

# Instance-level Node Profiling For Graph-structured Data

Tianqi Zhao

05.03.2026

# It's a common day...



Baseline A

Accuracy

60%



Baseline B

62%



Our method

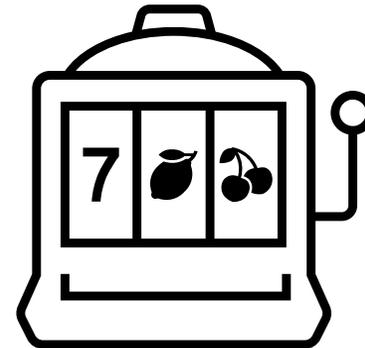
64%



Now what should I do?

# Experiment Analysis

- ▶ **How did the models perform?**
  - ▶ Compare to baseline A to B, B...
  - ▶ Our model achieves upto x% improvement...
- ▶ **Model-centric analysis**
  - ▶ **Evaluation metric based comparison**
    - ▶ Accuracy, F1-score, ...

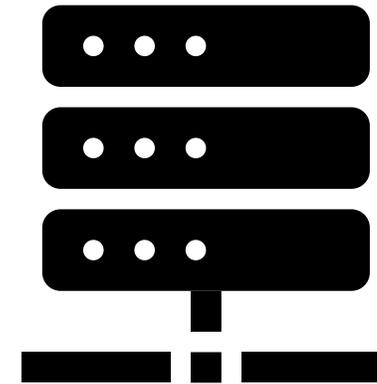


# Data Analysis

- ▶ **Graph data properties in a table**
  - ▶ Similarity measurements
  - ▶ Connectivity
  - ▶ ...

**And Hurray!! We did it!!**

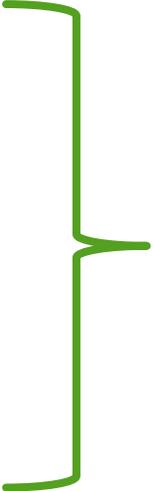
**wait..... did we?  
But WHY?**



Data-centric

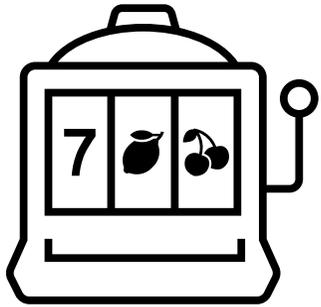
# Experiment Analysis

- ▶ **Model-centric analysis**
  - ▶ How did the baselines perform?
  - ▶ How did our method perform?
- ▶ **Data-centric analysis**
  - ▶ How do the datasets look like?
  - ▶ Where do they come from?



What about the connection?

# What does the connection tell?



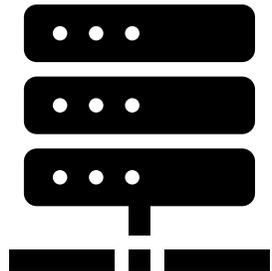
model behavior analysis



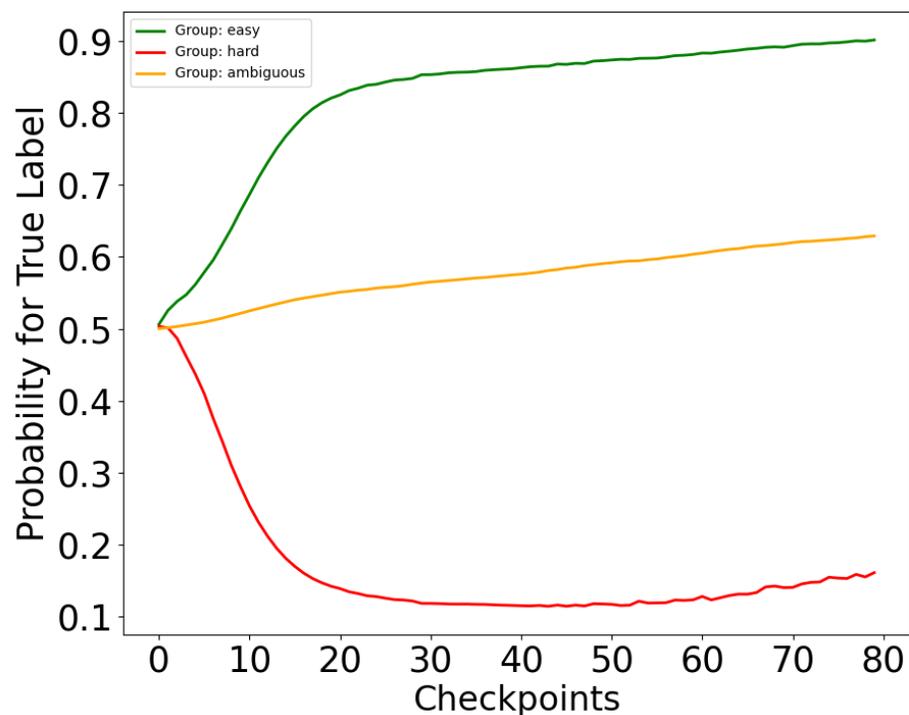
On this kind of data, models like A excel/fail dataset global view → a instance-level view  
Answers to “why”



data property measurement



# One dataset, different node-level behaviors



- ▶ **Easy nodes:**
  - ▶ model makes consistent **right** predictions
- ▶ **Hard nodes:**
  - ▶ model makes consistent **wrong** predictions
- ▶ **Ambiguous nodes:**
  - ▶ **others**

GCN trained on Bitcoin-Alpha.  
User, trust rating, and good/bad

# It's a REAL day...



Baseline A

60%



Baseline B

61%



Our method

61.5%

Accuracy ← Aggregated metric

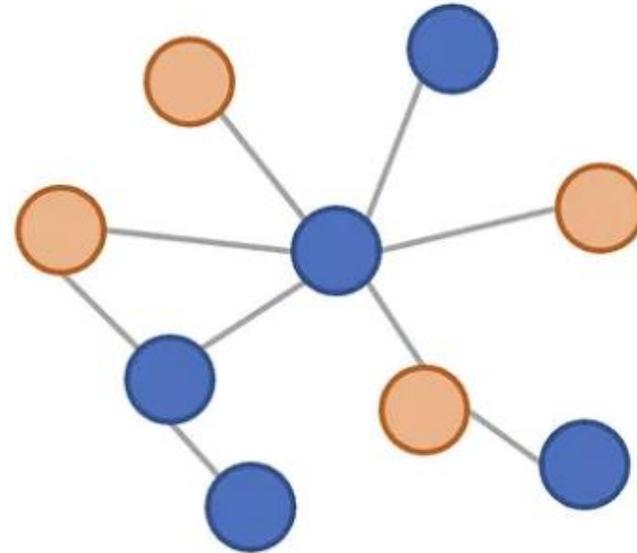


Would it still be possible to see differences in models behavior??

# Task Definition

- ▶ Given:
  - ▶ Graph topology
  - ▶ Input features
  - ▶ Incomplete labelling
- ▶ Task:
  - ▶ Predict labels of unlabelled nodes

## Node Classification



# NodePro

1

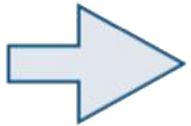
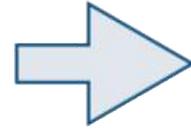
## Data-centric Profile

- Intra-class feature dissimilarity
- Neighbourhood class divergence
- Random walk class divergence

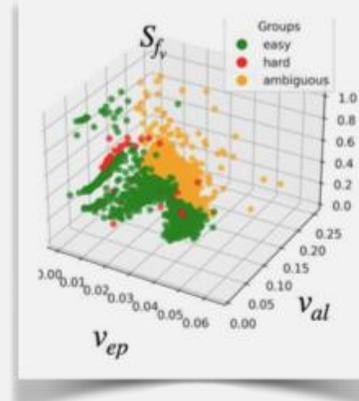
## Model-centric Profile

- Prediction uncertainty
- Prediction consistency

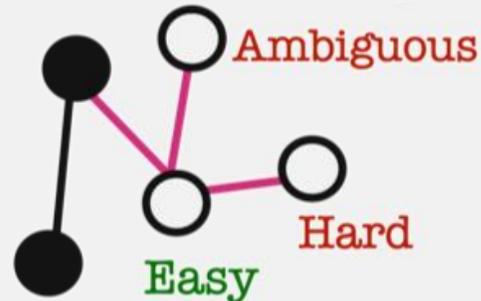
NodePro



## Node-level diagnostic



## Inductive Profiling



2

3

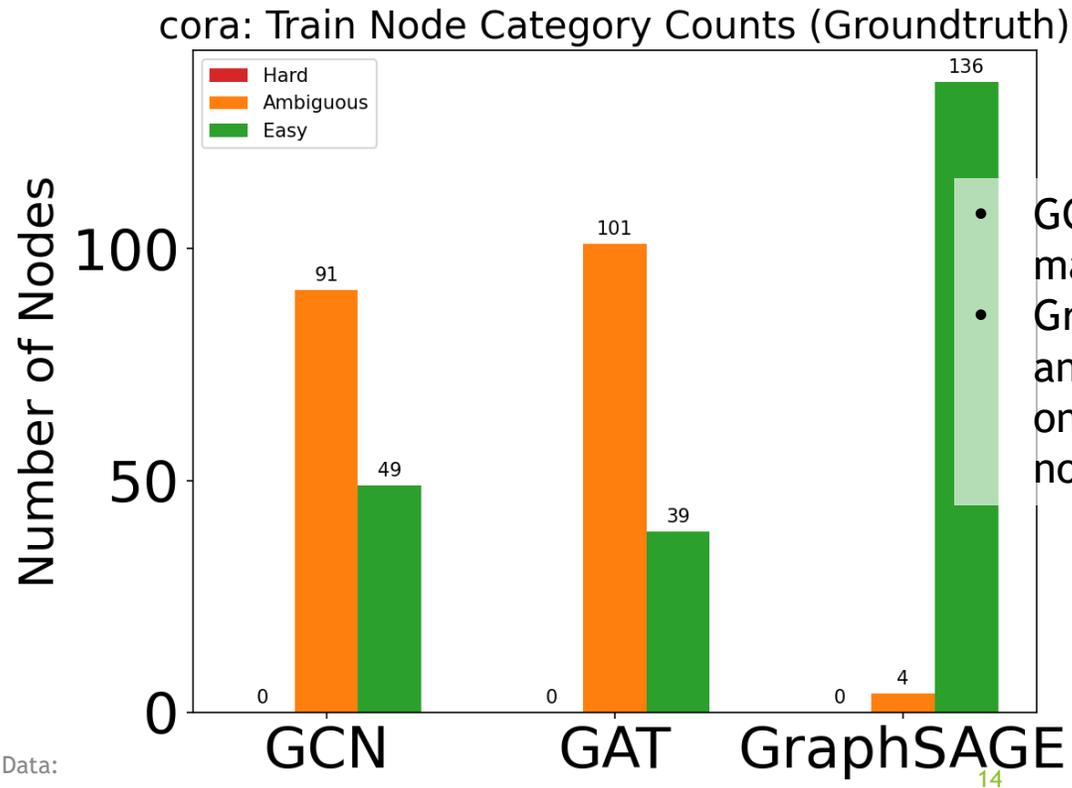
# 1. Similar accuracy, different behaviors

- ▶ For a given node-classification task, we've collected the following empirical results:

CORA	
GCN	0.801/0.981
GAT	0.780/0.997
GRAPHSAGE	0.780/1.000
MLP	0.569/1.000

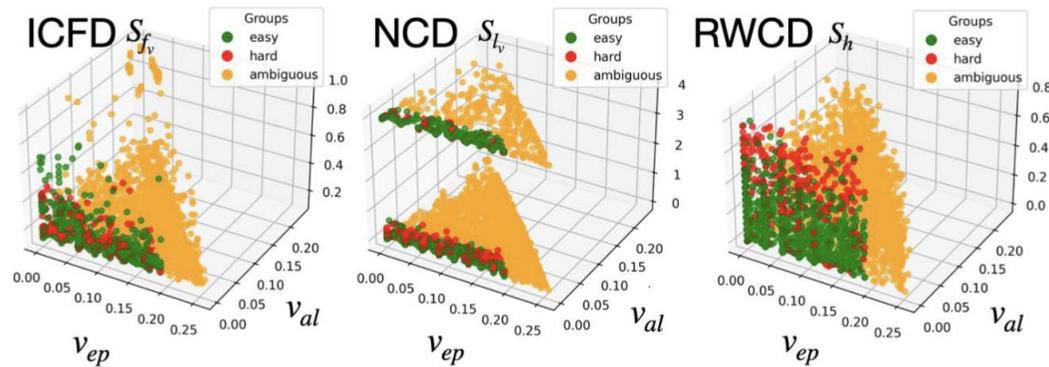
# 1. Similar accuracy, different behaviors

CORA	
GCN	0.801/0.981
GAT	0.780/0.997
GraphSAGE	0.780/1.000
MLP	0.569/1.000



- GCN and GAT struggles with majority of the nodes
- GraphSAGE make confident and consistent predictions on majority of the training nodes

## 2. Overconfidence of a model



(a) GRAPHSAGE trained on CREDIT with complete data.

- ▶ colors indicates easy and hard
- ▶ Low aleatoric uncertainty indicates high confidence
- ▶ Right and confident!
- ▶ Wrong but confident?

# 3. Why do models struggle?

- ▶ Full train set up in Credit

	CORA	CREDIT	BITCOINALPHA
GCN	0.801/0.981	0.613/0.661	0.850/0.850
GAT	0.780/0.997	0.594/0.641	0.594/0.732
GRAPHSAGE	0.780/1.000	0.614/0.690	0.810/0.850
MLP	0.569/1.000	0.623/0.680	0.738/0.740

customers, similar payment patterns, credit status

# 3. Why do models struggle on the data?

## ► Full train set up in Credit

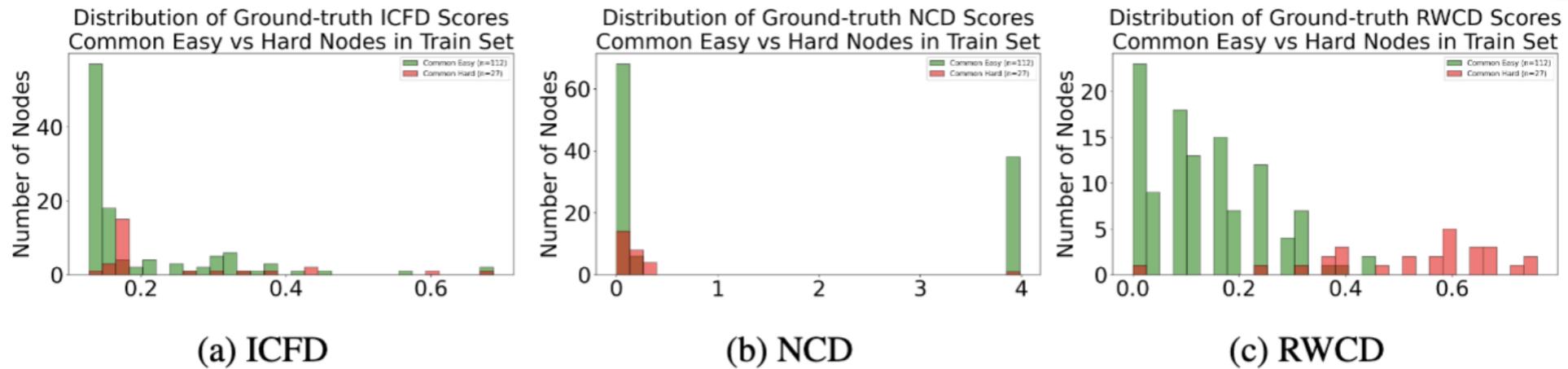


Figure 3: Common Easy vs Hard node across GNN models have different data-centric profiling score distributions in the training data in CREDIT.

## 4. Inductive profiling on test data

Category	Entity 1	Relation	Entity 2	Label flip
easy	Howl	film/music	Carter Burwell	No
easy	True Grit	film/music	Carter Burwell	No
ambiguous	bisexuality	netflix genre	Velvet Goldmine	Yes
ambiguous	Eddie Izzard	film	Velvet Goldmine	No
hard	Film score	music genre	Carter Burwell	Yes
hard	Velvet Goldmine	film location	London	Yes

- 97.3% of the hard nodes are with injected errors
- 48.9% error nodes are identified as hard

# Limitation and future work

- ▶ **NodePro: interpretable instance-level profiling**
  - ▶ Combines **data + model** profiles to uncover **hidden patterns** beyond accuracy/loss
  - ▶ Profiles **generalize to unseen nodes** --> **reliability assessment without labels**
- ▶ **Next:**
  - ▶ richer data profiles
  - ▶ apply to **model selection**
  - ▶ Multi-label graph



Thank you