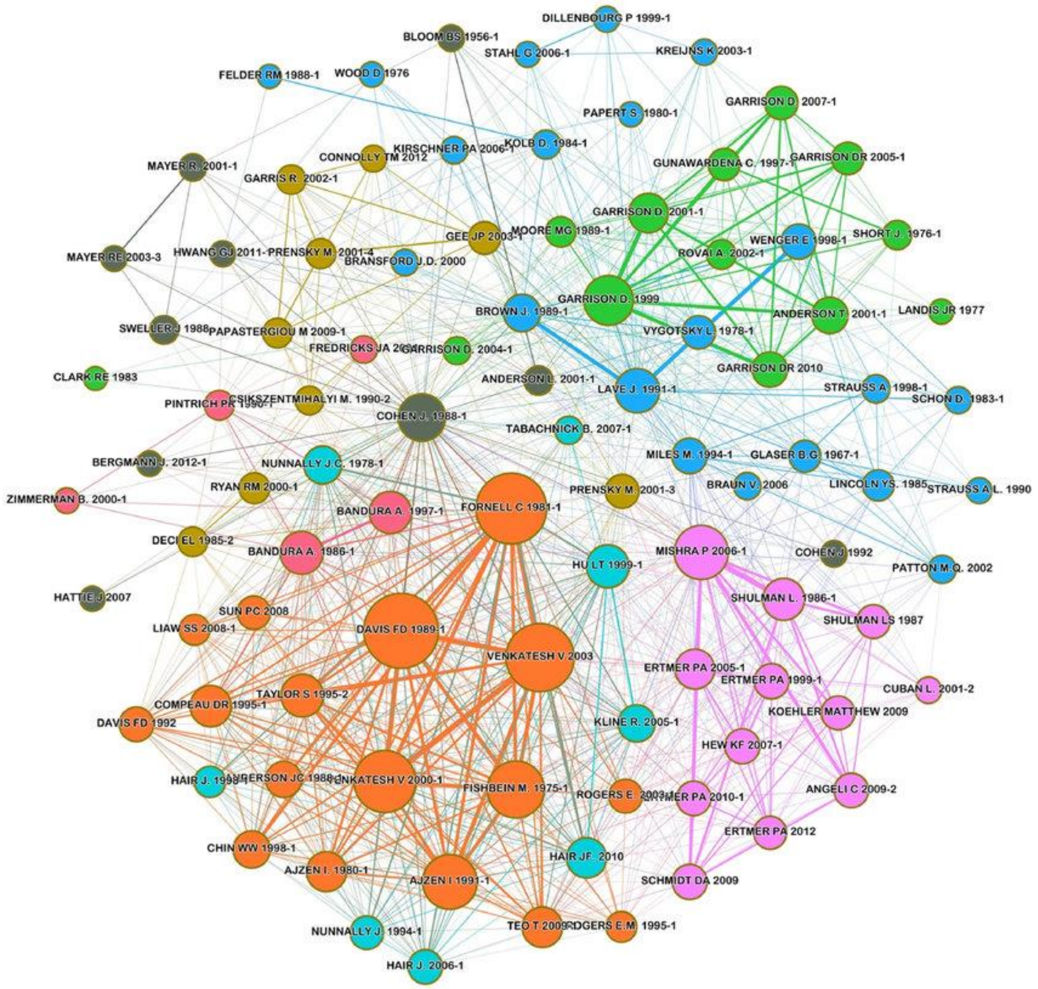


	Steel defect detection	Solar panel	Surface inspection
Baseline	76.2%	75.68%	85.05%
Model-centric	+0% (76.2%)	+0.04% (75.72%)	+0.00% (85.05%)
Data-centric	+16.9% (93.1%)	+3.06% (78.74%)	+0.4% (85.45%)

Data-centric improves more than model-centric!

[1] A Chat with Andrew on MLOps: From Model-centric to Data-centric AI: https://www.youtube.com/watch?v=06-AZXmwHjo

Graphs: A typical & vital instantiation in DCAI



Example: Citation Network [1]

A Graph has nodes/vertices and edges:

- Nodes/vertices

→ a paper in the citation network

- Edges

→ connections between papers

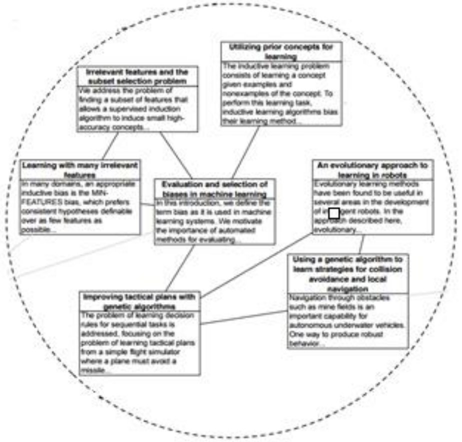
Graphs have the ability of:

- Representing complex structural relationships among massive diverse entities in the real world

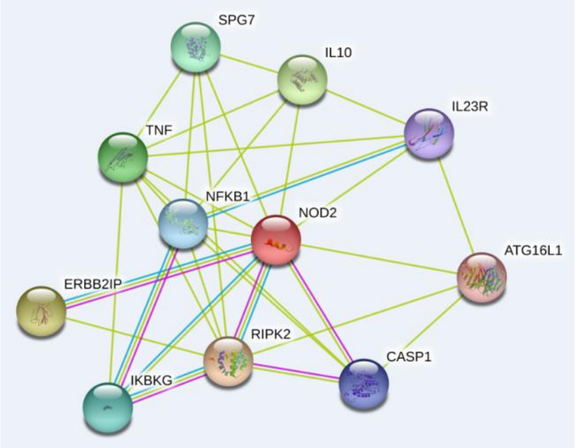
[1] Valtonen, Teemu, et al. "The nature and building blocks of educational technology research." Computers in Human Behavior 128 (2022): 107123.

Graphs in real-world applications

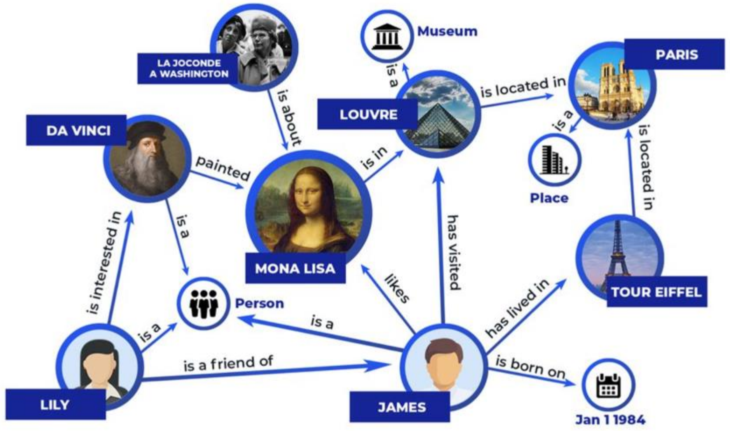
Social Networks



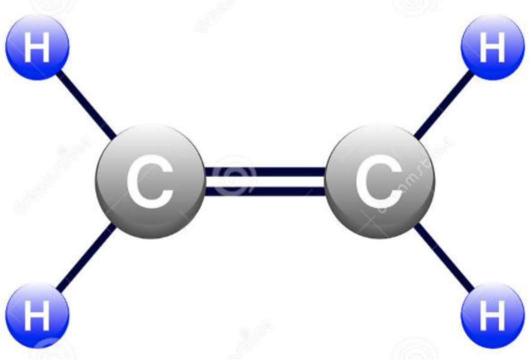
Bibliography Networks



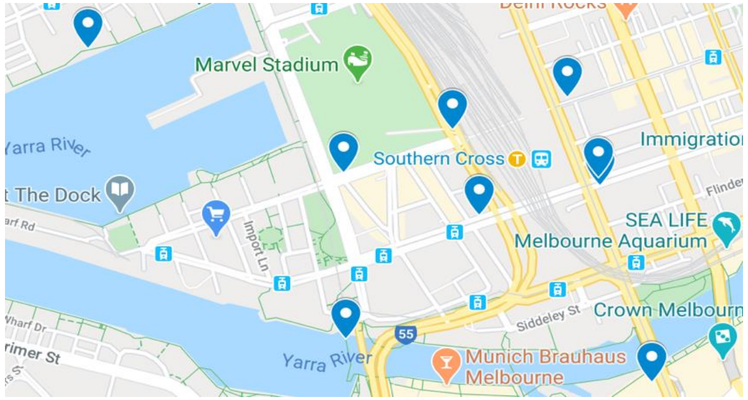
Protein Interaction Networks



Knowledge Graphs



Chemical Compounds



Traffic Networks

Graph Neural Networks (GNNs)

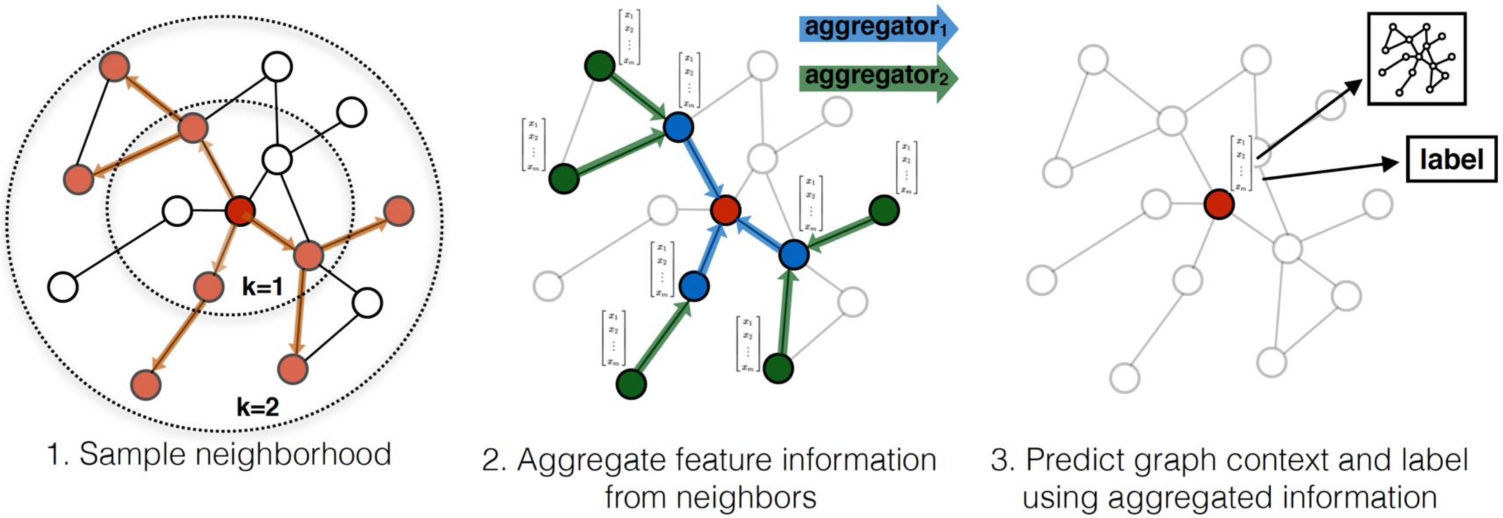


Figure 1: GraphSage pioneered powerful aggregation techniques for message passing in GNNs.

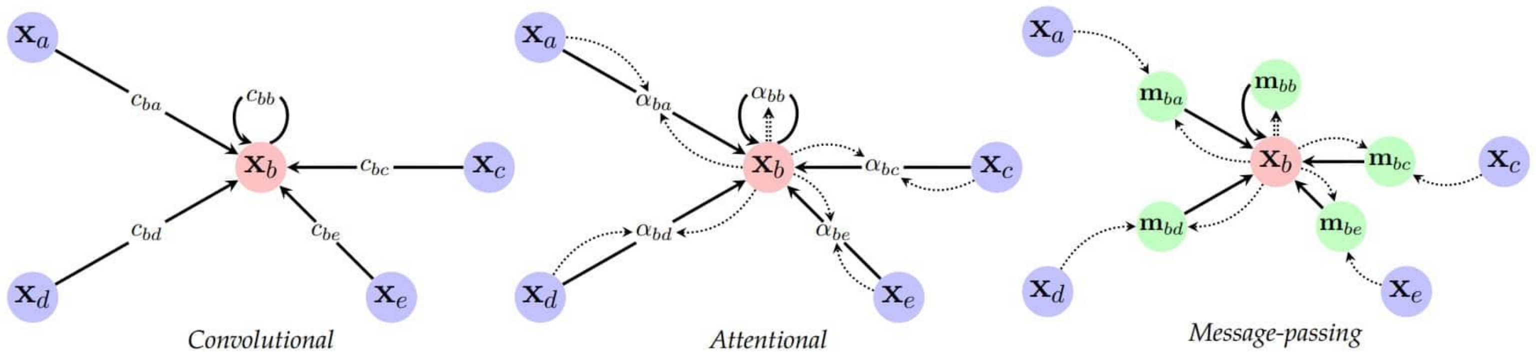
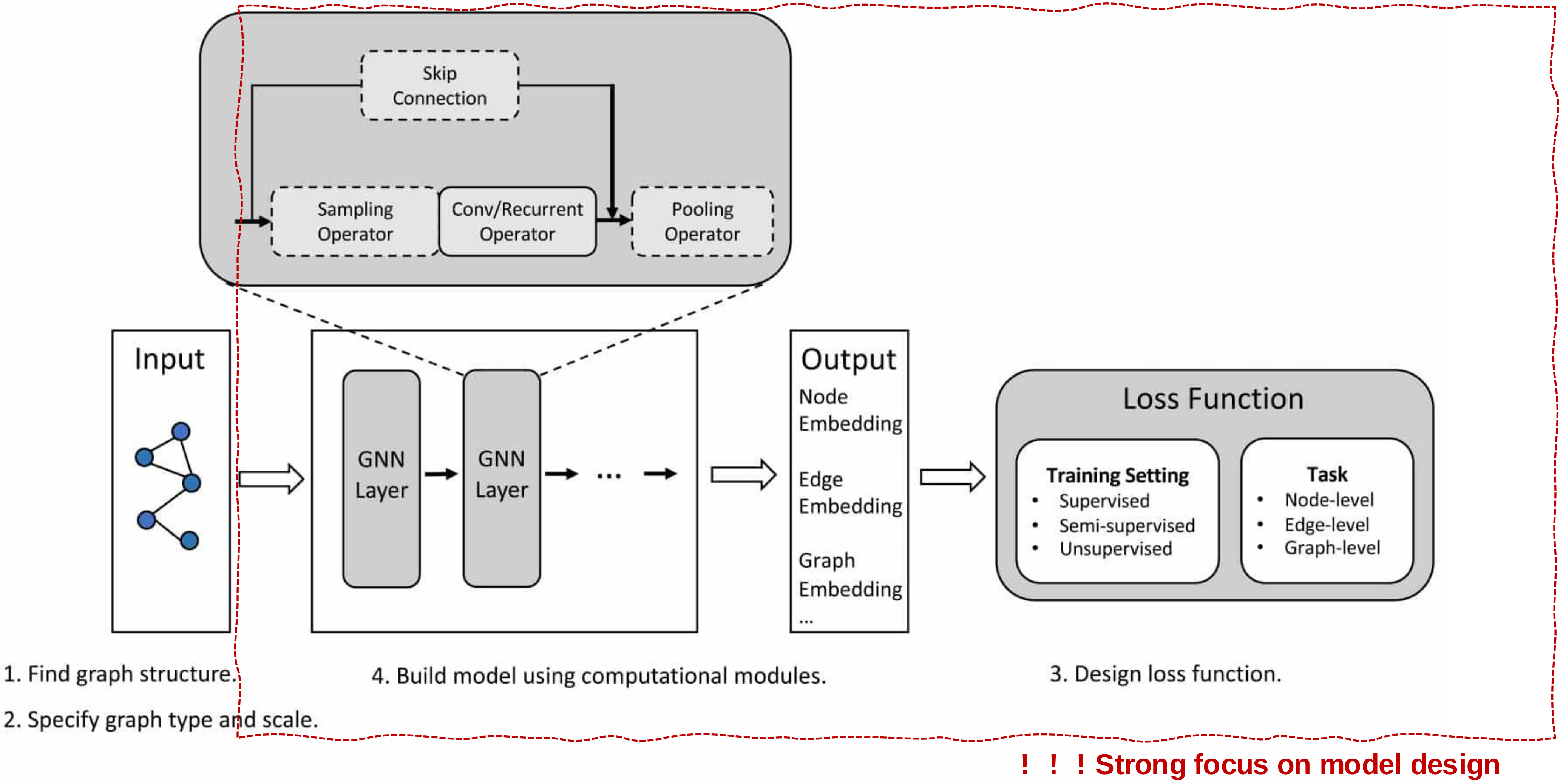


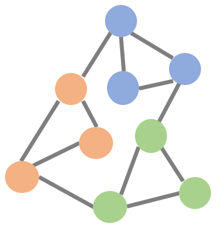
Figure 2: Example dataflows in three types of GNNs.

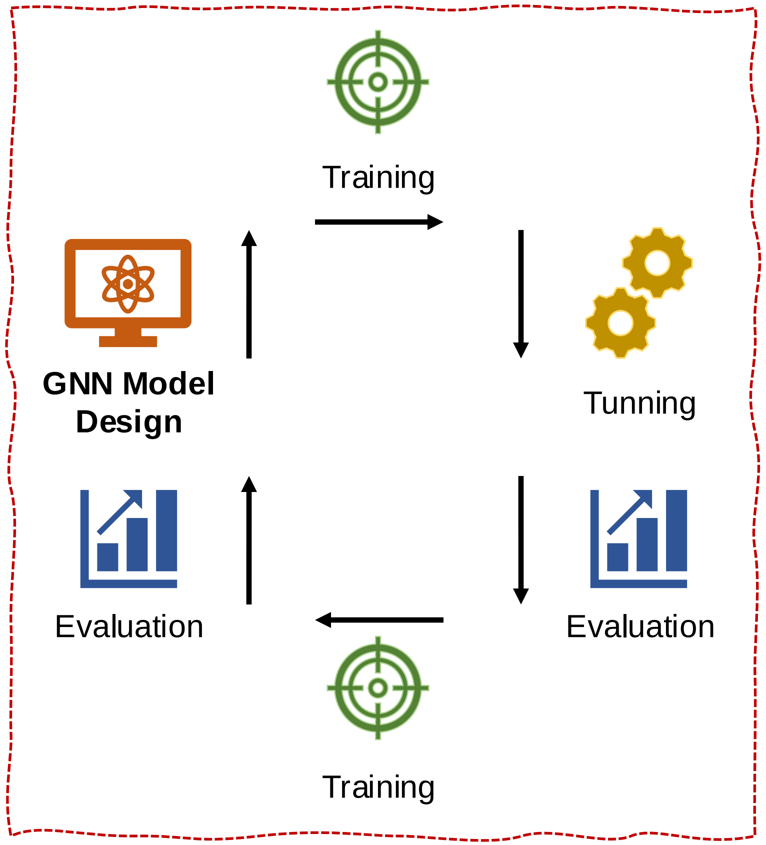
Graph Neural Networks (GNNs)



The general design pipeline for a GNN model.

Look at a Bigger Picture…





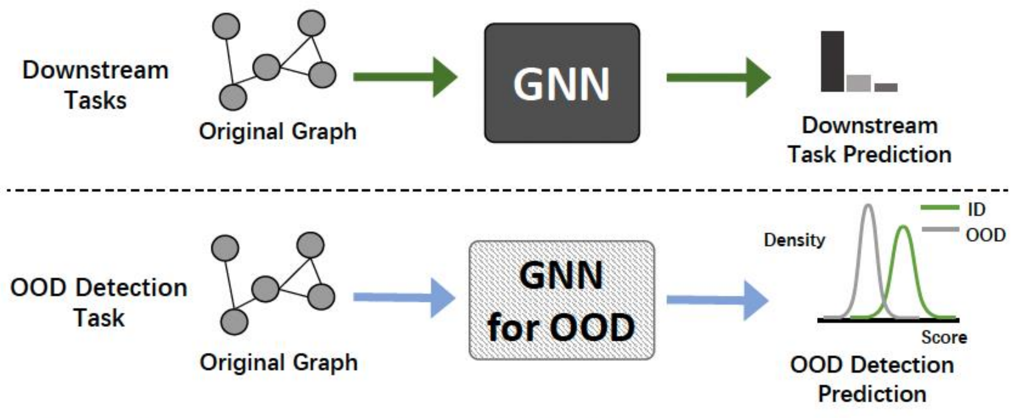
× 【Data-level】The important role of graph-structured data is overlooked (e.g., scale and types)

× **【Model-level】Human manually designed GNNs cannot well adapt specific graph data and tasks**

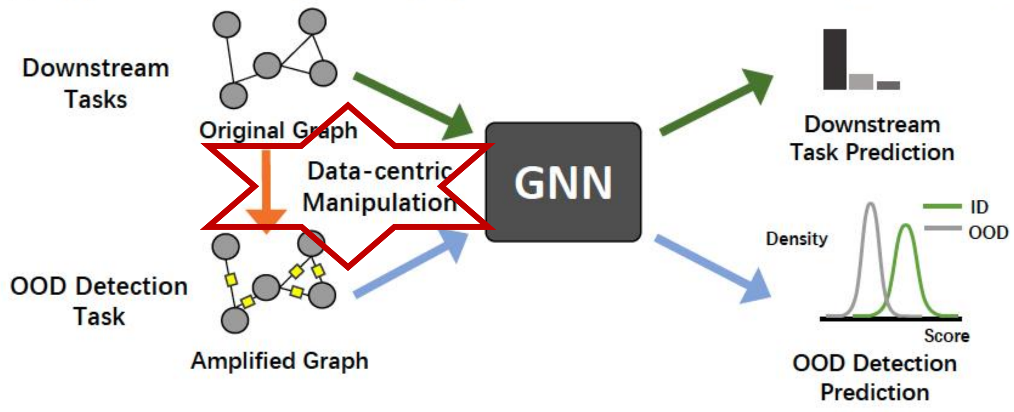
× **【Deployment-level】Difficult to evaluate well-developed GNNs on real-world test graph data**

Why Data-Centric Graph ML matters?

❖ **Taking graph OOD detection as example:**



(a) Typical retraining-based graph OOD detection methods



(b) Our proposed data-centric framework for graph OOD detection.

ID	OOD	Metric	GCL_S	GCL_S+	Improv.
ENZYMES	PROTEIN	AUC ↑	62.97	**73.76**	+17.14%
		AUPR ↑	62.47	**75.27**	+20.49%
		FPR95 ↓	93.33	**88.33**	-5.36%
IMDBM	IMDBB	AUC ↑	80.52	**83.84**	+4.12%
		AUPR ↑	74.43	**80.16**	+7.70%
		FPR95 ↓	38.67	**38.33**	-0.88%
BZR	COX2	AUC ↑	75.00	**97.31**	+29.75%
		AUPR ↑	62.41	**97.17**	+55.70%
		FPR95 ↓	47.50	**15.00**	-68.42%

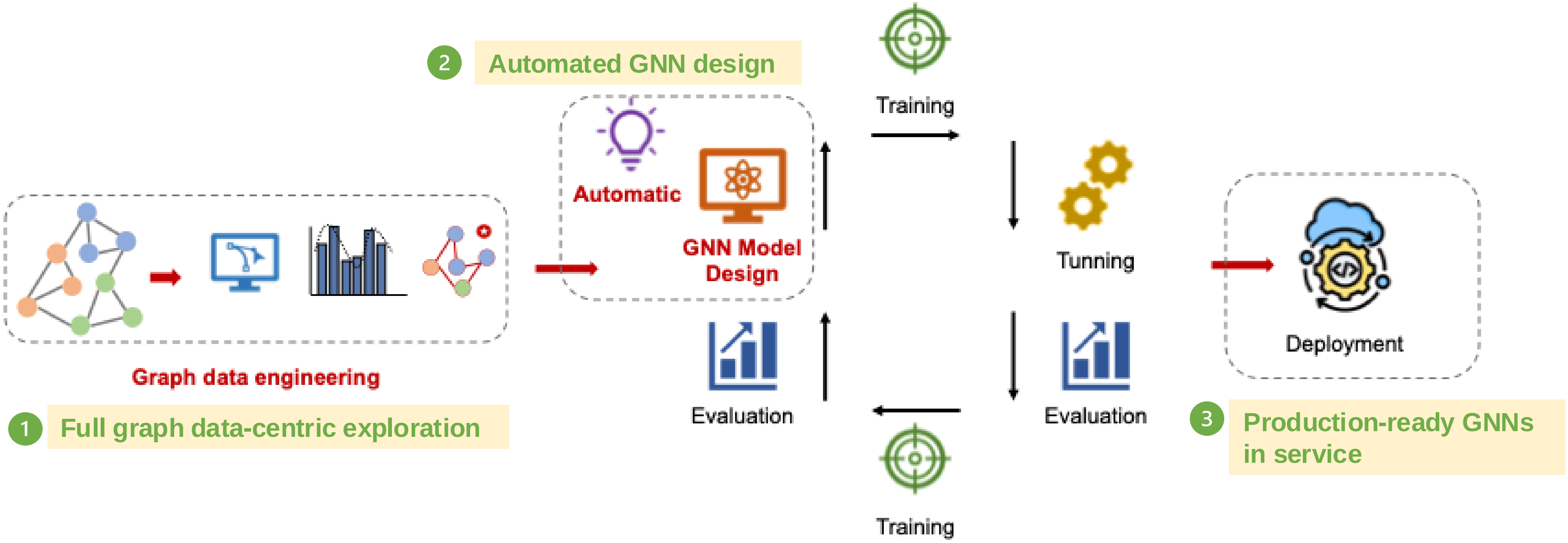
Model-centric GML method

✩ **Data-centric GML method and improvements**

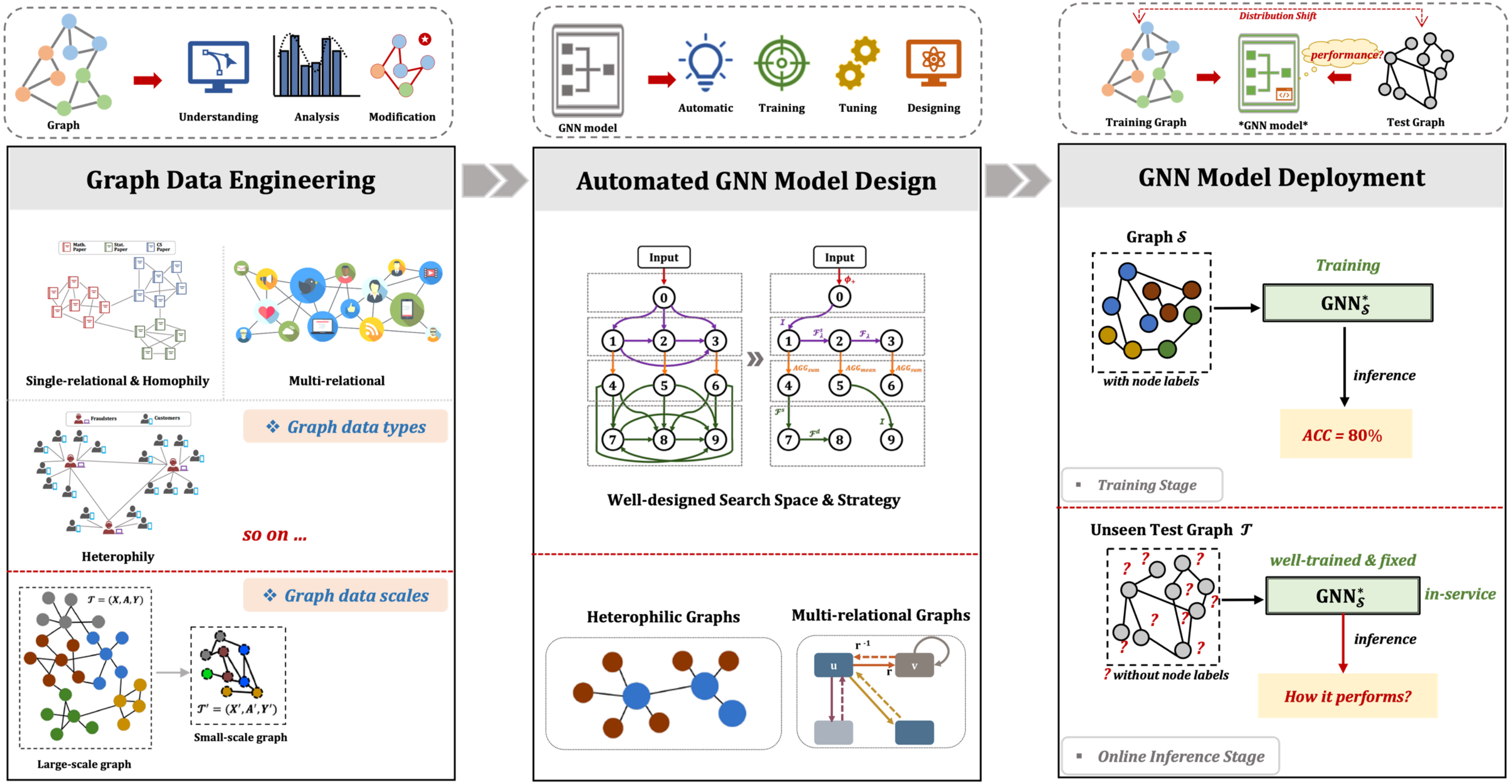
Guo, Y., Yang, C., Chen, Y., Liu, J., Shi, C., & Du, J. (2023). A Data-centric Framework to Endow Graph Neural Networks with Out-Of-Distribution Detection Ability.

What We Need is…

systematically consider the entire pipeline of building **"production-ready GNN models" from industrial perspective** in real-world application scenarios.



Automated Graph MLOps



GRIFFITH UNIVERSITY

Automated Graph MLOps--Data-Centric Focus

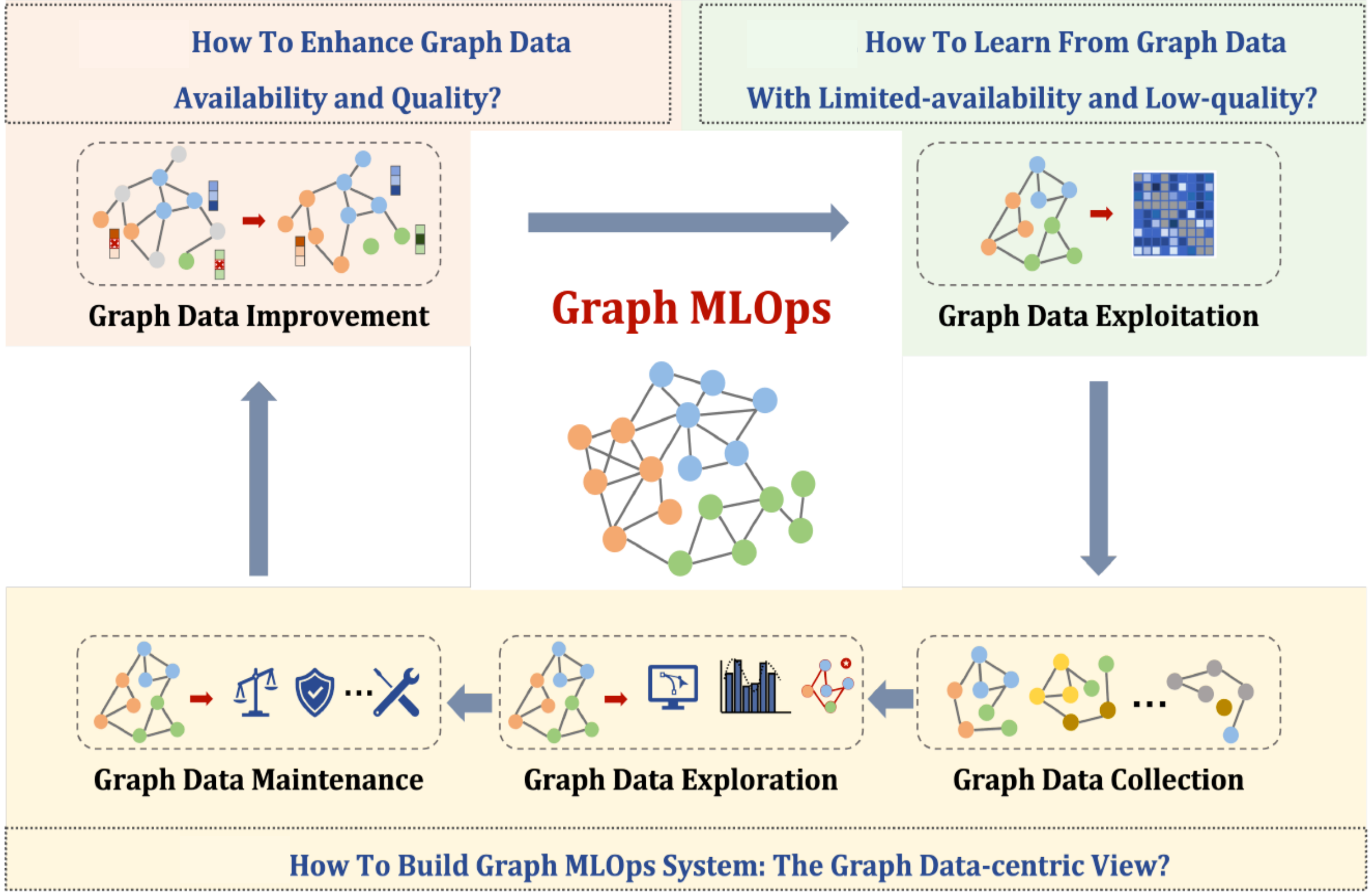
Automated Graph Machine Learning Operations (MLOps) Workflow

1. **Graph Data Engineering**
   - *(a) Graph data types: complexity and diversity*
   - *(b) Graph data scales: large graph data quantity*
2. **Automated GNN Model Design**
   - *Using graph neural architecture search (NAS) to automatically develop data-driven & task-driven GNNs*
3. **GNN Model Deployment**
   - *Understanding well-developed GNNs' performance when serving real-world unseen test graphs*

Automated Graph MLOps--Data-Centric Focus

Data-centric graph machine learning (DC-GML) aims to:

- Process, analyze, and understand graph data in entire lifecycle
- Enhancing the quality
- Uncovering the insights
- Developing comprehensive representations
- Working collaboratively with graph ML models under graph MLOps



Data-centric Graph ML Review & Outlook

- Survey paper: Towards Data-centric Graph Machine Learning: Review and Outlook
- Github collection: https://github.com/Data-Centric-GraphML/awesome-papers

Contents

- **Introduction & Overview**
  - *Automated Graph MLOps in Data-Centric AI*
- **Graph Data-Centric Exploitation**
  - ***Graph Scale***
  - *Graph Type*
- **Graph Data-Centric Model Deployment**
  - *GNN Model Evaluation*
- **Future Opportunities**

GRIFFITH UNIVERSITY

Graph Data Scale Issue

In real-world application,

a) Graph data scale can be very large;

b) Modelling large-scale graphs hinders GNN development with heavy costs.

Table 1: Model serving space

Datasets	Model size	Training graph size	Training feature size	Total serving size
Arxiv	1.4MB	5.9MB	46.5MB	53.8MB
Reddit	7.6MB	86.0MB	370.7MB	464.3MB
Product	4.8MB	87.2MB	78.6MB	170.6MB
Amazon2M	3.0MB	485.4MB	684.0MB	1.17GB

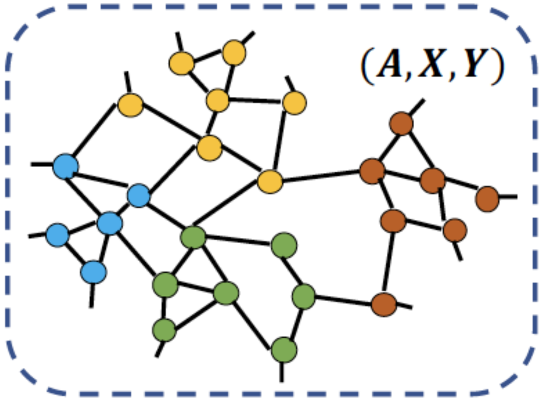
!!! Model size << Graph data size

Graph Data Scale Issue

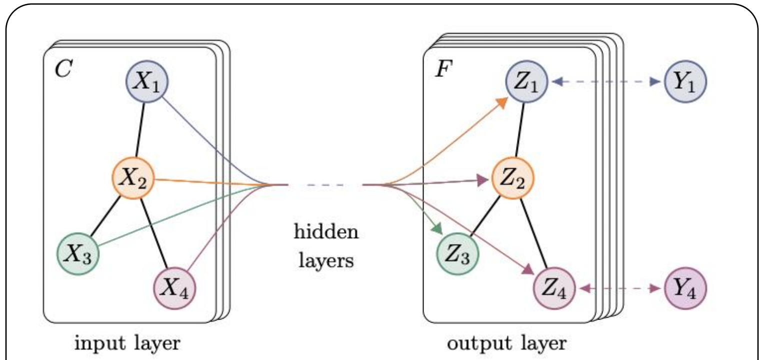
Training optimal GNN models on large-scale graphs would:

a) Require repeat training & finetuning for optimality

b) Heavy costs on: graph data storage, computation, and memory



153,932 training nodes

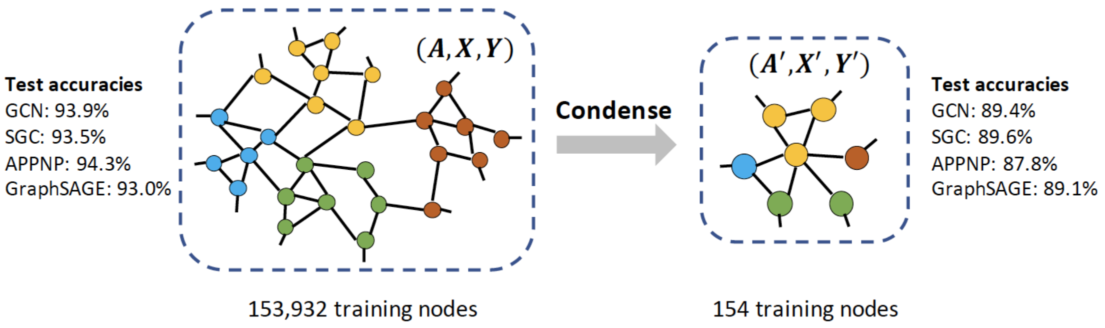


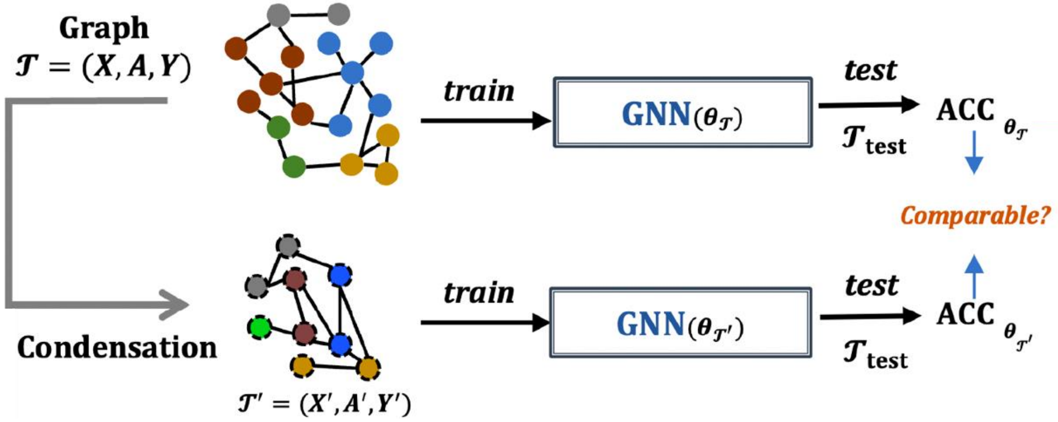
Graph Convolutional Networks (GCN)

Solution: Graph Condensation

aim to reduce the size of a large-scale graph by synthesizing a small-scale condensed graph

→ → the small-scale condensed graph achieves comparable test performance as the large-scale graph when training the same GNN model.





Zheng, X., Zhang, M., Chen, C., Nguyen, Q. V. H., Zhu, X., & Pan, S. (2023). Structure-free Graph Condensation: From Large-scale Graphs to Condensed Graph-free Data. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Benefits of Graph Condensation

Using condensed graph as substitution to facilitate GNN training:

- Alleviated graph data storage/computation/memory costs

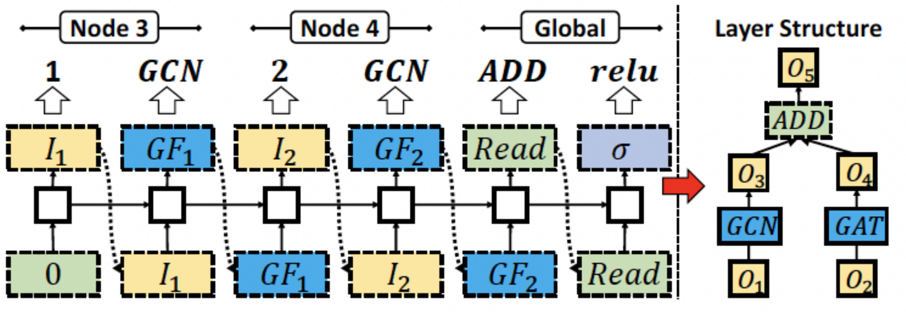
❖ Practical applications of GC?

- Graph Neural Architecture Search (GraphNAS)

  By searching on a small-scale condensed graph, accelerating new GNN architecture development in GraphNAS

  ...

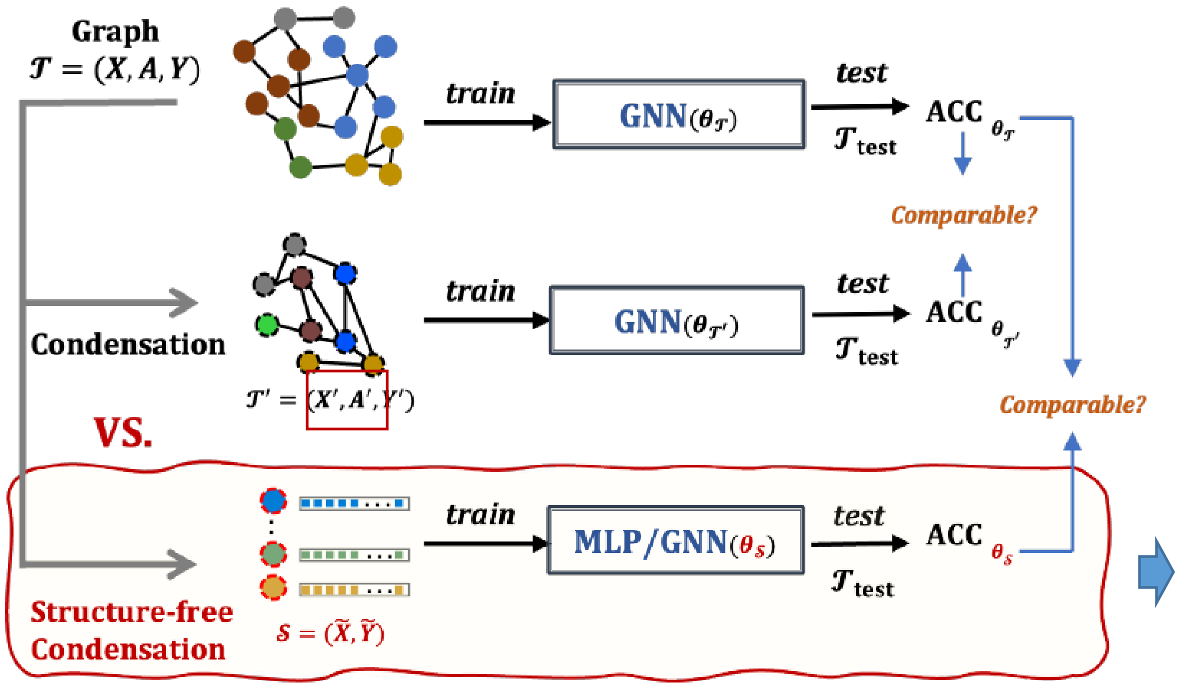
- Privacy Protection
- Adversarial Robustness



Example: Graph NAS

Zheng, X., Zhang, M., Chen, C., Nguyen, Q. V. H., Zhu, X., & Pan, S. (2023). Structure-free Graph Condensation: From Large-scale Graphs to Condensed Graph-free Data. Advances in Neural Information Processing Systems (NeurIPS), 2023.

The Proposed: Structure-free Graph Condensation



- **Existing works :**

$$\mathcal{T} = (\mathbf{X}, \mathbf{A}, \mathbf{Y}) \rightarrow \mathcal{T}' = (\mathbf{X}', \mathbf{A}', \mathbf{Y}'), \quad \text{GC.}$$

- **Our SFGC:**

$$\boxed{\mathcal{T} = (\mathbf{X}, \mathbf{A}, \mathbf{Y}) \rightarrow \mathcal{S} = (\widetilde{\mathbf{X}}, \mathbf{I}, \widetilde{\mathbf{Y}}) = \mathcal{S} = (\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}), \quad \text{SFGC.}}$$

- **Our Solution:**
  - ✓ **Structure-free paradigm** → • Only synthesizes a small scaled node set to train a GNN/MLP
  - ✓ **Long-range parameter matching schema** → • Implicitly encodes topology structure into node attributes

Zheng, X., Zhang, M., Chen, C., Nguyen, Q. V. H., Zhu, X., & Pan, S. (2023). Structure-free Graph Condensation: From Large-scale Graphs to Condensed Graph-free Data. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Framework: Structure-free Graph Condensation

Condensing large-scale graph into only node set without structures!

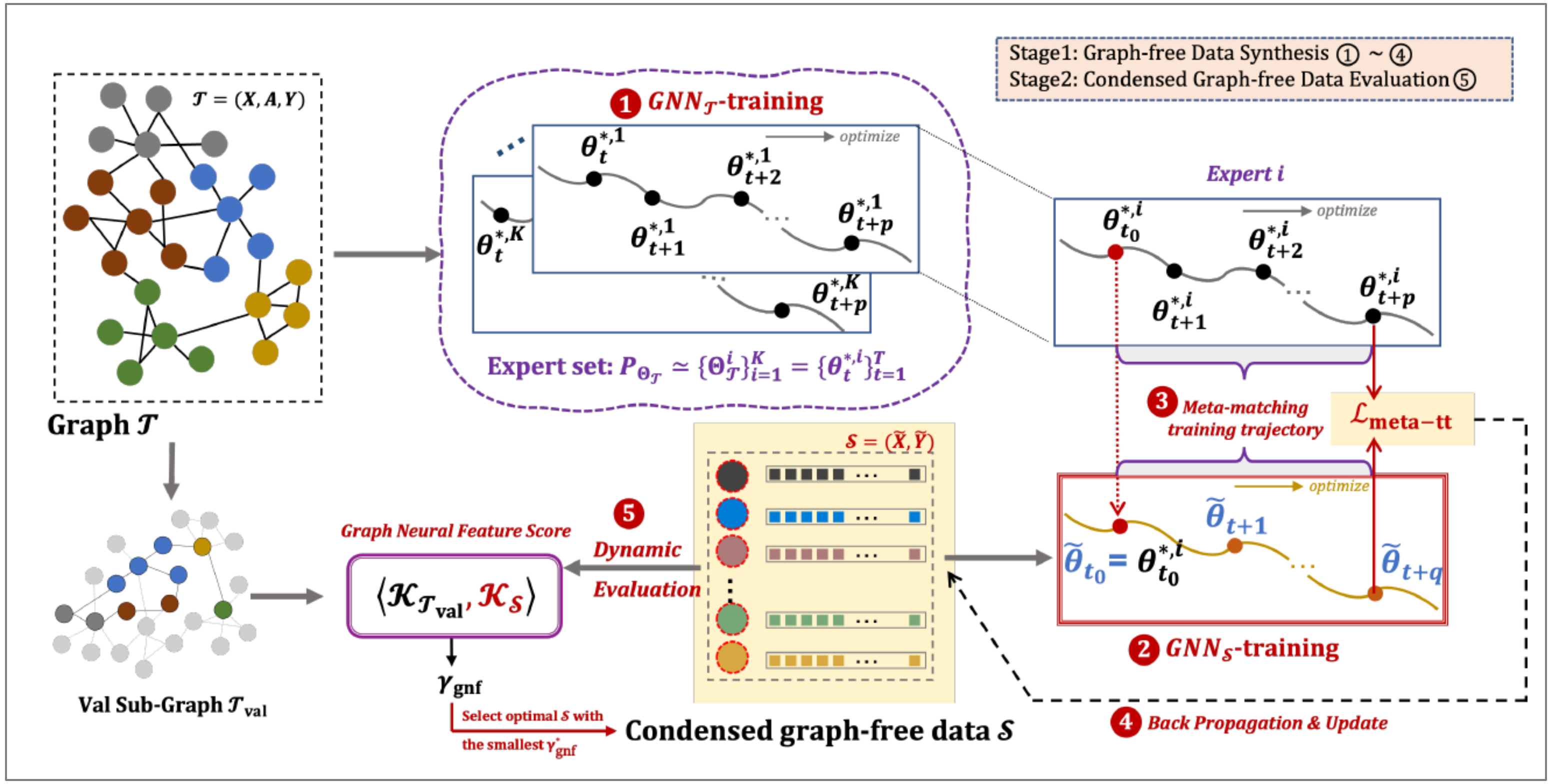


Figure 1. Overall pipeline of the proposed Structure-Free Graph Condensation (SFGC) framework

Input: large-scale T, GNN(T)

Output: small-scale condensed S

- **S1: train expert GNN on large-scale T**
- **S2-3: <u>long-term meta training trajectory matching</u> with condensed S**
- **S4: update S**
- **S5: dynamically evaluates S with a <u>GNTK-based score</u>**

Zheng, X., Zhang, M., Chen, C., Nguyen, Q. V. H., Zhu, X., & Pan, S. (2023). Structure-free Graph Condensation: From Large-scale Graphs to Condensed Graph-free Data. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Results: Structure-free Graph Condensation

Table 1: Node classification performance (ACC%±std) comparison between condensation methods and other graph size reduction methods with different condensation ratios. (Best results are in bold, and the second-bests are underlined.)

Datasets	Ratio (r)	*Other Graph Size Reduction Baselines*				*Condensation Methods*				Whole Dataset
		Coarsening [13]	Random [31]	Herding [31]	K-Center [28]	DC-Graph [42]	GCOND-X [18]	GCOND [18]	SFGC (ours)	
Citeseer	0.9%	52.2±0.4	54.4±4.4	57.1±1.5	52.4±2.8	66.8±1.5	71.4±0.8	70.5±1.2	**71.4**±0.5	71.7±0.1
	1.8%	59.0±0.5	64.2±1.7	66.7±1.0	64.3±1.0	66.9±0.9	69.8±1.1	70.6±0.9	**72.4**±0.4	
	3.6%	65.3±0.5	69.1±0.1	69.0±0.1	69.1±0.1	66.3±1.5	69.4±1.4	69.8±1.4	**70.6**±0.7	
Cora	1.3%	31.2±0.2	63.6±3.7	67.0±1.3	64.0±2.3	67.3±1.9	75.9±1.2	79.8±1.3	**80.1**±0.4	81.2±0.2
	2.6%	65.2±0.6	72.8±1.1	73.4±1.0	73.2±1.2	67.6±3.5	75.7±0.9	80.1±0.6	**81.7**±0.5	
	5.2%	70.6±0.1	76.8±0.1	76.8±0.1	76.7±0.1	67.7±2.2	76.0±0.9	79.3±0.3	**81.6**±0.8	
Ogbn-arxiv	0.05%	35.4±0.3	47.1±3.9	52.4±1.8	47.2±3.0	58.6±0.4	61.3±0.5	59.2±1.1	**65.5**±0.7	71.4±0.1
	0.25%	43.5±0.2	57.3±1.1	58.6±1.2	56.8±0.8	59.9±0.3	64.2±0.4	63.2±0.3	**66.1**±0.4	
	0.5%	50.4±0.1	60.0±0.9	60.4±0.8	60.3±0.4	59.5±0.3	63.1±0.5	64.0±0.4	**66.8**±0.4	
Flickr	0.1%	41.9±0.2	41.8±2.0	42.5±1.8	42.0±0.7	46.3±0.2	45.9±0.1	46.5±0.4	**46.6**±0.2	47.2±0.1
	0.5%	44.5±0.1	44.0±0.4	43.9±0.9	43.2±0.1	45.9±0.1	45.0±0.2	**47.1**±0.1	47.0±0.1	
	1%	44.6±0.1	44.6±0.2	44.4±0.6	44.1±0.4	45.8±0.1	45.0±0.1	**47.1**±0.1	**47.1**±0.1	
Reddit	0.05%	40.9±0.5	46.1±4.4	53.1±2.5	46.6±2.3	88.2±0.2	88.4±0.4	88.0±1.8	**89.7**±0.2	93.9±0.0
	0.1%	42.8±0.8	58.0±2.2	62.7±1.0	53.0±3.3	89.5±0.1	89.3±0.1	89.6±0.7	**90.0**±0.3	
	0.2%	47.4±0.9	66.3±1.9	71.0±1.6	58.5±2.1	**90.5**±1.2	88.8±0.4	90.1±0.5	90.3±0.3	

- Generally, SFGC achieves the best performance on the node classification task with 13 of 15 cases (five datasets and three condensation ratios for each of them), illustrating the high quality and expressiveness of the condensed graph-free data synthesized by our SFGC

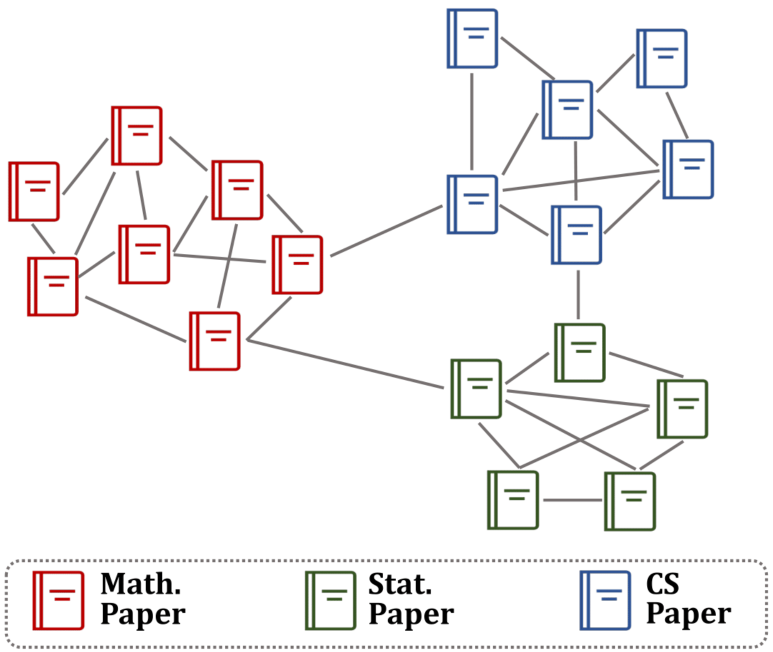
Zheng, X., Zhang, M., Chen, C., Nguyen, Q. V. H., Zhu, X., & Pan, S. (2023). Structure-free Graph Condensation: From Large-scale Graphs to Condensed Graph-free Data. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Contents

- **Introduction & Overview**
  - *Automated Graph MLOps in Data-Centric AI*
- **Graph Data-Centric Exploitation**
  - *Graph Scale*
  - ***Graph Type***
- **Graph Data-Centric Model Deployment**
  - *GNN Model Evaluation*
- **Future Opportunities**

GRIFFITH UNIVERSITY

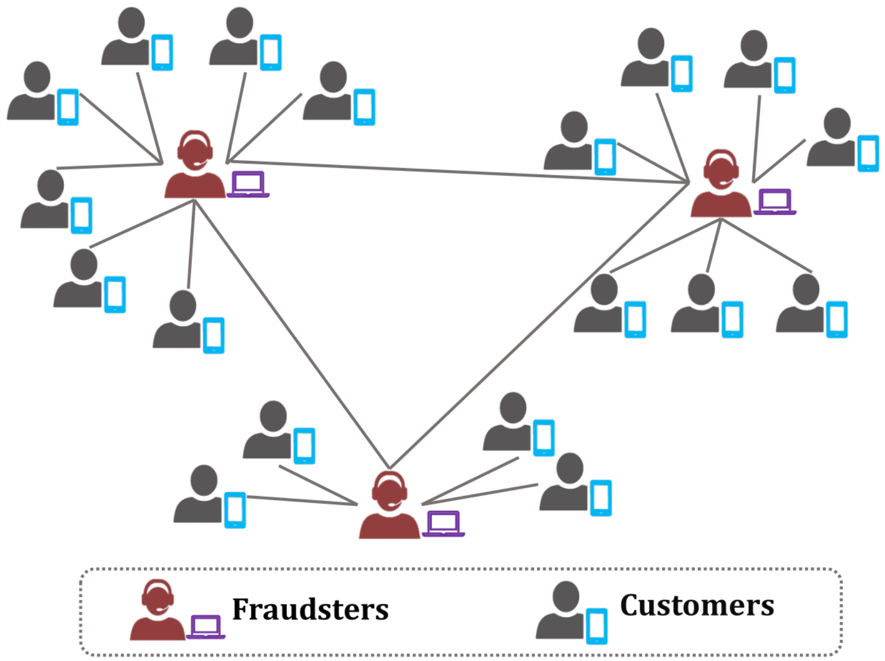
Homophilic vs. Heterophilic Graphs



(a) Homophilic Graph

Nodes with similar features or same class labels are linked together.

- *E.g., in citation networks, a study usually cites reference papers from the same research.*



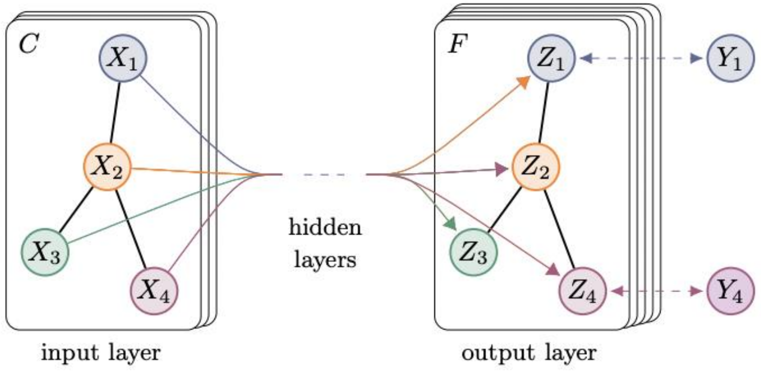
(b) Heterophilic Graph

Linked nodes have dissimilar features and different class labels.

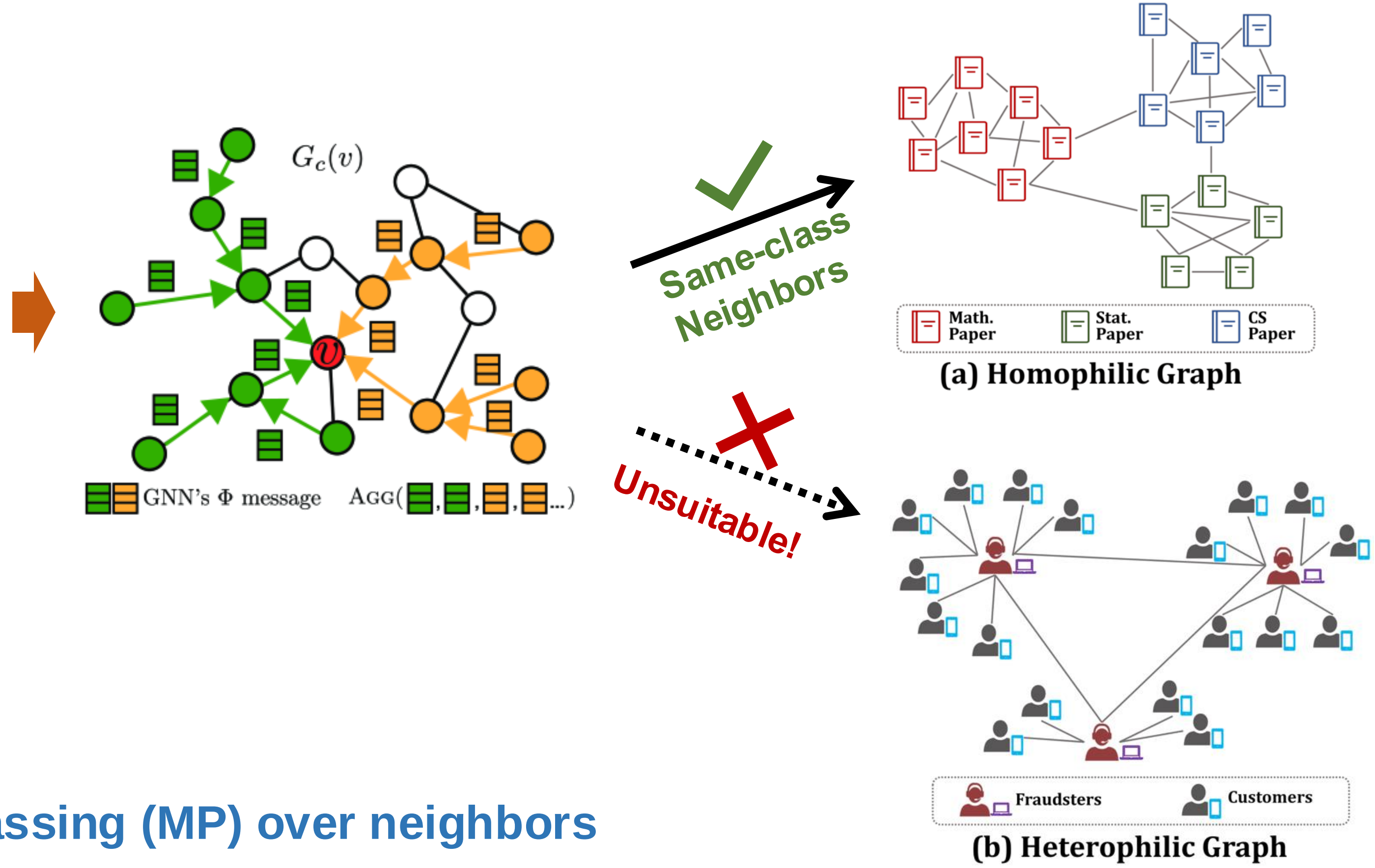
- *E.g., in online transaction networks, fraudsters are more likely to build connections with customers instead of other fraudsters.*

Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

Homophilic GNNs Unsuitable for Heterophily



Graph Convolutional Networks (GCN)



❖ Core idea of GNNs： Message Passing (MP) over neighbors

But on heterophilic graphs, neighbors might not in the same class !

Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

Question

Is that possible to design

❶ **heterophily-friendly GNNs &**

❷ **automatically &**

❸ **driven by heterophilic graphs?**

Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

Solution: Graph Neural Architecture Search (NAS)

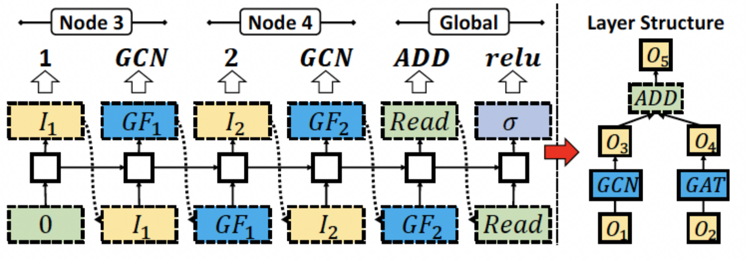
❖ **Human-designed—Non-automated:**

- ✕ Too much human-effort cost
- ✕ Model performance heavily relies on expertise

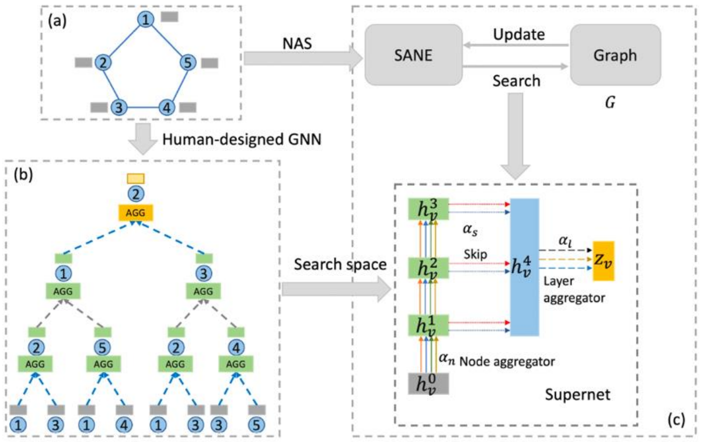
A promising solution…

❖ **Graph Neural Architecture Search (NAS) – Automated:**

- ✓ Relieving human efforts
- ✓ Powerful GNN models driven by data and tasks



Graph Neural Architecture Search (GNAS)



Search to Aggregate Neighborhood (SANE)

Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

Solution: Graph Neural Architecture Search (NAS)

Despite promising performance, the mainstream graph NAS is limited by

➢ **Graph-structured data level:**

- Simple-relational graph-structure data → • Real-world graphs are complex and diverse
- Homophily assumption of graphs → • **Heterophily?**

➢ **NAS algorithm level:**

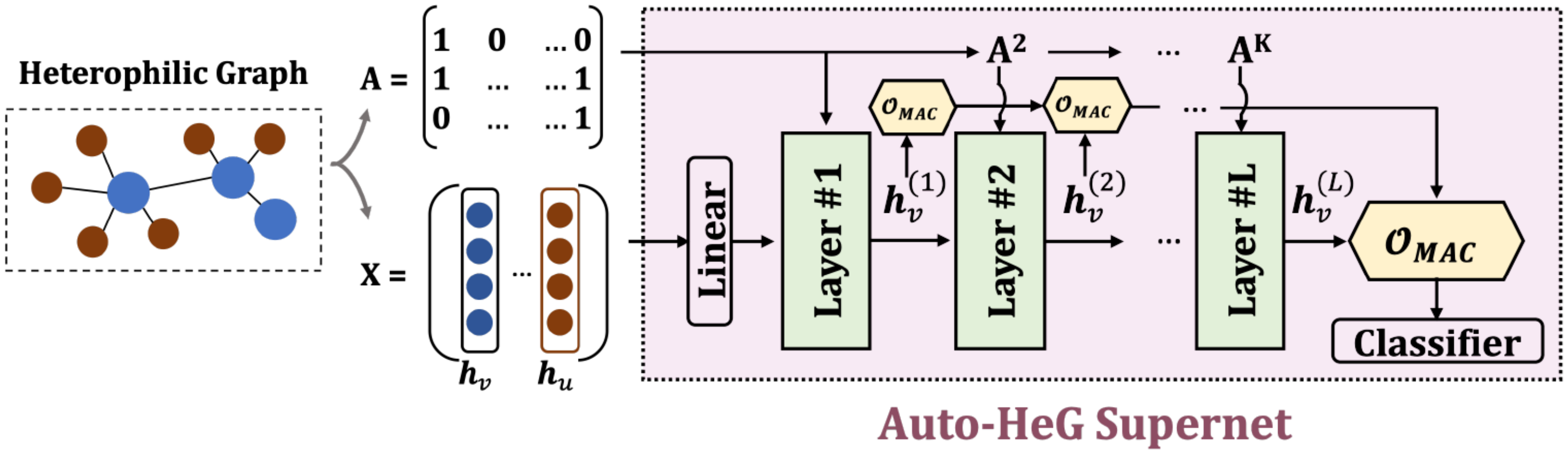
- Coarse-grained GNN search space → • Simple ensemble learning of existing GNNs
- Simple search strategy → • Require specific search strategy

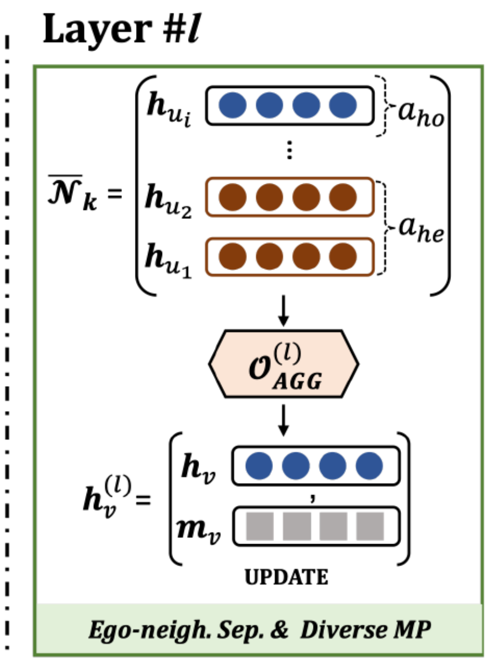
Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

Our Proposed: Auto-HeG

--- your good choice to automatically construct heterophily-aware GNNs!

By fully exploring Graph Neural Architecture Search (GraphNAS) for heterophilic graphs:





Auto-HeG has the capability to:

- **Automatically customize GNNs for heterophilic graphs**
- **Comprehensive GNN architecture components friendly to heterophily →"Heterophilic Search Space"**
- **Efficiently & Effectively derive data-drive and graph-specific GNNs → "Heterophilic Search Strategy"**

Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

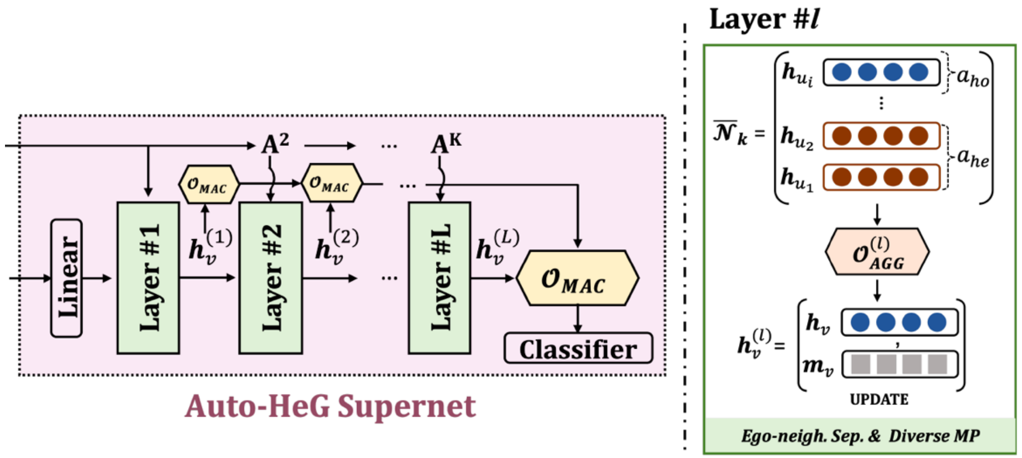
GRIFFITH UNIVERSITY

In-depth Look at Auto-HeG

• Heterophilic Search Space

Table 1: Heterophilic search space details of the proposed Auto-HeG. 'homo.' and 'hete.' indicate homophily-related and heterophily-related aggregation functions, respectively.

Search Space	Modules		Operations
Micro-level	Neighbors		$\{A, A^2, \cdots, A^K\}$
	O_{AGG}	homo.	{ SAGE, SAGE_SUM, SAGE_MAX, GCN, GIN, GAT, GAT_SYM, GAT_COS, GAT_LIN, GAT_GEN_LIN, GeniePATH}
		hete.	{ GCNII, FAGCN, GPRGNN, SUPERGAT, GCN_CHEB, APPNP, SGC}
Macro-level	O_{MAC}		$l\_skip, l\_zero, l\_concat, l\_max, l\_lstm$



• Heterophilic Search Strategy

(1) Progressive Heterophilic Supernet Training

Algorithm 1 Progressive Heterophilic Supernet Training

Require: Initial heterophilic supernet $\mathcal{S}_0$, number of shrinking iterations T, number of candidate operations C to be dropped per iteration t.

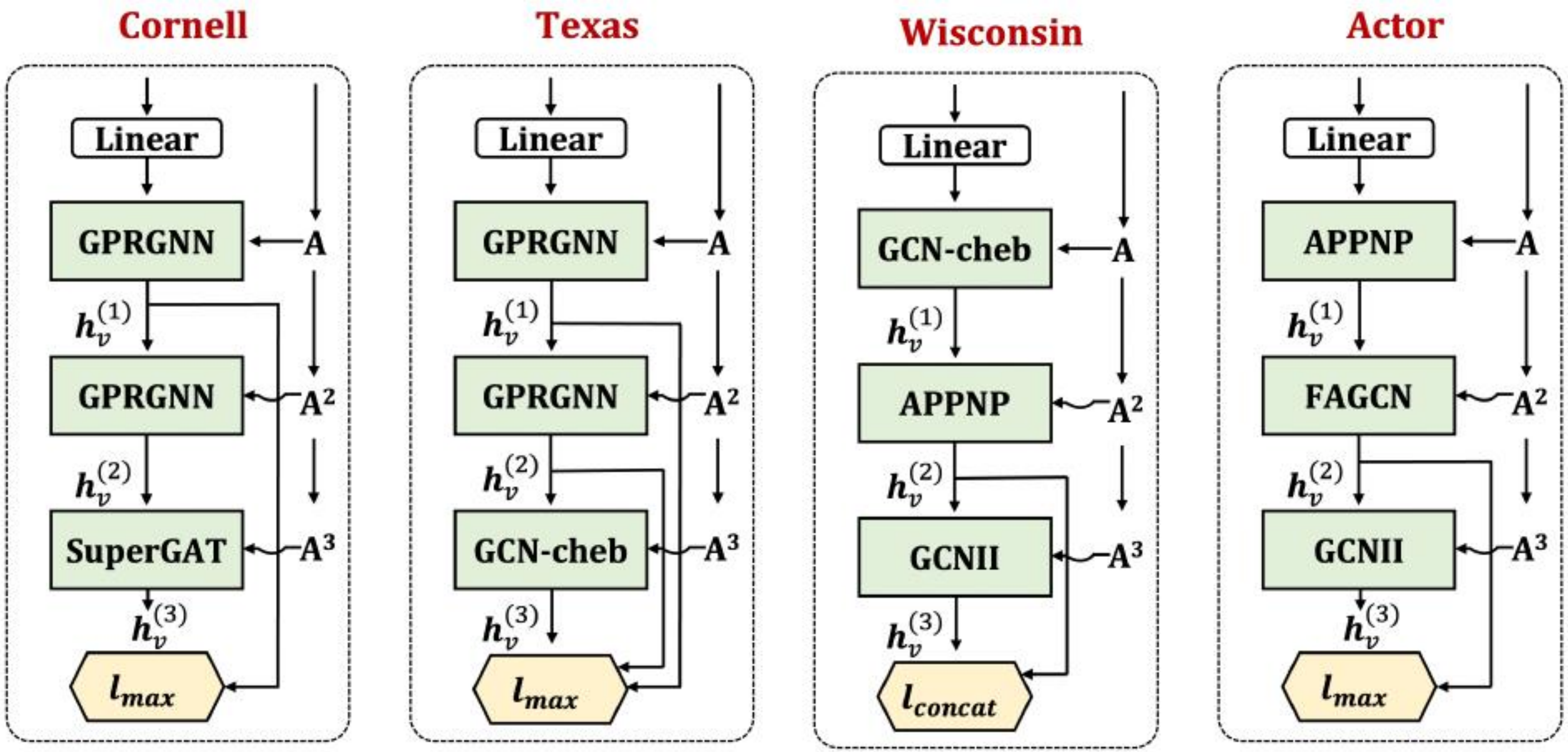
Ensure: $\boxed{\text{Compact heterophilic supernet } \mathcal{S}_c.}$

Let $\mathcal{S}_c \leftarrow \mathcal{S}_0$;
while $t < T$ **do**
  Training $\mathcal{S}_c$ for several epochs as Eq. (6) and (7);
  Ranking the magnitudes of the architecture $\boldsymbol{\alpha}$;
  Dropping C operations from $\mathcal{S}_c$ with the smallest C architecture weights;
end while

(2) Heterophily Guided Architecture Selection

Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

Auto-HeG Designed Heterophilic GNNs



Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

Experiments of Auto-HeG

--High-heterophily Graphs

Table 2: Performance (ACC%±std) of the proposed Auto-HeG compared with human-designed and graph NAS models on high-heterophily datasets. The best results are in bold and the second-best results are underline. Superscript * represents the officially reported results with the same dataset splits, where Geom-GCN and GCNII do not provide the std; And the remains are our reproduced results if official methods do not test under the same dataset splits.

Methods	Datasets	Cornell	Texas	Wisconsin	Actor
Human-designed models	H2GCN-1*	<u>82.16±4.80</u>	<u>84.86±6.77</u>	<u>86.67±4.69</u>	<u>35.86±1.03</u>
	H2GCN-2*	82.16±6.00	82.16±5.28	85.88±4.22	35.62±1.30
	MixHop*	73.51±6.34	77.84±7.73	75.88±4.90	32.22±2.34
	GPR-GNN	81.89±5.93	83.24±4.95	84.12±3.45	35.27±1.04
	GCNII*	76.49	77.84	81.57	-
	Geom-GCN-I*	56.76	57.58	58.24	29.09
	Geom-GCN-P*	60.81	67.57	64.12	31.63
	Geom-GCN-S*	55.68	59.73	56.67	30.30
	FAGCN	81.35±5.05	84.32±6.02	83.33±2.01	35.74±0.62
Graph NAS models	GraphNAS	58.11±3.87	54.86±6.98	56.67±2.99	25.47±1.32
	SNAG	57.03±3.48	62.70±5.52	62.16±4.63	27.84±1.29
	SANE	56.76±6.51	66.22±10.62	86.67±5.02	33.41±1.41
	SANE-hete	77.84±5.51	77.84±7.81	83.92±4.28	35.88±1.30
	Auto-HeG (ours)	**83.51±6.56**	**86.76±4.60**	**87.84±3.59**	**37.43±1.37**

Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

Experiments of Auto-HeG

-- Low-heterophily Graphs

Table 3: Performance (ACC%±std) of the proposed Auto-HeG compared with human-designed and graph NAS models on low-heterophily datasets.

Methods	Datasets	Cora	Citeseer	Pubmed
Human-designed models	GCN	85.69±1.80	75.38±1.75	86.08±0.64
	GAT	86.52±1.41	75.51±1.85	84.75±0.51
	GraphSAGE	80.60±3.63	67.18±5.46	81.18±1.12
	SGC	85.88±3.61	73.86±1.73	84.87±2.81
	GCNII*	88.01	77.13	90.30
	Geom-GCN-I*	85.19	77.99	90.05
	Geom-GCN-P*	84.93	75.14	88.09
	Geom-GCN-S*	85.27	74.71	84.75
Graph NAS models	GraphNAS	84.10±0.79	68.83±2.09	82.28±0.64
	SNAG	81.01±1.31	70.14±2.40	83.24±0.84
	SANE	84.25±1.82	74.33±1.54	87.82±0.57
	SANE-hete	85.05±0.90	74.46±1.59	88.99±0.42
	Auto-HeG (ours)	86.88±1.10	75.81±1.52	89.29±0.27

- ❖ Both high-heterophily and low-heterophily graphs' node classification results show the superior of the proposed Auto-HeG to existing human designed GNNs & automated GNNs

Zheng, Xin, et al. "Auto-HeG: Automated graph neural network on heterophilic graphs." Proceedings of the ACM Web Conference 2023. 2023.

GRIFFITH UNIVERSITY

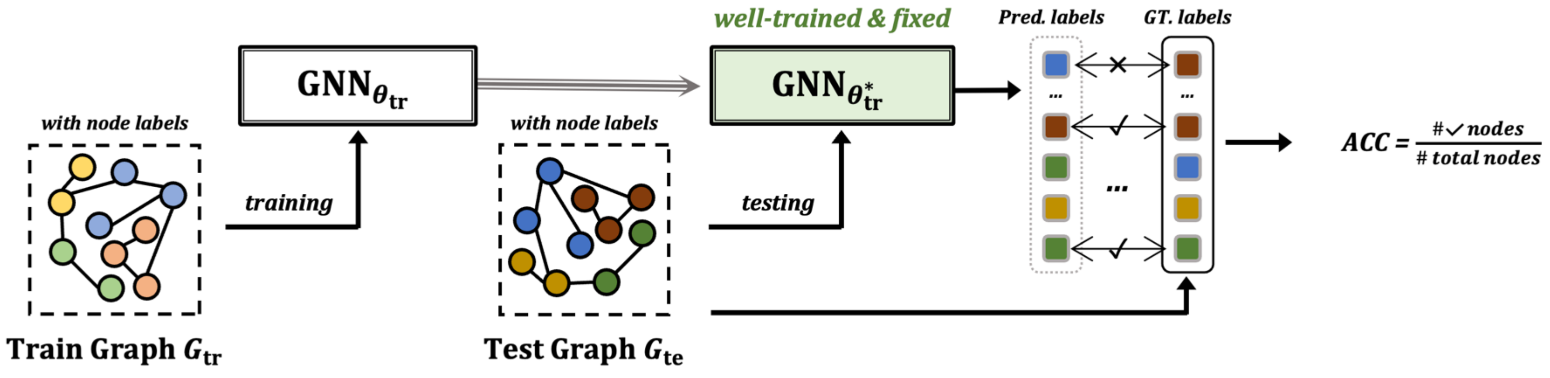
Contents

- **Introduction & Overview**
  - *Automated Graph MLOps in Data-Centric AI*
- **Graph Data-Centric Exploitation**
  - *Graph Type*
  - *Graph Scale*
- **Graph Data-Centric Model Deployment**
  - *GNN Model Evaluation*
- **Future Opportunities**

CRICOS: 00233E | TEQSA: PRV12076

Conventional Model Evaluation

Understanding and evaluating GNN models' performance is a vital step for GNN model deployment and serving.



(a) Conventional GNN Model Evaluation

For instance,

in financial transaction networks:

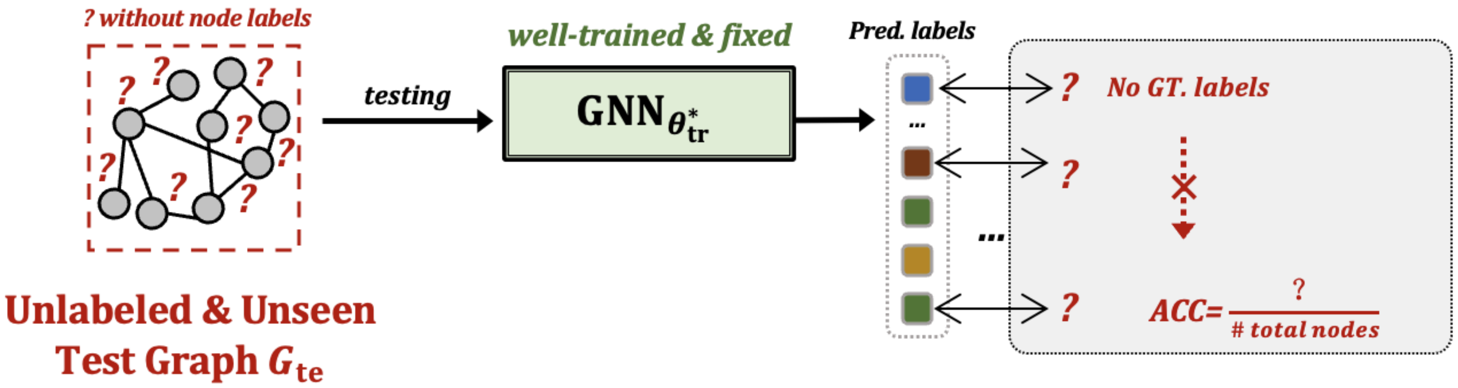
- **GNN model designers:** expect their developed GNNs to excel in identifying newly emerging suspicious transactions
- **Users:** ensure how they could trust well-trained GNNs to know suspicious transactions within their own data

In conventional model evaluation of GNNs, we have:

1) Seen test graph $\boldsymbol{G}_{\mathbf{te}}$ in the same distribution as the train graph $\boldsymbol{G}_{\mathbf{tr}}$
2) Known test graph labels for computing performance metric, e.g., Accuracy (ACC)

Zheng, X., Zhang, M., Chen, C., Molaei, S., Zhou, C., & Pan, S. (2023). GNNEvaluator: Evaluating GNN Performance On Unseen Graphs Without Labels. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Real-world Model Evaluation



(b) Real-world GNN Model Evaluation

In real-world model evaluation of GNNs, we:

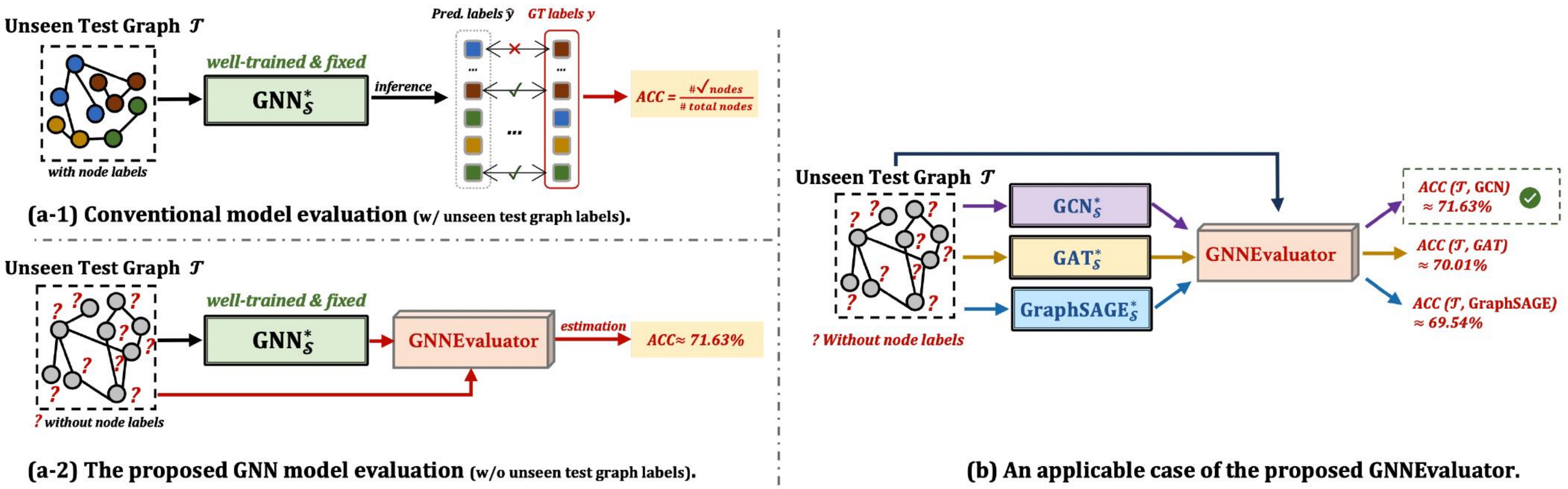
- X CAN NOT access the ground-truth labels of the test graph G_{te}
- X CAN NOT compute performance metric, e.g., Accuracy (ACC)
- X DO NOT know whether potential distribution shifts from the train graph G_{tr}

Zheng, X., Zhang, M., Chen, C., Molaei, S., Zhou, C., & Pan, S. (2023). GNNEvaluator: Evaluating GNN Performance On Unseen Graphs Without Labels. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Test-Time GNN Model Evaluation

Given above scenarios, a natural question, i.e., **“GNN model evaluation problem”** arises:

In the absence of labels in an unseen test graph, can we estimate the performance of a well-trained GNN model?



Zheng, X., Zhang, M., Chen, C., Molaei, S., Zhou, C., & Pan, S. (2023). GNNEvaluator: Evaluating GNN Performance On Unseen Graphs Without Labels. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Definition of GNN Model Evaluation

Definition of GNN Model Evaluation. Given the observed training graph $\mathcal{S}$, its well-trained model $\text{GNN}^*_{\mathcal{S}}$, and an unlabeled unseen graph $\mathcal{T}$ as inputs, the **goal** of GNN model evaluation aims to learn an accuracy estimation model $f_\phi(\cdot)$ parameterized by ϕ as:

$$\text{Acc}(\mathcal{T}) = f_\phi(\text{GNN}^*_{\mathcal{S}}, \mathcal{T}), \tag{2}$$

where $f_\phi : (\text{GNN}^*_{\mathcal{S}}, \mathcal{T}) \rightarrow a$ and $a \in \mathbb{R}$ is a scalar denoting the overall node classification accuracy $\text{Acc}(\mathcal{T})$ for all unlabeled nodes of $\mathcal{T}$. When the context is clear, we will use $f_\phi(\mathcal{T})$ for simplification.

To solve above problems,

We propose a two-stage GNN model evaluation framework with a "GNNEvaluator"

Note that our principal goal is to estimate well-trained GNN models' performance, rather than improve the generalization ability of new GNN models. In the whole evaluation process, the in-service GNN model is fixed

Zheng, X., Zhang, M., Chen, C., Molaei, S., Zhou, C., & Pan, S. (2023). GNNEvaluator: Evaluating GNN Performance On Unseen Graphs Without Labels. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Our Proposed GNNEvaluator

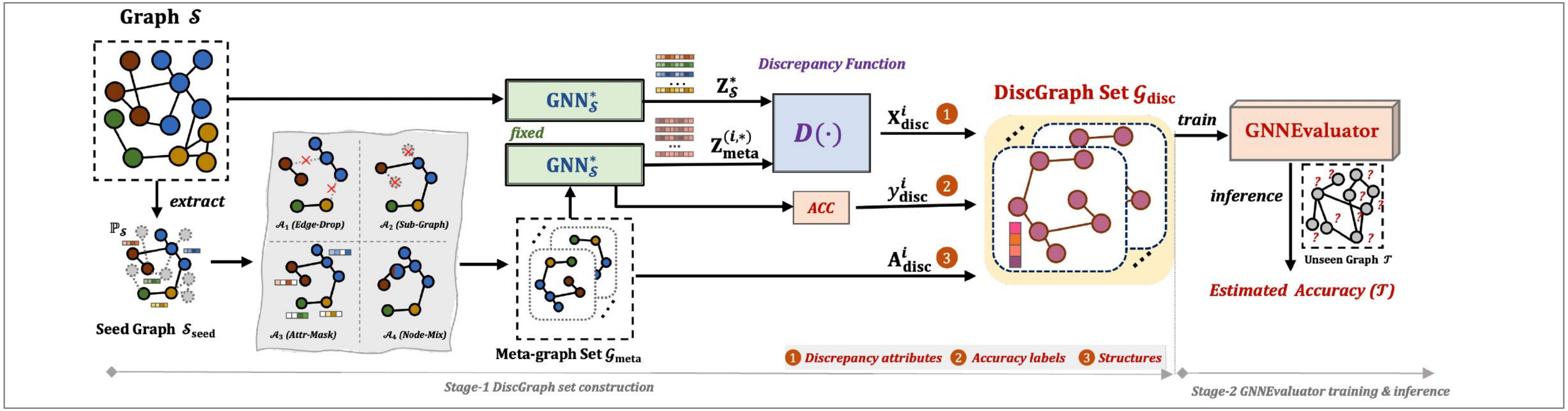


Figure.1 Overall two-stage framework of the proposed GNN model evaluation with GNNEvaluator

- **Stage-1: DiscGraph set construction**

  incorporating training-test graph discrepancies into DiscGraph node attributes $\mathbf{X}^i_{\text{disc}}$, structures $\mathbf{A}^i_{\text{disc}}$, and accuracy labels y^i_{disc}

- **Stage-2: GNNEvaluator training and inference**

  GNNEvaluator, train on DiscGraphs and output estimated ACC on the real-world test graph $\mathcal{T}$.

Zheng, X., Zhang, M., Chen, C., Molaei, S., Zhou, C., & Pan, S. (2023). GNNEvaluator: Evaluating GNN Performance On Unseen Graphs Without Labels. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Experiments of GNNEvaluator

The performance of our proposed GNNEvaluator in evaluating well-trained GNNs' node classification accuracy under all test evaluation cases and models

Table 1: Mean Absolute Error (MAE) performance of different GNN models across five random seeds. (GNNs are well-trained on the ACMv9 dataset and evaluated on the unseen and unlabeled Citationv2 and DBLPv8 datasets, *i.e.*, A→C and A→D, respectively. Best results are in bold.)

Methods	ACMv9→Citationv2						ACMv9→DBLPv8					
	GCN	SAGE	GAT	GIN	MLP	*Avg.*	GCN	SAGE	GAT	GIN	MLP	*Avg.*
ATC-MC [8]	4.49	8.40	4.37	18.40	34.33	14.00	21.96	24.20	30.30	24.06	26.62	25.43
ATC-MC-c [8]	**2.41**	5.74	4.67	22.00	51.41	17.25	31.15	30.55	30.18	29.71	45.81	33.48
ATC-NE [8]	3.97	8.02	4.28	17.35	38.87	14.50	22.93	24.78	30.50	23.74	31.13	26.62
ATC-NE-c [8]	4.44	6.09	**3.30**	23.95	44.62	16.48	34.42	28.31	27.02	30.28	39.28	31.86
Thres. ($\tau = 0.7$) [6]	32.64	35.81	33.63	50.76	35.28	37.63	9.59	12.14	14.30	32.67	39.72	21.68
Thres. ($\tau = 0.8$) [6]	26.30	29.60	26.18	49.25	35.87	33.44	**2.63**	7.44	14.47	32.20	40.31	19.41
Thres. ($\tau = 0.9$) [6]	17.56	21.34	16.38	46.53	36.08	27.58	8.20	7.42	16.07	31.47	40.56	20.74
AutoEval-G [6]	18.94	26.19	26.12	50.86	32.40	30.90	2.77	**2.54**	7.25	48.68	29.95	18.24
GNNEvaluator **(Ours)**	4.85	**4.11**	12.23	**10.14**	**22.20**	**10.71**	11.80	14.88	**6.36**	**13.78**	**17.49**	**12.86**

Table 2: Mean Absolute Error (MAE) performance of different GNN models across five random seeds. (GNNs are well-trained on the Citationv2 dataset and evaluated on the unseen and unlabeled ACMv9 and DBLPv8 datasets, i.e., C→A and C→D, respectively.Best results are in bold.)

Methods	Citationv2→ACMv9						Citationv2→DBLPv8					
	GCN	SAGE	GAT	GIN	MLP	*Avg.*	GCN	SAGE	GAT	GIN	MLP	*Avg.*
ATC-MC [8]	9.50	13.40	8.28	35.51	43.40	22.02	22.57	**1.37**	21.87	29.24	35.20	22.05
ATC-MC-c [8]	6.93	11.75	**6.70**	38.93	57.43	24.35	33.67	4.92	28.23	30.89	52.59	30.06
ATC-NE [8]	8.86	13.04	7.87	34.88	47.49	22.42	23.97	1.86	23.74	28.96	39.72	23.65
ATC-NE-C [8]	7.73	13.94	7.63	41.17	62.96	26.69	37.16	4.66	29.43	31.66	58.95	32.37
Thres. ($\tau = 0.7$) [6]	37.33	36.61	31.68	58.91	34.33	39.77	10.70	23.05	12.74	34.60	38.29	23.88
Thres. ($\tau = 0.8$) [6]	29.62	28.95	22.77	57.48	34.53	34.67	5.65	15.01	7.61	34.36	38.43	20.21
Thres. ($\tau = 0.9$) [6]	19.59	19.06	11.37	55.72	34.56	28.06	10.65	8.28	8.07	34.00	38.44	19.89
AutoEval-G [6]	23.01	31.24	26.74	59.66	35.02	28.28	**2.57**	16.52	6.96	19.20	32.24	24.59
GNNEvaluator **(Ours)**	**5.45**	**8.53**	9.61	**29.77**	**28.52**	**16.38**	11.64	7.02	**5.58**	**6.46**	**22.87**	**10.71**

- ❖ Experiments on 3 real-world graph datasets in 6 cases potential domain shift, each evaluating 5 models:
- ❖ *Consistent outstanding performance over all GNN models and cases!*

43

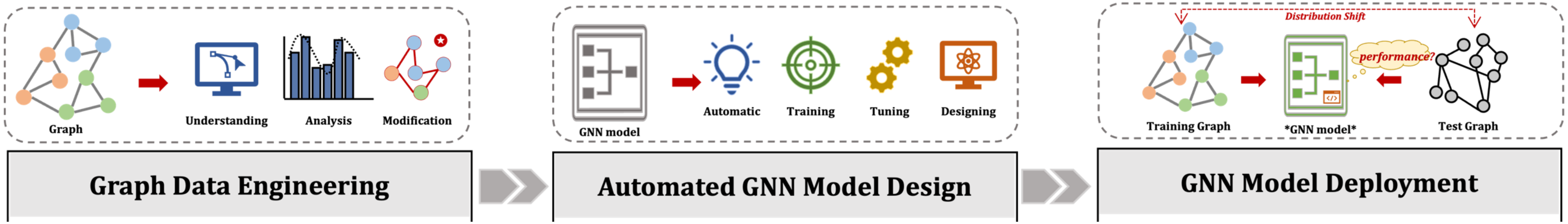
Zheng, X., Zhang, M., Chen, C., Molaei, S., Zhou, C., & Pan, S. (2023). GNNEvaluator: Evaluating GNN Performance On Unseen Graphs Without Labels. Advances in Neural Information Processing Systems (NeurIPS), 2023.

Contents

- **Introduction & Overview**
  - *Automated Graph MLOps in Data-Centric AI*
- **Graph Data-Centric Exploitation**
  - *Graph Type*
  - *Graph Scale*
- **Graph Data-Centric Model Deployment**
  - *GNN Model Evaluation*
- **Future Opportunities**

GRIFFITH UNIVERSITY

Futures of Automated Graph MLOps



A. 【Graph Data Level】In-depth exploration, understanding, and management of various graph data types and characteristics

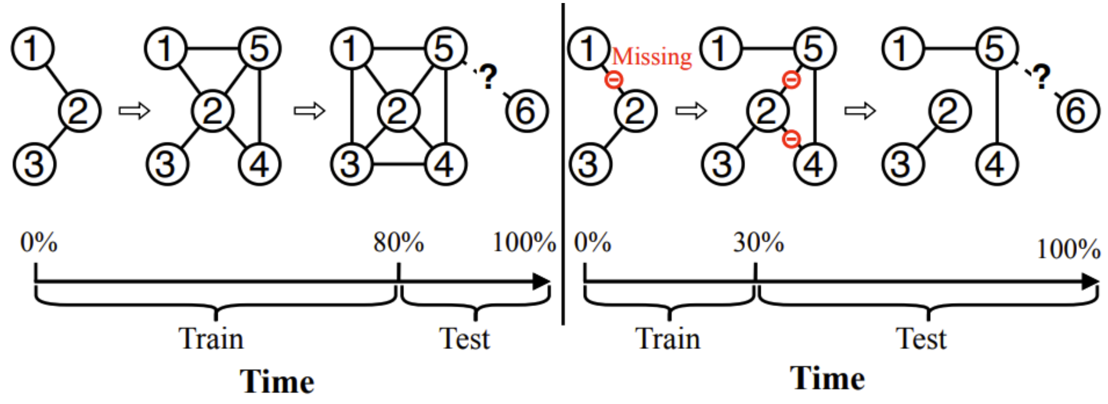
B. 【GNN Model Level】Automated and explainable GNN models with robustness, fairness

C. 【Data + Model Level】Good generalization and transfer learning abilities of both graph data and models

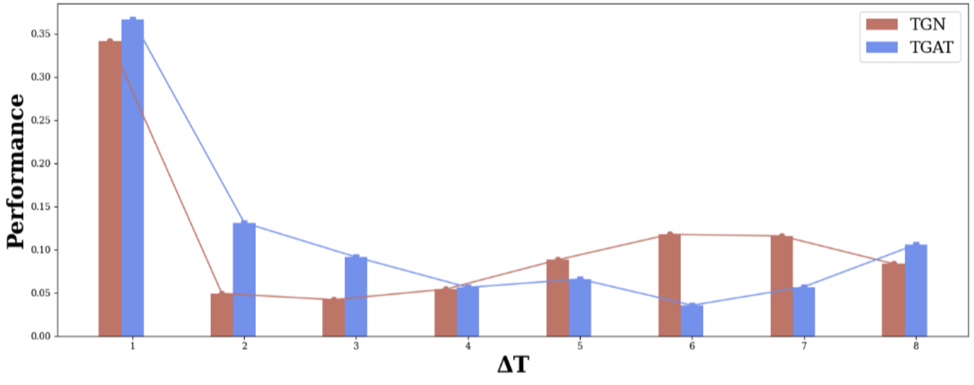
D. 【GNN Application Level】Continuous evaluation, integration, deployment, and monitor of GNN models

Futures of Automated Graph MLOps

A. 【Graph Data Level】In-depth exploration, understanding, and management of various graph data types and characteristics



Example: Dynamic Graph

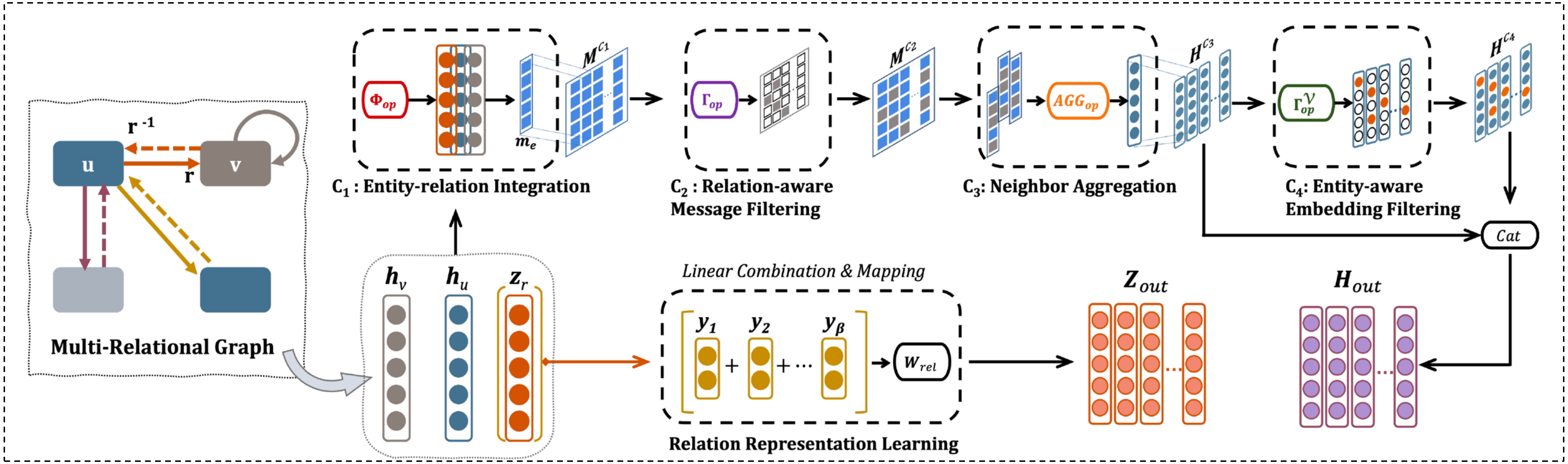


(b) Inference uncertainty on unseen dynamic graphs

Futures of Automated Graph MLOps

B. 【GNN Model Level】Automated and explainable GNN models with robustness, fairness

【ICDM'2022】 "Multi-Relational Graph Neural Architecture Search (MR-GNAS)"



Zheng, X., Zhang, M., Chen, C., Li, C., Zhou, C., & Pan, S. (2022, November). Multi-relational graph neural architecture search with fine-grained message passing. ICDM'2022

Futures of Automated Graph MLOps

C. 【Data + Model Level】Good generalization and transfer learning abilities of both graph data and models

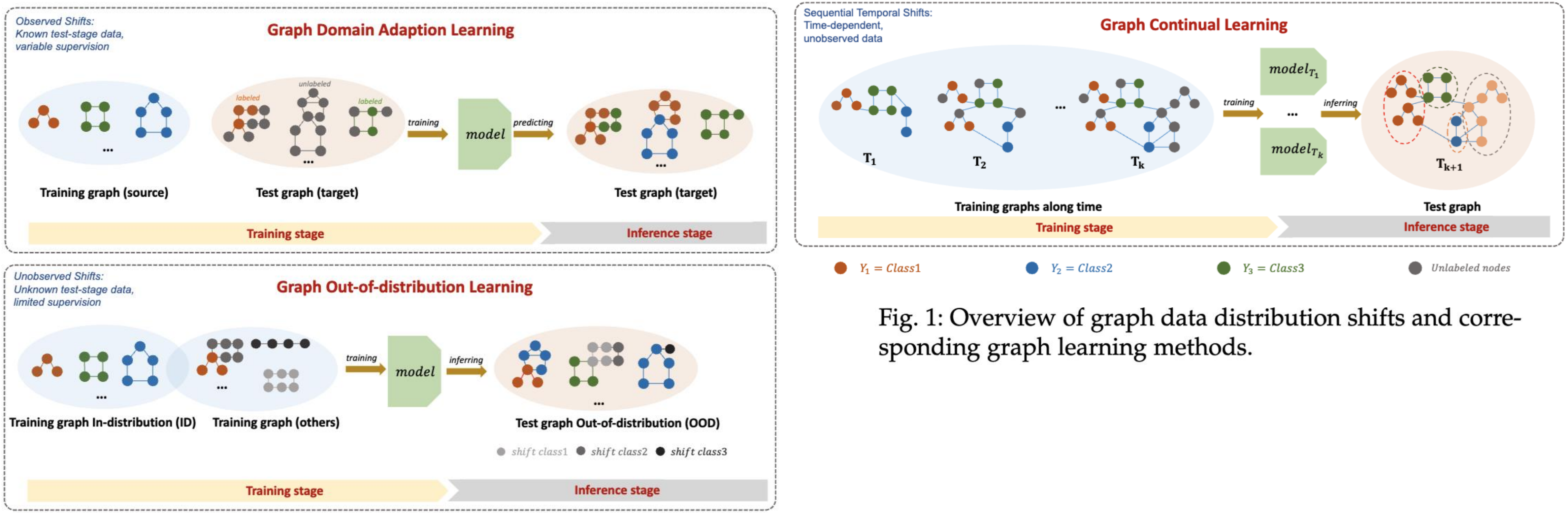


Fig. 1: Overview of graph data distribution shifts and corresponding graph learning methods.

Wu, M., Zheng, X., Zhang, Q., Shen, X., Luo, X., Zhu, X., & Pan, S. (2024). Graph Learning under Distribution Shifts: A Comprehensive Survey on Domain Adaptation, Out-of-distribution, and Continual Learning. arXiv preprint arXiv:2402.16374.

Futures of Automated Graph MLOps

D. 【GNN Application Level】 Continuous evaluation, integration, deployment, and monitor of GNN models

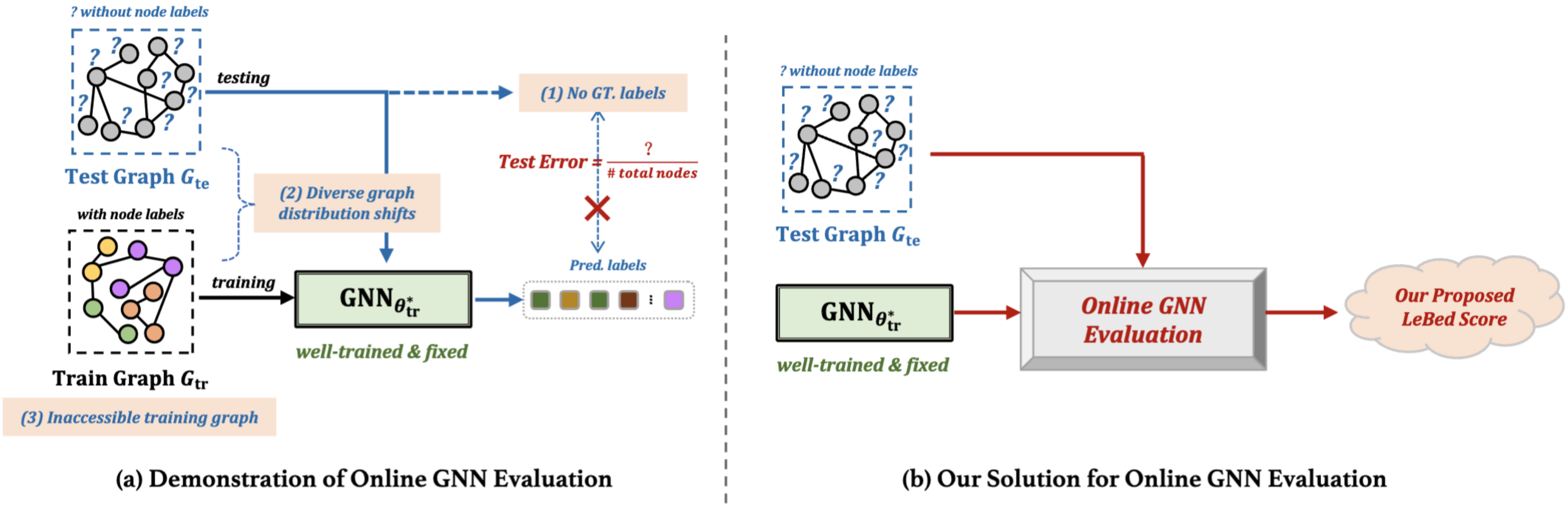


Figure 1: Illustration of the proposed online GNN evaluation problem and our solution.

Zheng, X., Song, D., Wen, Q., Du, B., & Pan, S (2024). Online GNN Evaluation Under Test-time Graph Distribution Shifts. In The Twelfth International Conference on Learning Representations, (ICLR).

Thanks!

Automated Graph Machine Learning Operations (MLOps) Workflow:

A Data-Centric Perspective

Dr. Xin Zheng

Griffith University