KDD 2022

UD-GNN: Uncertainty-aware Debiased Training on Semi-Homophilous Graphs

Yang Liu[1]; Xiang Ao[1*]; Fuli Feng[2]; Qing He[1*]

柳 阳[1]; 敖 翔[1*]; 冯福利[2]; 何 清[1*]

1 中国科学院计算技术研究所 INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

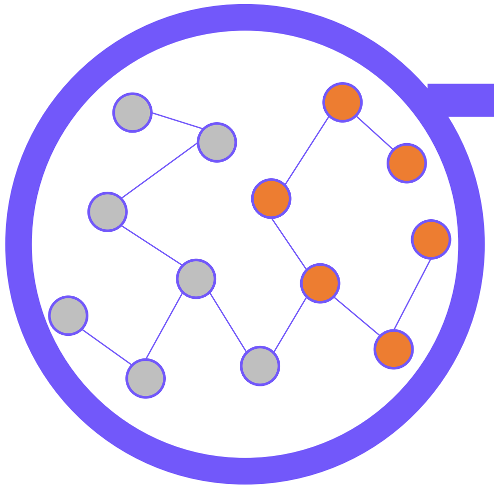
2 中国科学技术大学 University of Science and Technology of China

* *denotes corresponding author.*

Content

- Background and Motivation
- Method – UD-GNN
- Experiment
- Conclusion and Future Work

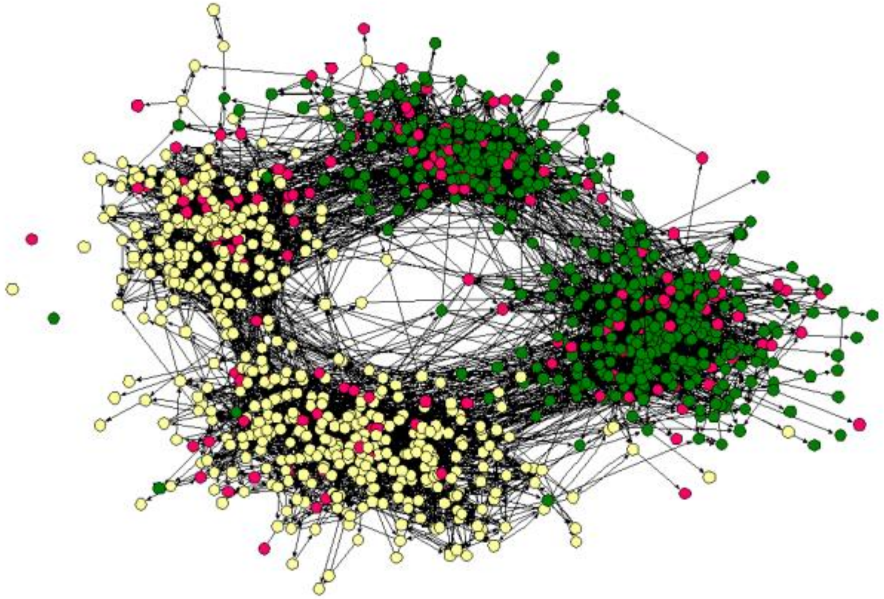
Background

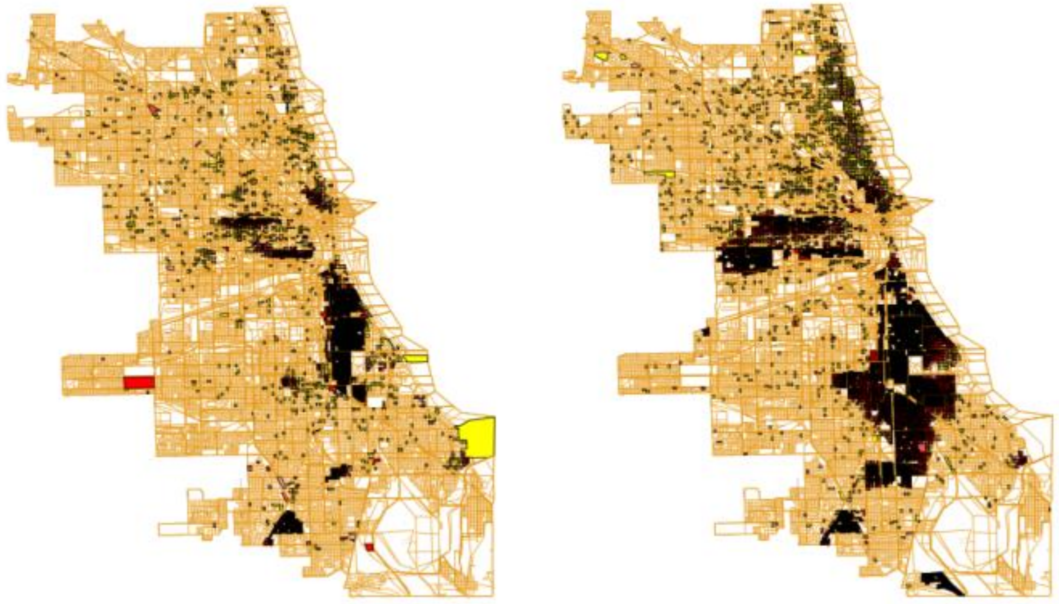


Homophily: The tendency of individuals to associate and bond with **similar** others

- People with the same interest are more closely connected
- Researchers who focus on the same research area are more likely to establish a connection

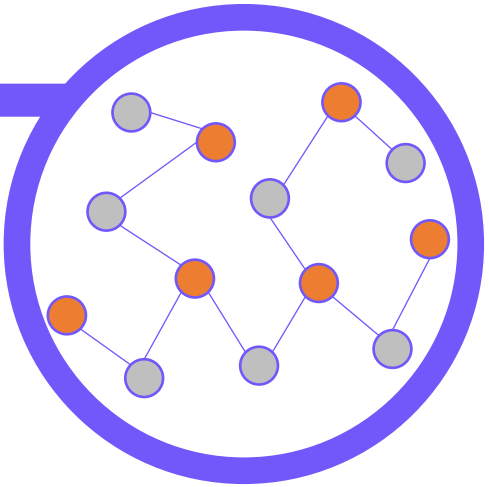
Homophilous Graph





(a) *Chicago, 1940* (b) *Chicago, 1960*

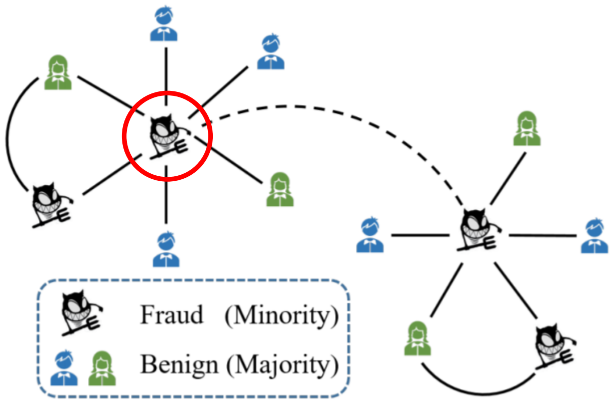
Images credit to Easley and Kleinberg, Networks, crowds, and markets. 2010



Heterophily: Nodes from different classes tend to connect to each other

- Fraudsters connect to benign users to camouflage themselves
- Interdisciplinary papers cite papers from other research areas

Heterophilous Graph



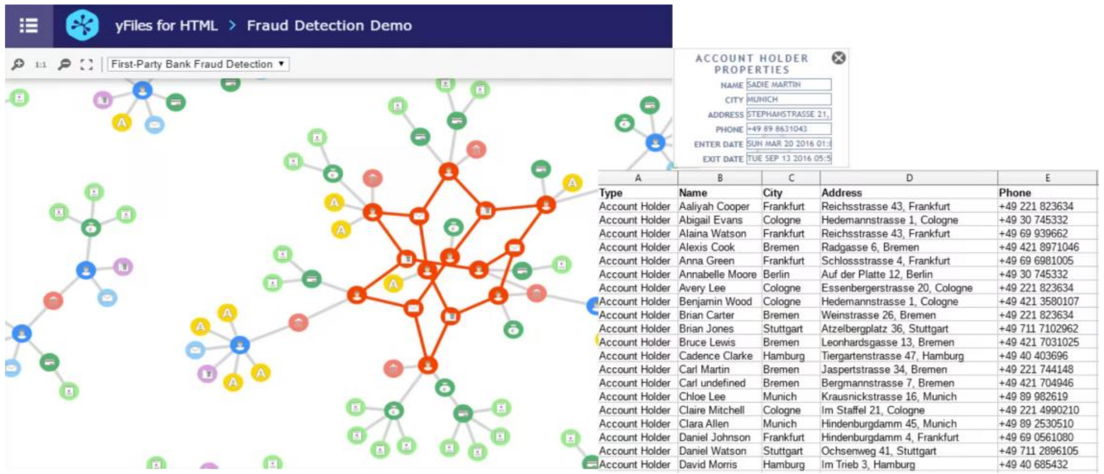
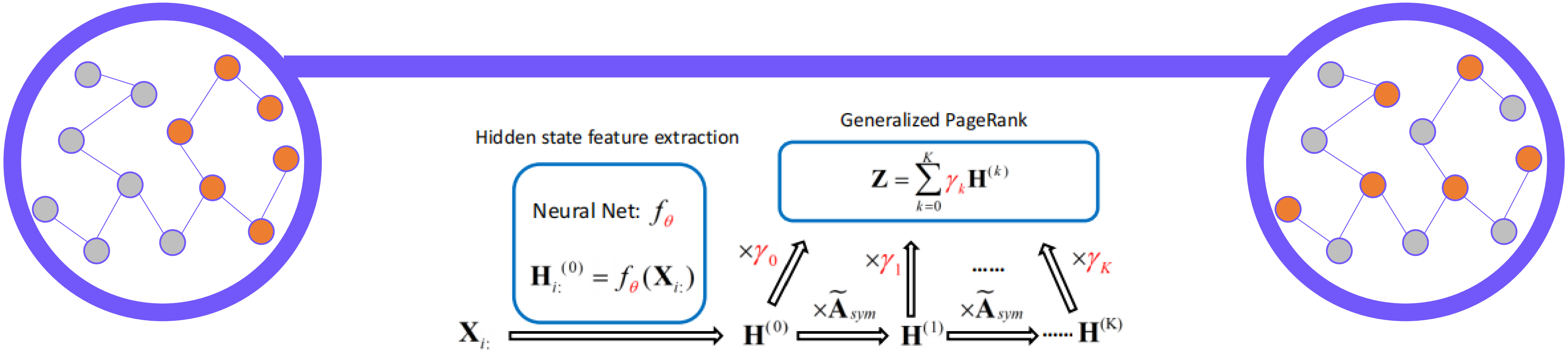


Image (Right) credit to yworks.com

Related Work

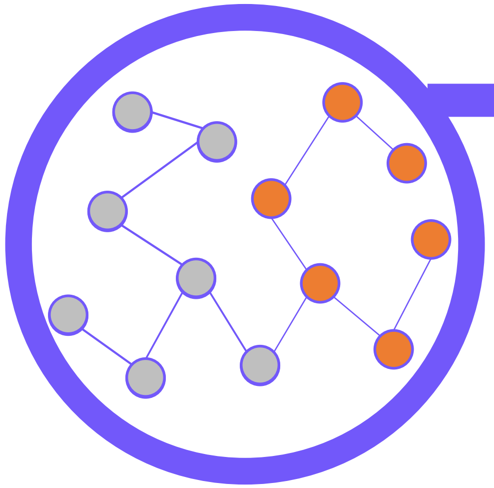


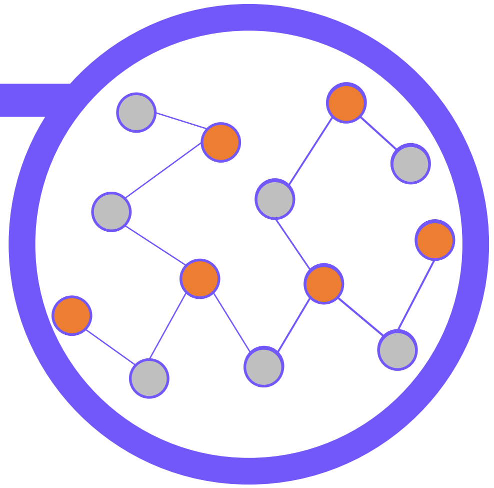
GNNs for Homophilous Graphs

GNNs for Heterophilous Graphs

- Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." ICLR 2017.
- Hamilton, W. L.; Ying, R.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. NeurIPS 2017.
- Veličković, Petar, et al. "Graph attention networks." ICLR 2018.
- Johannes Klicpera et al. "Predict then propagate: Graph neural networks meet personalized pagerank." ICLR 2018.

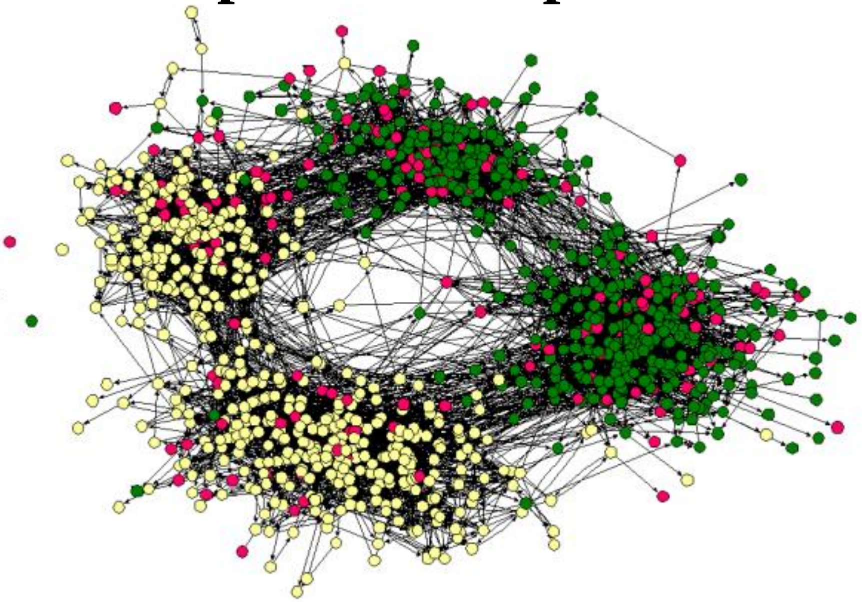
- Zhu, Jiong, et al. "Beyond homophily in graph neural networks: Current limitations and effective designs." NeurIPS 2020.
- Chien, Eli, et al. "Adaptive universal generalized pagerank graph neural network." ICLR 2020.
- Zhu, Jiong, et al. "Graph neural networks with heterophily." AAAI 2021.
- Lim, Derek, et al. "Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods." NeurIPS 2021.



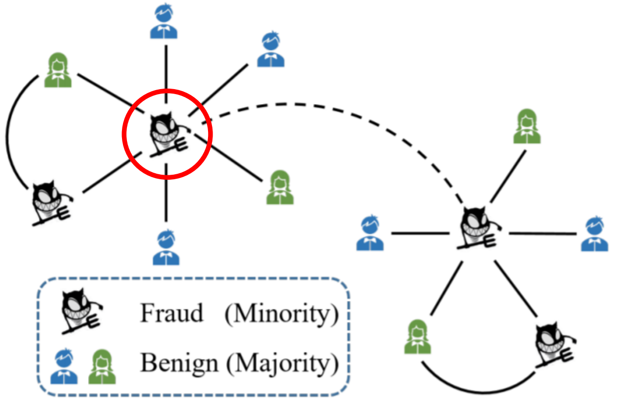


Homophilous Graph

Heterophilous Graph

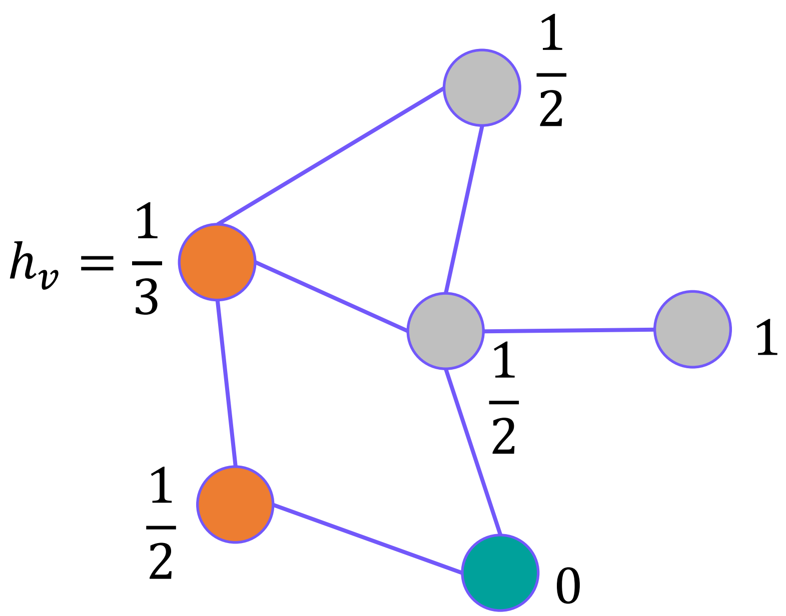


Semi-Homophilous Graph



Lim, Derek, et al. "Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods." NeurIPS 2021.

Homophily Ratio



Definitions:

$$\mathcal{G} = (\mathcal{V}, \mathbf{A}, \mathbf{X}, \mathbf{Y})$$

$$\mathcal{E}_{\text{intra}} = \{(u,v) | \mathbf{A}_{uv} = 1 \wedge y_u = y_v\}$$

$$\mathcal{E}_v = \{(u,v) | \mathbf{A}_{uv} = 1\}$$

- Graph-level Homophily Ratio:

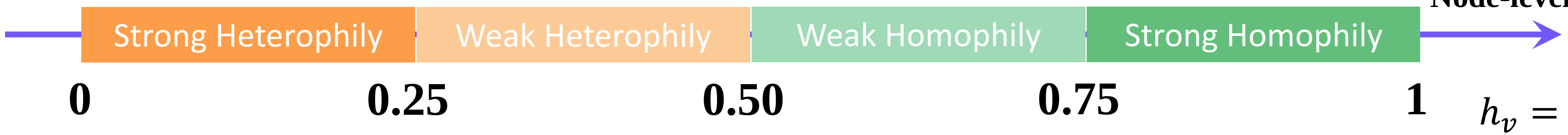
$$h_{\mathcal{G}} = \frac{|\mathcal{E}_{\text{intra}}|}{|\mathbf{A}|} = \frac{3}{7}$$

- Node-level Homophily Ratio:

$$h_v = \frac{|\mathcal{E}_v \cap \mathcal{E}_{\text{intra}}|}{|\mathcal{E}_v|}$$

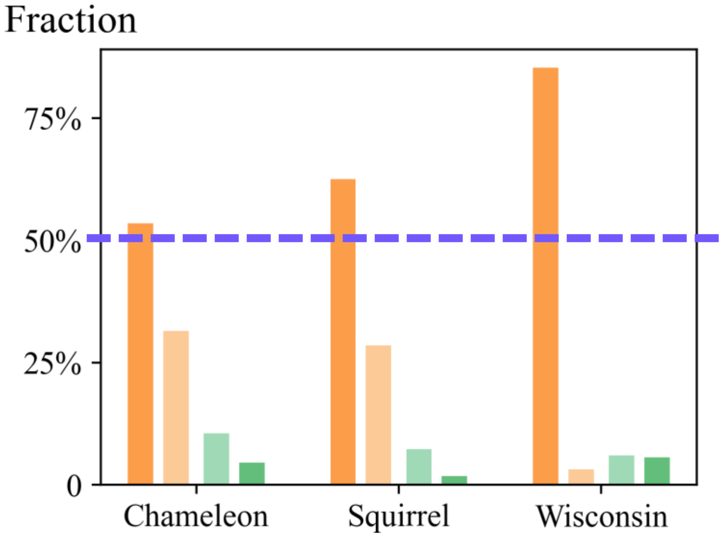
Preliminaries -- Categorization of Graph Benchmarks

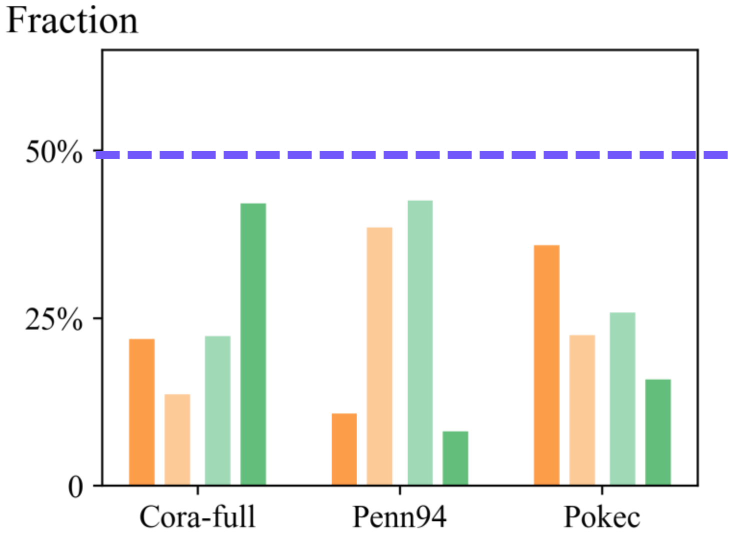
Node-level Homophily Ratio



$$h_v = \frac{|\mathcal{E}_v \cap \mathcal{E}_{\text{intra}}|}{|\mathcal{E}_v|}$$

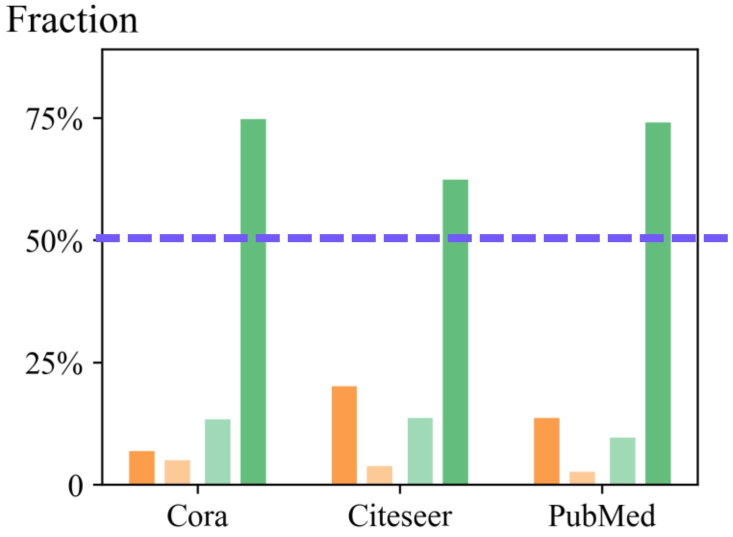
Heterophilous Graph





Semi-Homophilous Graph

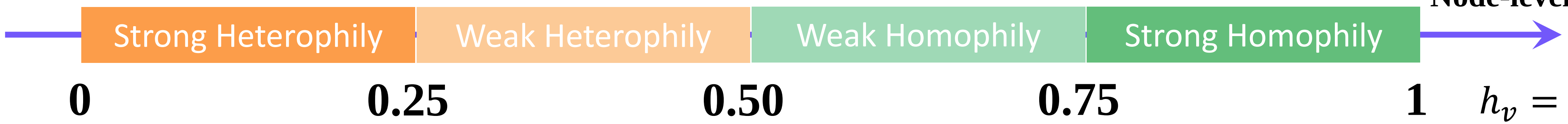
Homophilous Graph



Lim, Derek, et al. "Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods." NeurIPS 2021.

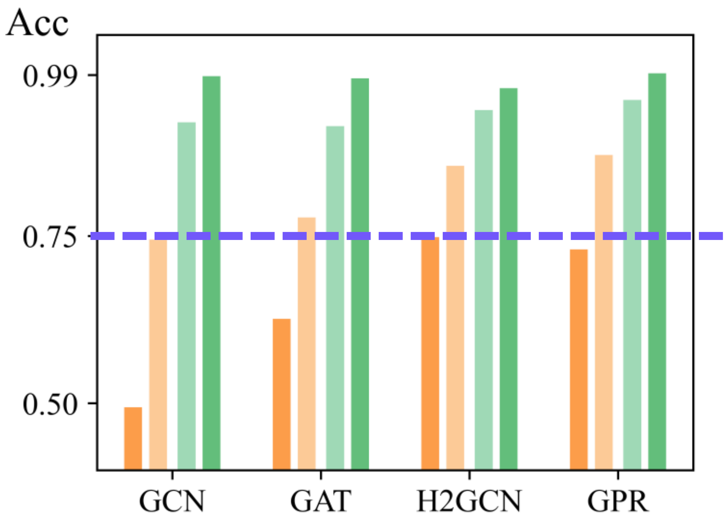
Motivation

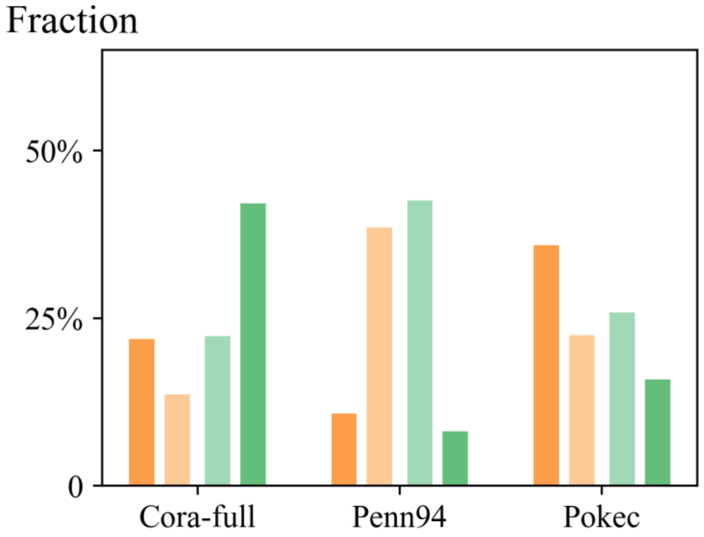
Node-level Homophily Ratio



$$h_v = \frac{|\mathcal{E}_v \cap \mathcal{E}_{\text{intra}}|}{|\mathcal{E}_v|}$$

Group-wise Performance





Semi-Homophilous Graph

- For strong homophilous nodes, the accuracy is close to 0.99.
- For strong heterophilous nodes, the accuracy ranges from 0.48 to 0.74.
- For the four GNNs, the performance gap exists with a range from 0.25 to 0.5.

Content

- **Background and Motivation**

- **Method – UD-GNN**
  - Observation and Overview
  - Uncertainty Estimation
  - Debiased Training

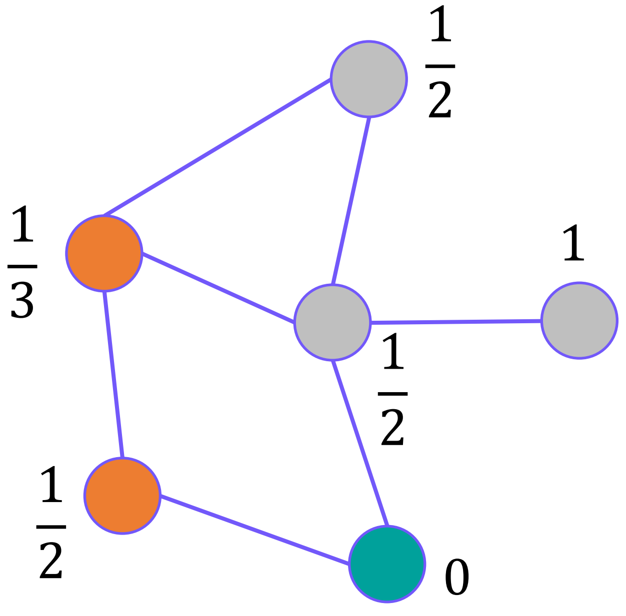
- Experiment

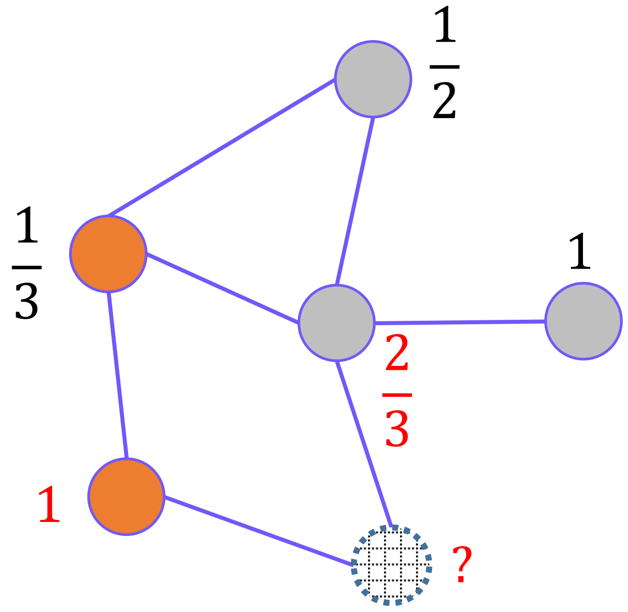
- Conclusion and Future Work

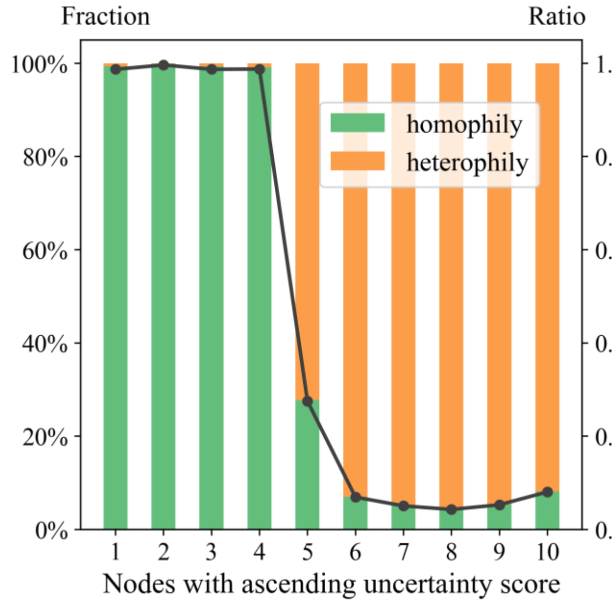
Observation

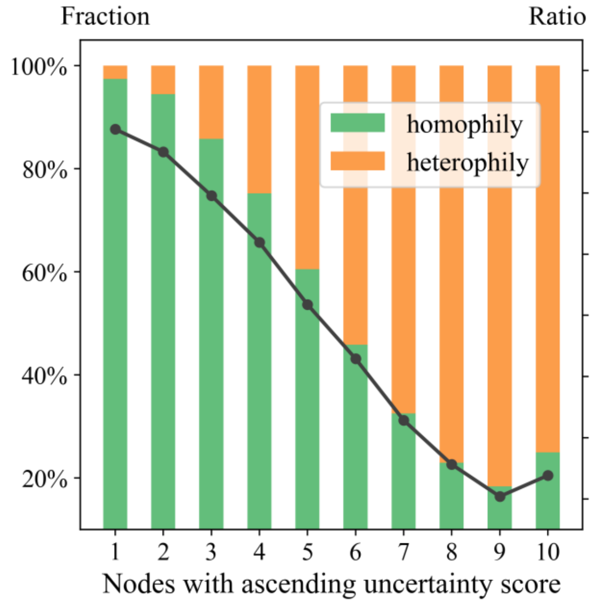
For transductive node classification

- ➢ The node-level homophily ratio computed from incomplete labels are unreliable
  - Some node labels are unavailable during the training phase









- ➢ The output of GNNs exhibit high uncertainty for heterophilous nodes
  - The output uncertainty may help to identify heterophilous nodes

Method - Overview

➢ **UD-GNN: U**ncertainty-aware **D**ebiased **G**raph **N**eural **N**etwork

③ Debiased Training

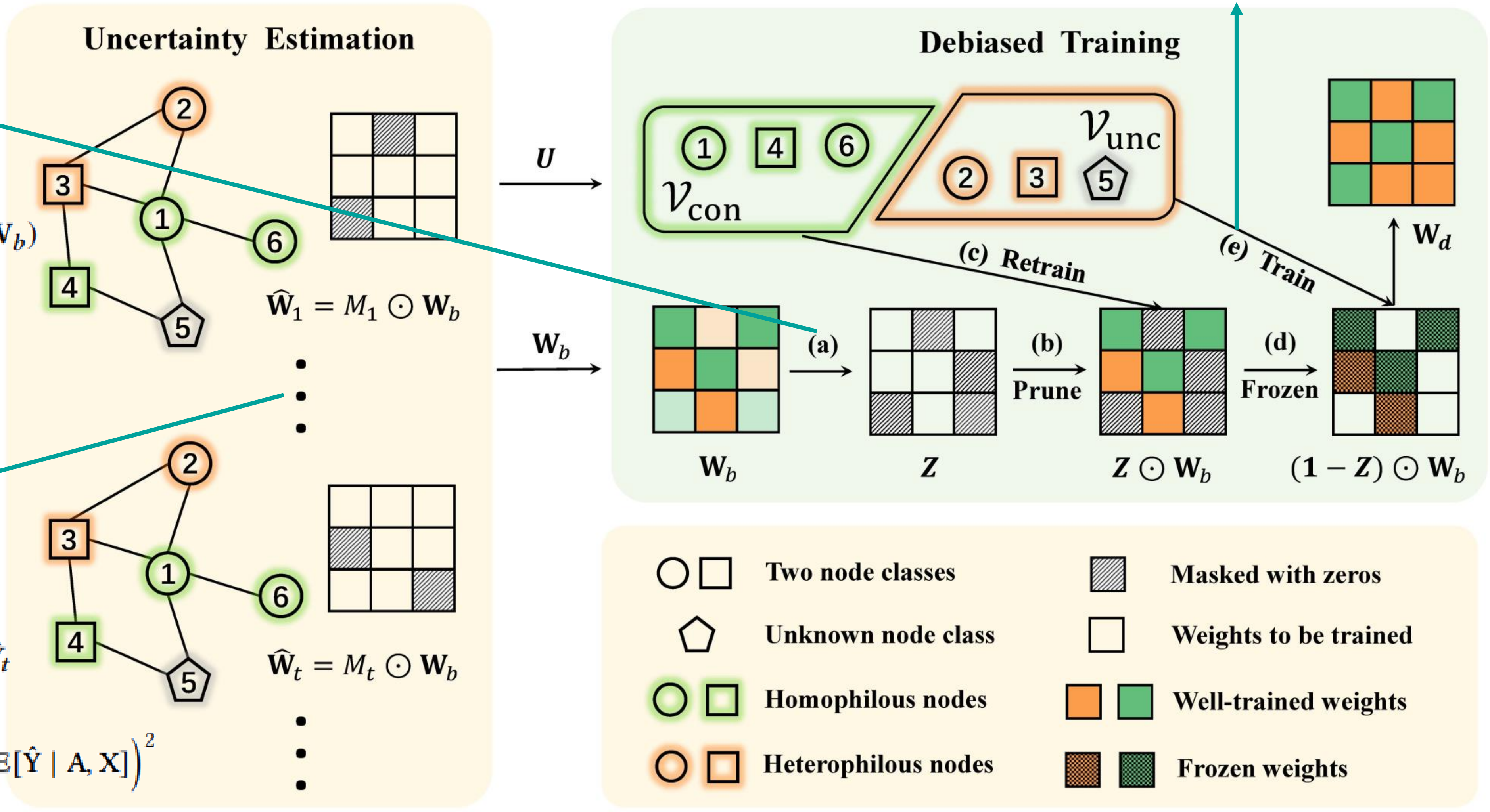
② Mask parameters for $\mathcal{V}_{\text{con}}$

$$\mathbf{W}_d = (1 - Z) \odot \mathbf{W}_b + \text{stop\_gradient}(Z \odot \mathbf{W}_b)$$

① Uncertainty Estimation

$$\mathbb{E}[\hat{\mathrm{Y}} \mid \mathrm{A}, \mathrm{X}] = \int \hat{\mathrm{Y}} p(\hat{\mathrm{Y}} \mid \mathrm{A}, \mathrm{X}) d\hat{\mathrm{Y}} \approx \frac{1}{T} \sum_{t=1}^{T} \hat{\mathrm{Y}}_t$$

$$U(\hat{\mathrm{Y}} \mid \mathrm{A}, \mathrm{X}) = \mathrm{Var}(\hat{\mathrm{Y}} \mid \mathrm{A}, \mathrm{X}) \approx \frac{1}{T} \sum_{t=1}^{T} \left(\hat{\mathrm{Y}}_t - \mathbb{E}[\hat{\mathrm{Y}} \mid \mathrm{A}, \mathrm{X}] \right)^2$$



➢ Uncertainty Estimation

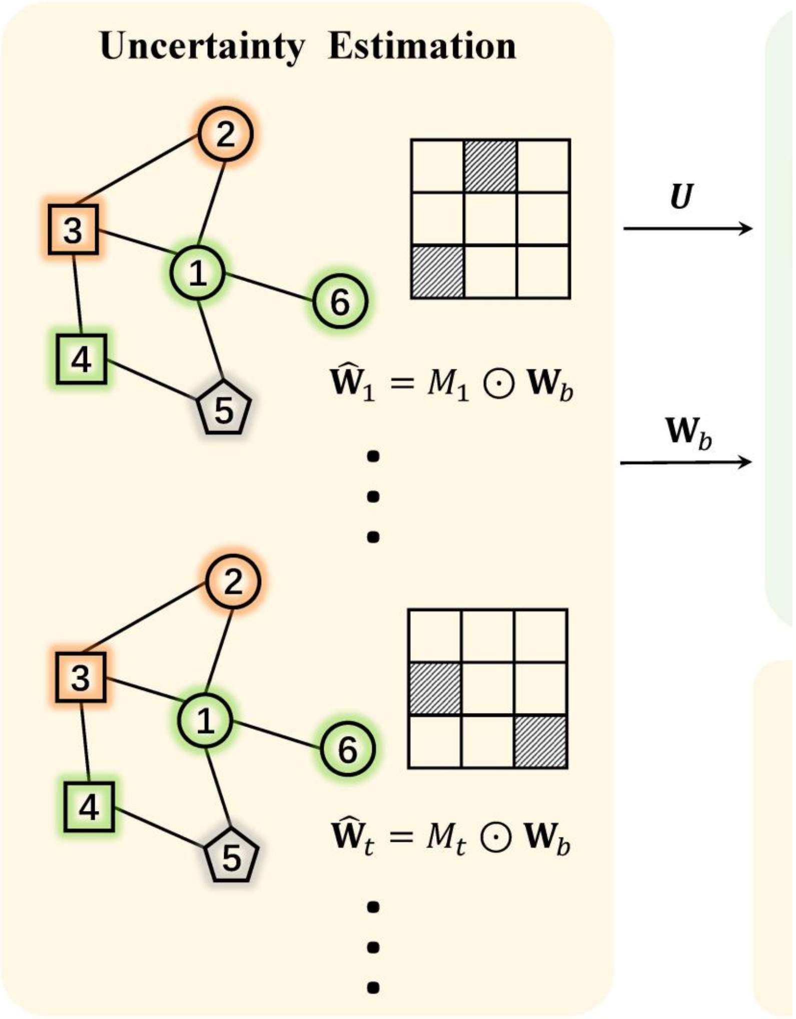
- Monte Carlo dropout variational inference
- Estimated from $\{\widehat{\mathbf{W}}_t\}_{t=1}^{T}$ GNN predictors

$$\mathcal{L}(\mathbf{W}_b) = -\frac{1}{T}\sum_{t=1}^{T} \mathbf{Y}\log(f_{\widehat{\mathbf{W}}_t}(\mathbf{A},\mathbf{X})) + \frac{1-\theta}{2T}\|\mathbf{W}_b\|^2$$

$$\hat{Y}_t = f_{\widehat{\mathbf{W}}_t}(\mathbf{A},\mathbf{X})$$

$$\mathbb{E}[\widehat{\mathbf{Y}}|\mathbf{A},\mathbf{X}] = \int \widehat{\mathbf{Y}}\, p(\widehat{\mathbf{Y}}|\mathbf{A},\mathbf{X})\mathrm{d}\widehat{\mathbf{Y}} \approx \frac{1}{T}\sum_{t=1}^{T}\hat{Y}_t$$

$$U[\widehat{\mathbf{Y}}|\mathbf{A},\mathbf{X}] = \mathrm{Var}(\widehat{\mathbf{Y}}|\mathbf{A},\mathbf{X}) \approx \frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - \mathbb{E}[\widehat{\mathbf{Y}}|\mathbf{A},\mathbf{X}])^2$$

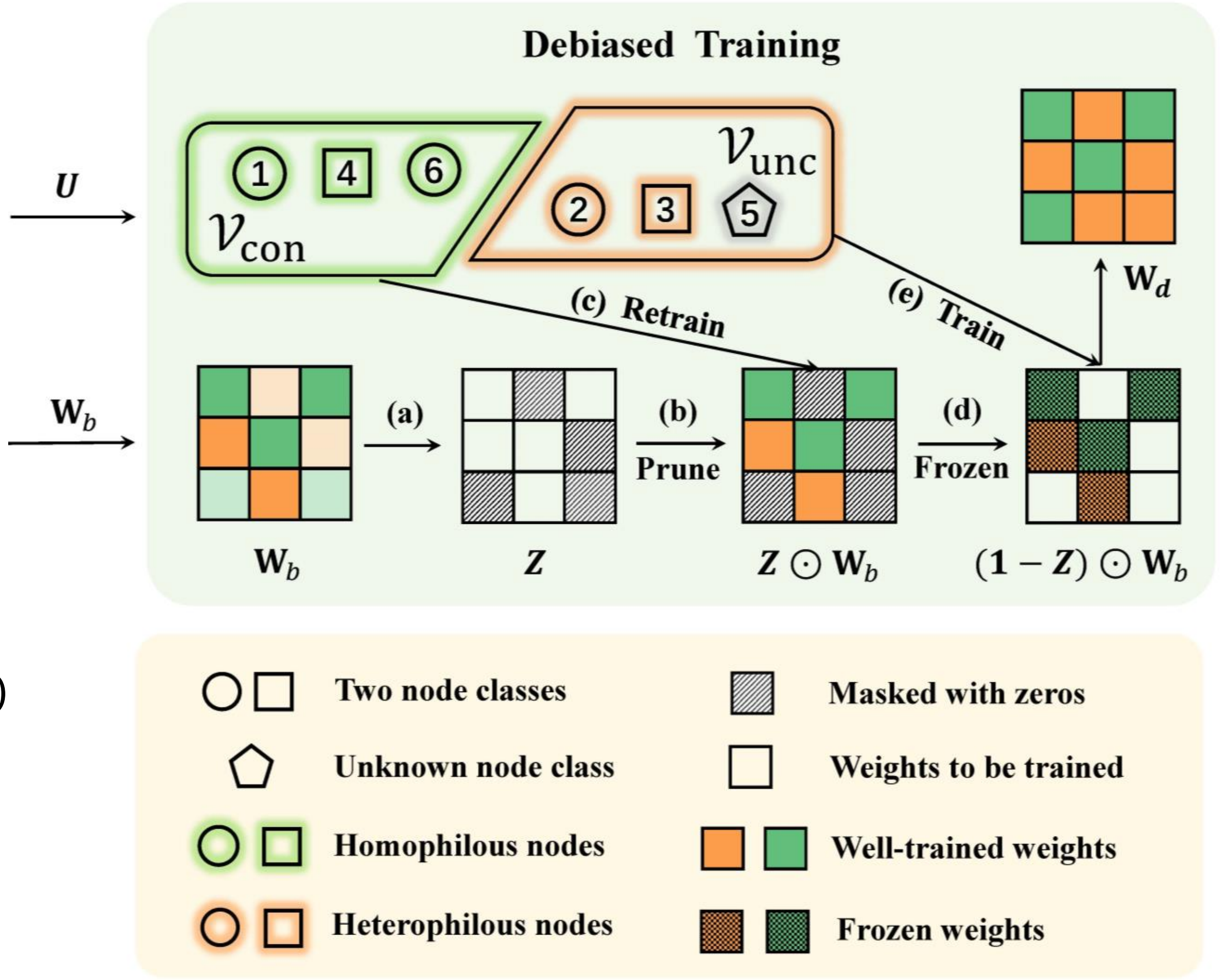


➢ Debiased Training

- Divide $\mathcal{V}_{\text{train}}$ into $\mathcal{V}_{\text{con}}$ and $\mathcal{V}_{\text{unc}}$ according to U such that debiasing ratio $\gamma = \frac{|\mathcal{V}_{\text{con}}|}{|\mathcal{V}_{\text{unc}}|}$
- Prune the parameters close to zero in $\mathbf{W}_b$ with 0-1 mask Z and retrain the remained parameters with $\mathcal{V}_{\text{con}}$
- Freeze $Z \odot \mathbf{W}_b$ and train $(1-Z) \odot \mathbf{W}_b$ with $\mathcal{V}_{\text{unc}}$ to obtain $\mathbf{W}_d$

$$\mathbf{W}_d = (1-Z) \odot \mathbf{W}_b + \text{stop\_gradient}(Z \odot \mathbf{W}_b)$$

$$\mathcal{L}(\mathbf{W}_d) = -\frac{1}{|\mathcal{V}_{\text{unc}}|} \sum_{v \in \mathcal{V}_{\text{unc}}} y_v \log(f_{\mathbf{W}_d}(\mathbf{A}, \mathbf{X}_{v \cup \mathcal{N}_v}))$$



Content

- Background and Motivation
- Method – UD-GNN
- **Experiment**
  - RQ1: Does UD-GNN outperform the state-of-the-art methods on semi-homophilous graphs?
  - RQ2: How do the key components contribute to the results?
  - RQ3: Does UD-GNN work well on heterophilous graphs?
  - RQ4: What is the sensitivity of UD-GNN with respect to different debiasing ratios, mixing ratios and number of classes?
- Conclusion and Future Work

Data

➢ Public benchmark

- **cSBM**: contextual stochastic block models
- **Penn94**: a friendship network from the Facebook100 networks
- **Cora-full:** a citation network labeled on the paper topic
- **Ogbn-arxiv:** the citation network between all Computer Science (CS) arXiv papers indexed by MAG

Dataset	#Node	#Edge	#Class	#Feat	%Heter
cSBM	20,000	998,766	10	1,024	50%
Penn94	41,554	1,362,229	2	4,814	49%
Cora-full	19,793	126,842	70	8,710	44%
Ogbn-arxiv	169,343	1,166,243	40	128	37%

➢ Train/Valid/Test:

- **cSBM**: 40%/20%/40%
- **Penn94**: 80%/10%/10%
- **Cora-full:** 70%/10%/20%
- **Ogbn-arxiv:** 2017/2018/2019

Experiment

- Compared methods
  - **GCN, GAT:** Traditional graph convolutional network and graph attention network
  - **Mixhop:** Repeatedly mixing feature representations of neighbors at various distances
  - **GPR-GNN:** Generalized PageRank GNN
  - **JK-Net:** Jump Knowledge
  - **H2GCN, CPGNN:** GNNs for heterophilous graphs
  - **WRGAT**: Improving the assortativity of graphs with local mixing patterns
  - **U-GNN**: Universal GCN extracting information from 1-hop, 2-hop and kNN networks
- Metrics
  - **Accuracy:** $\mathrm{eval}(\mathbf{Y}, f_{\mathbf{W}}(\mathbf{A}, \mathbf{X}))$
  - **Relative bias:** $\delta = \mathrm{eval}(\mathbf{Y}, f_{\mathbf{W}^*}(\mathbf{A}, \mathbf{X})) - \mathrm{eval}(\mathbf{Y}, f_{\mathbf{W}}(\mathbf{A}, \mathbf{X}))$

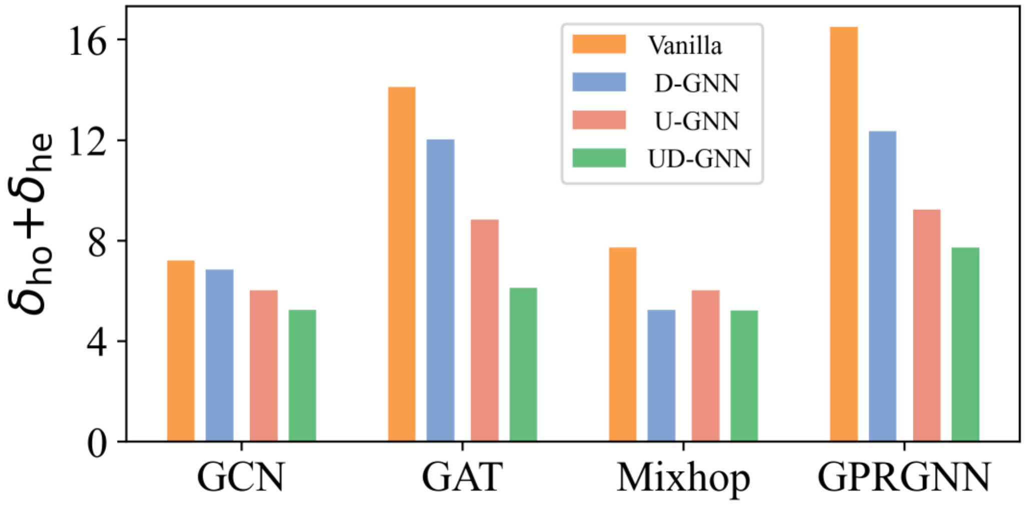
Experiment

➢RQ1: Does UD-GNN outperform the state-of-the-art methods on semi-homophilous graphs?

- UD-GNN achieves the best accuracy with the lowest relative bias

Dataset		cSBM			Penn94			Cora-full			Ogbn-arxiv		
Metric		Acc ↑	δ_{ho} ↓	δ_{he} ↓	Acc ↑	δ_{ho} ↓	δ_{he} ↓	Acc ↑	δ_{ho} ↓	δ_{he} ↓	Acc ↑	δ_{ho} ↓	δ_{he} ↓
GCN	VA	47.30±0.43	1.23	9.11	82.06±0.19	3.89	3.32	68.81±0.29	2.8	3.11	71.17±0.11	**0.03**	**7.76**
	UD	49.39±0.16	2.31	4.42	82.96±0.21	3.93	1.31	69.70±0.24	3.51	1.12	72.06±0.17	**1.88**	**0.28**
GAT	VA	50.44±0.29	1.68	8.27	79.85±0.73	5.48	8.63	69.24±0.27	1.82	3.27	69.90±0.12	0.43	6.24
	UD	52.66±0.31	2.04	3.12	83.54±0.20	5.24	0.89	70.85±0.34	2.21	1.18	71.07±0.35	1.65	0.65
Mixhop	VA	51.31±0.41	0.93	8.25	82.14±0.21	2.37	5.36	69.11±0.31	2.26	3.94	72.80±0.23	0.7	5.63
	UD	52.51±0.17	1.84	3.17	**83.68±0.65**	3.22	2.01	70.01±0.34	2.94	2.67	**73.20±0.09**	2.13	0.2
GPRGNN	VA	53.79±0.15	**1.29**	**9.23**	76.77±0.25	**5.58**	**10.92**	70.15±0.30	**1.61**	**4.92**	70.98±0.17	0.95	5.84
	UD	**54.80±0.20**	**1.62**	**2.48**	80.70±0.24	**6.82**	**0.91**	**71.09±0.30**	**1.65**	**2.36**	71.32±0.13	2.48	1.04
JK-NET		50.68±0.38	2.67	6.21	81.26±0.23	3.35	8.24	68.12±0.23	2.34	3.23	70.66±0.19	0.72	8.23
H2GCN		51.93±0.25	2.46	8.26	81.63±0.16	4.16	7.12	70.82±0.82	3.35	3.67	70.14±0.00	0.28	6.82
CPGNN		51.84±0.67	2.03	5.26	80.92±0.67	4.12	8.21	70.38±0.23	2.35	3.26	69.24±0.52	0.52	6.27
WRGAT		52.48±0.28	1.89	6.26	82.32±0.83	3.12	6.23	71.32±0.92	2.92	4.38	71.23±0.67	0.46	6.23
U-GCN		52.74±0.72	1.73	8.23	82.31±0.89	5.23	4.21	70.47±0.78	1.92	4.21	70.27±0.81	0.61	5.98

- RQ2: How do the key components contribute to the results?
  - D-GNN removes uncertainty estimation and trains a separate model to discriminate homophilous nodes from heterophilous nodes for debiasing.
  - U-GNN removes debiased training and applies Focal loss based on the estimated uncertainty scores.



➢RQ3: Does UD-GNN work well on heterophilous graphs?

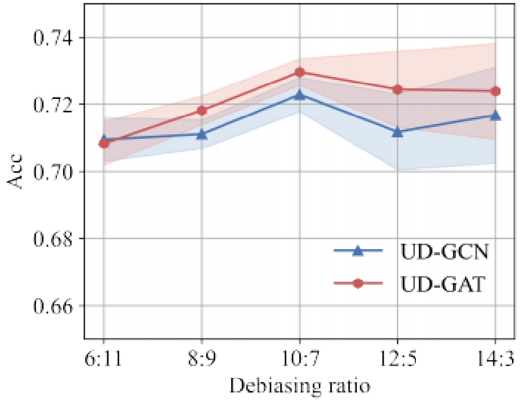
- UD-GNN improves the performance due to the refining of uncertain nodes in the debiased training.
- UD-GNN achieves the best results on Chameleon and Squirrel, with comparable performance on Wisconsin.

Dataset	#Node	#Edge	#Class	#Feat
Chameleon	2,277	36,101	5	2,325
Squirrel	5,201	217,073	5	2,089
Wisconsin	251	515	5	1,703

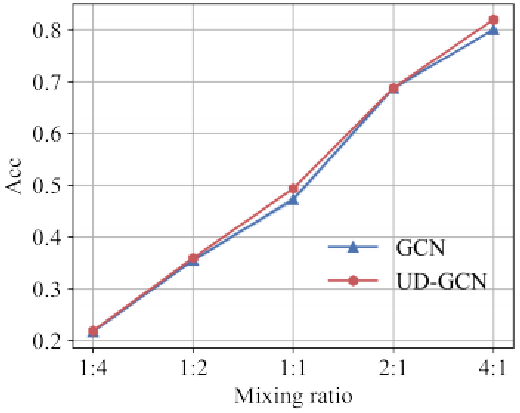
Dataset		Chameleon	Squirrel	Wisconsin
Mixhop	VA	58.25±1.83	42.86±1.48	73.83±6.82
	UD	59.23±1.24	43.92±1.42	74.23±5.92
GPRGNN	VA	64.36±0.87	46.83±0.84	79.23±3.81
	UD	66.23±1.03	47.92±0.25	81.29±2.34
JK-NET		53.95±1.14	33.51±1.32	48.39±5.28
H2GCN		57.39±1.96	35.23±1.35	85.88±4.92
CPGNN		59.11±1.23	36.27±1.29	86.29±4.28
WRGAT		63.26±1.67	41.26±1.37	86.28±2.46
U-GCN		54.07±1.57	34.39±1.34	69.89±2.54

Experiment

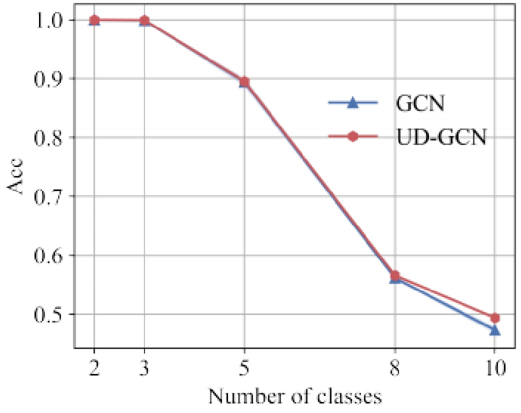
➢RQ4: What is the sensitivity of UD-GNN with respect to different debiasing ratios, mixing ratios and number of classes?



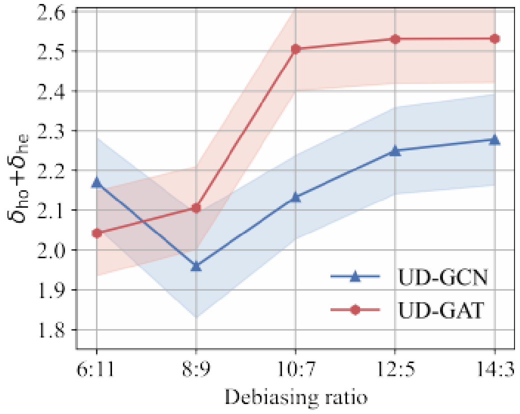
(a) Acc on ogbn-arxiv with different debiasing ratios



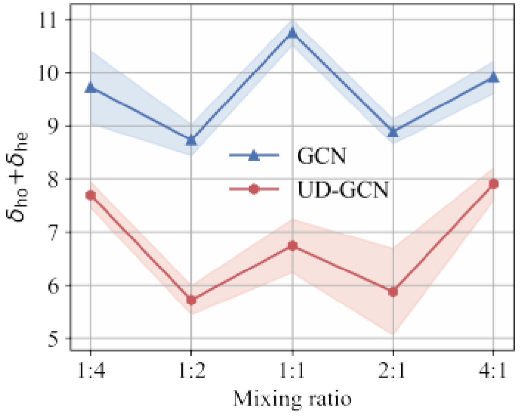
(b) Acc on cSBM with different mixing ratios



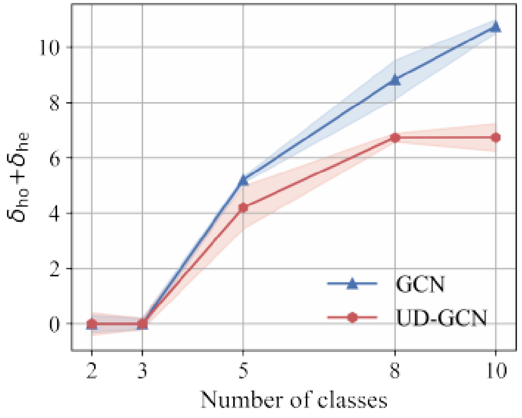
(c) Acc on cSBM with different classes



(d) δ on ogbn-arxiv with different debiasing ratios



(e) δ on cSBM with different mixing ratios



(f) δ on cSBM with different classes

Content

- Background and Motivation
- Method – UD-GNN
- Experiment
- **Conclusion and Future Work**

Conclusion and Future work

➢Conclusion

- We investigate the bias issue between homophily and heterophily on semi-homophilous graphs.
- We propose an Uncertainty-aware Debiasing framework to mitigate the bias.
- Experiments on four benchmark semi-homophilous graph datasets demonstrate the effectiveness of the proposed framework.

➢Future Work

- New message passing architecture for semi-homophilous graphs
- Spectral filter for semi-homophilous graphs

Acknowledgment

Thanks for listening!

If you have any question, feel free to contact us at
liuyang520ict@gmail.com
aoxiang@ict.ac.cn

Paper and slides are available at
https://ponderly.github.io/