Few-shot Node Classification with Extremely Weak Supervision

Song Wang
University of Virginia
sw3wv@virginia.edu

Yushun Dong
University of Virginia
yd6eb@virginia.edu

Kaize Ding
Arizona State University
kding9@asu.edu

Chen Chen
University of Virginia
zrh6du@virginia.edu

Jundong Li
University of Virginia
jundong@virginia.edu

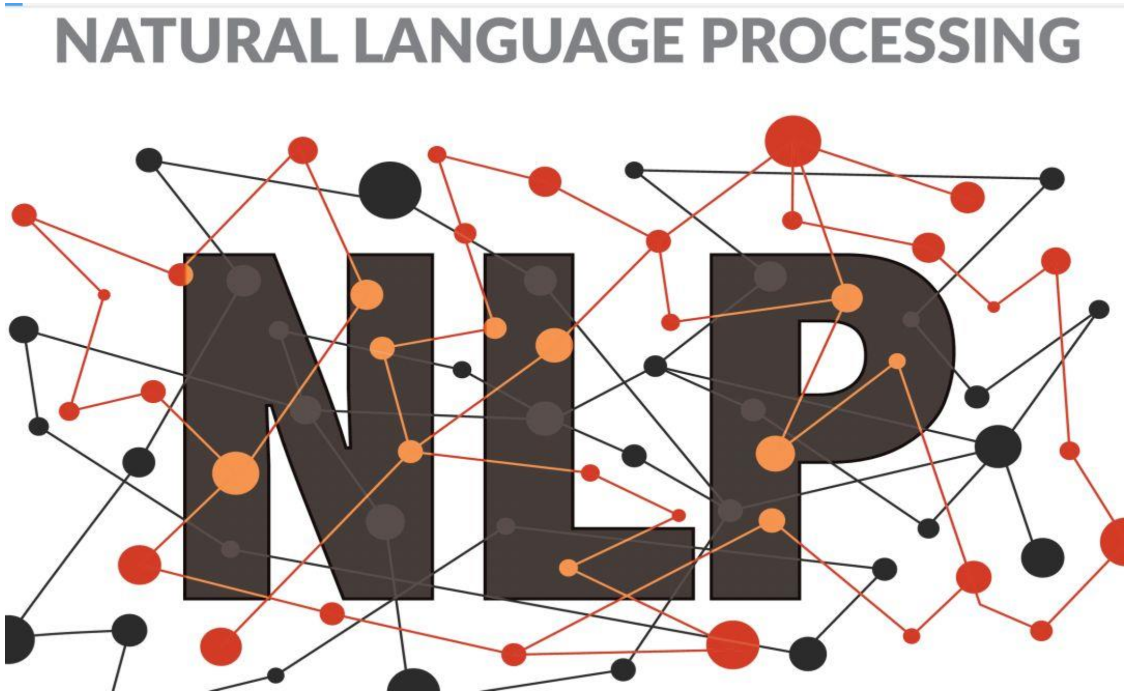
code: https://github.com/SongW-SW/X-FNC

2023. 2. 19 • ChongQing

2023_WSDM

gesis
Leibniz-Institut
für Sozialwissenschaften

Reported by Junhao Cao



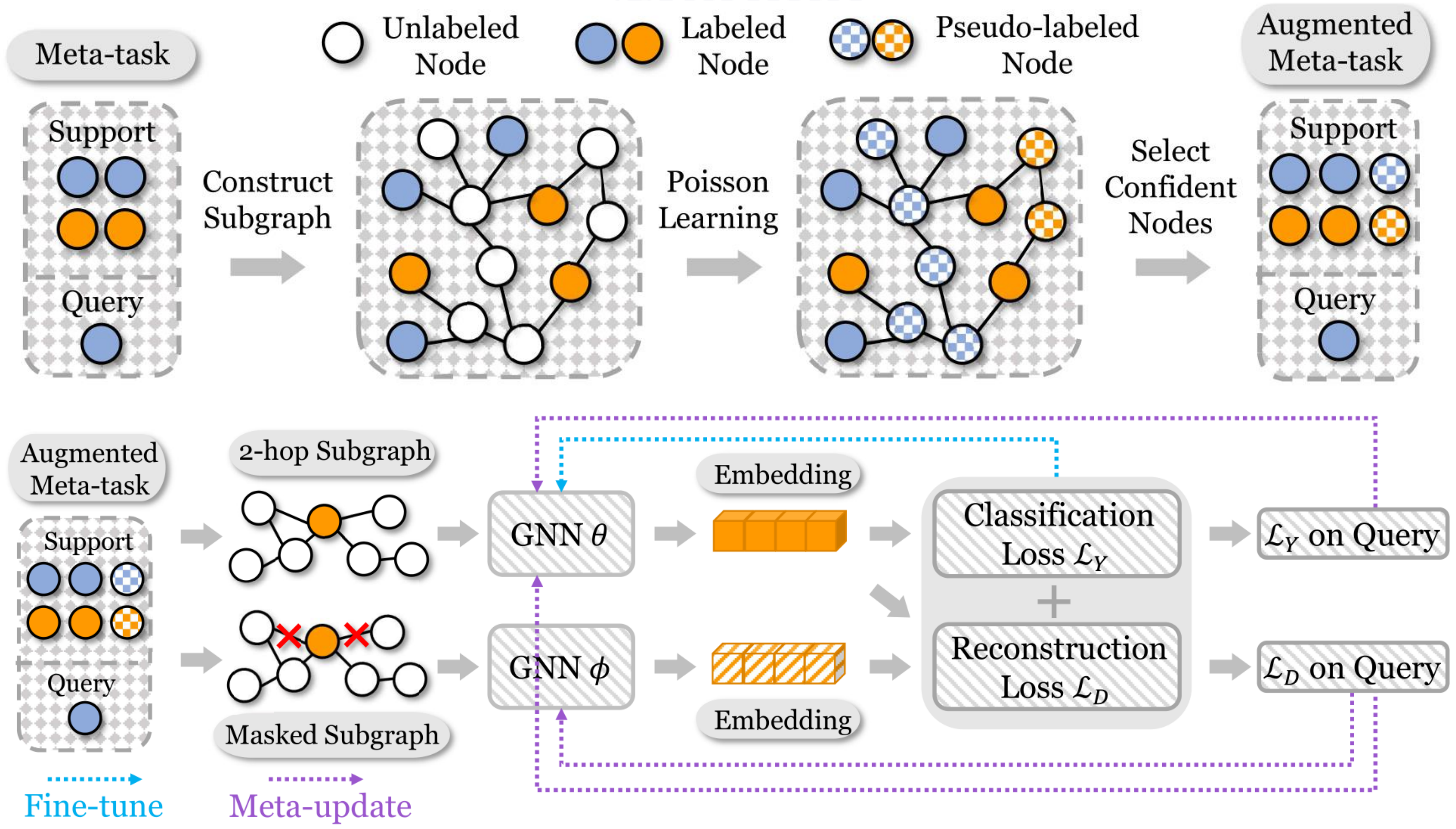
1.Introduction
2.Method
3.Experiments

Introduction

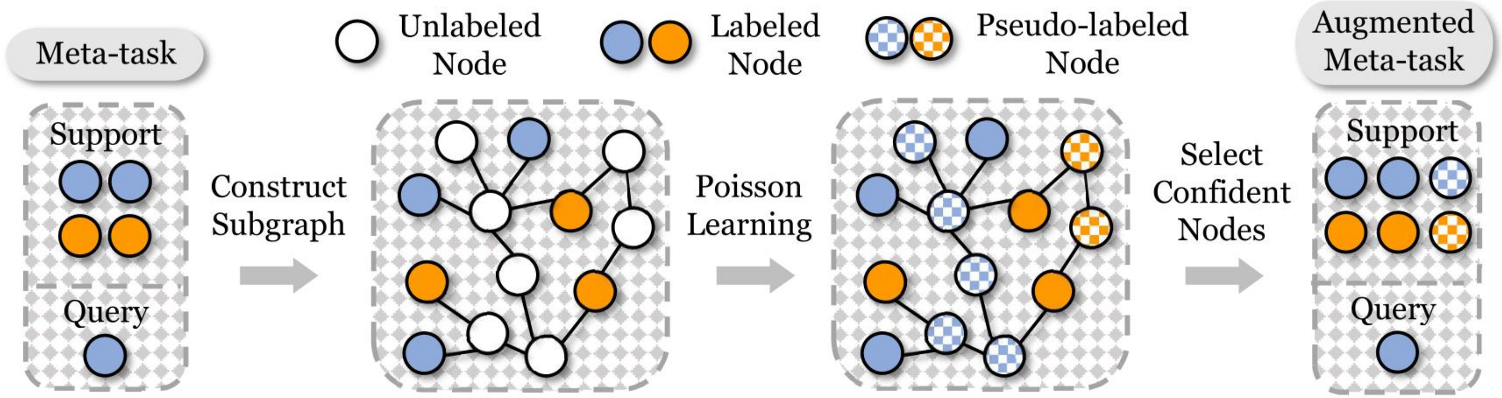
scarce labeled nodes---under-generalizing--- obtain abundant pseudo-labeled nodes based on Poisson Learning

inadequate query nodes---over-fitting---optimize the model by filtering out irrelevant information based on Information Bottleneck (IB)

Method


Meta-task
Unlabeled Node
Labeled Node
Pseudo-labeled Node
Augmented Meta-task
Support
Query
Construct Subgraph
Poisson Learning
Select Confident Nodes
Support
Query
Augmented Meta-task
Support
Query
2-hop Subgraph
Masked Subgraph
GNN θ
GNN ϕ
Embedding
Embedding
Classification Loss $\mathcal{L}_Y$
+
Reconstruction Loss $\mathcal{L}_D$
$\mathcal{L}_Y$ on Query
$\mathcal{L}_D$ on Query
Fine-tune
Meta-update

Method



Poisson Label Propagation

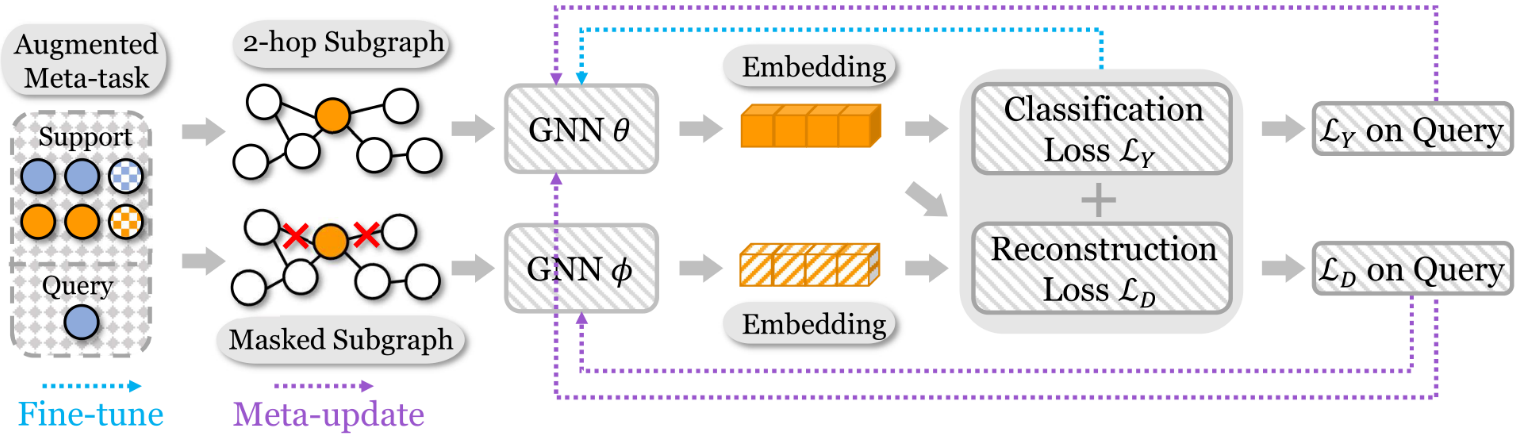
$$\mathbf{A}''_{ij} = \exp\left(-\eta\|\mathbf{x}_i - \mathbf{x}_j\|\right), \tag{1}$$

$$\begin{cases} \sum_{j=1}^{|\mathcal{V}_s|} \mathbf{A}_{ij}\left(\mathbf{u}_i - \mathbf{u}_j\right) = 0, & \text{if } NK+1 \leq i \leq |\mathcal{V}_s|, \\ \mathbf{u}_i = \mathbf{y}_i - \bar{\mathbf{y}}, & \text{if } 1 \leq i \leq NK, \end{cases}, \tag{2}$$

$$\mathbf{U}^{(t)} \leftarrow \mathbf{U}^{(t-1)} + \mathbf{D}^{-1}\left(\mathbf{B}^{\top} - \mathbf{L}\mathbf{U}^{(t-1)}\right), \tag{3}$$

$$c_i = -\sum_{j=1}^{N} u_{ij} \log u_{ij}, \tag{4}$$

Method



Information Bottleneck Fine-tuning

$$\min \mathrm{IB}(D, Y; Z) \triangleq [-I(Y;Z) + \beta I(D;Z)], \tag{5}$$

$$-I(Y;Z) = -\left[H(Y) - H(Y|Z)\right], \tag{6}$$

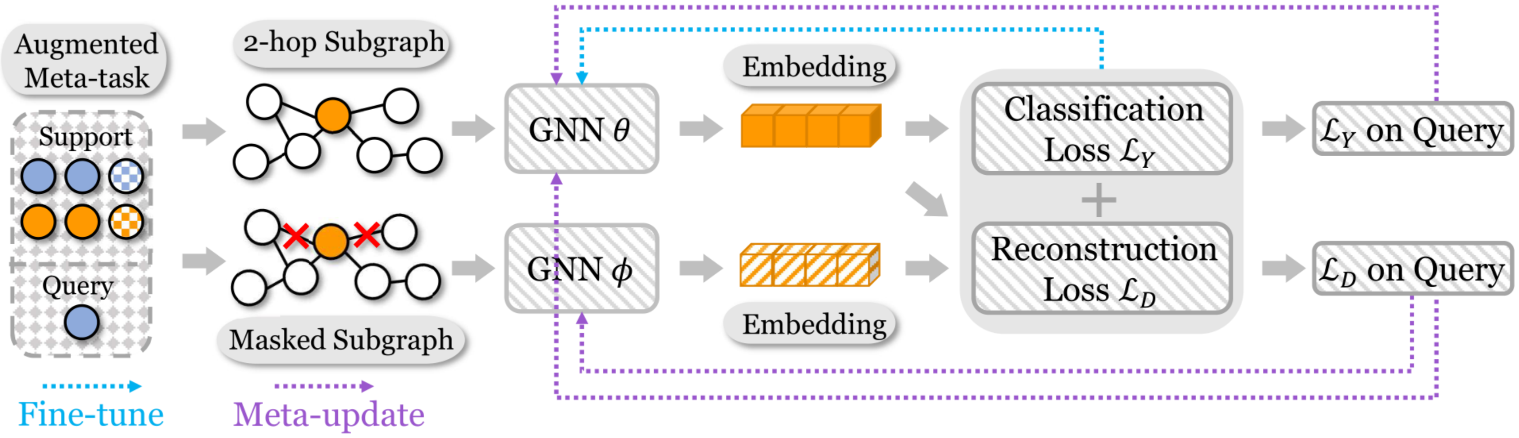
$$\begin{aligned} H(Y|Z) &= -\sum_{i=1}^{N}\sum_{j=1}^{|\widetilde{S}|} p(y_i, z_j)\log\frac{p(y_i, z_j)}{p(z_j)} \\ &= -\sum_{i=1}^{N}\sum_{j=1}^{|\widetilde{S}|} p(z_j|y_i)p(y_i)\log p(y_i|z_j), \end{aligned} \tag{7}$$

$$-I(Y;Z) \rightarrow \mathcal{L}_Y = -\sum_{i=1}^{\widetilde{S}} \log p(y_i'|z_i), \tag{8}$$

$$\mathbf{s}_i = \mathrm{MLP}_\theta\left(\mathrm{GNN}_\theta(\mathbf{A}_i, \mathbf{X}_i)\right), \tag{9}$$

$$I(D;Z) = \mathbb{E}\left(\log\frac{p(Z|D)}{p(Z)}\right), \tag{10}$$

Method



Information Bottleneck Fine-tuning

$$\mathbb{E}\left(\log\left(\frac{p(Z|D)}{q(Z)}\frac{q(Z)}{p(Z)}\right)\right) = \mathbb{E}\left(\log\frac{p(Z|D)}{q(Z)}\right) - \mathrm{KL}\left(p(Z)||q(Z)\right)$$
$$\le \mathbb{E}\left(\log\frac{p(Z|D)}{q(Z)}\right), \tag{11}$$

$$\mathbf{h}_i = \mathrm{GNN}_\theta(\mathbf{A}_i,\mathbf{X}_i), \quad \widetilde{\mathbf{h}}_i = \mathrm{GNN}_\phi(\widetilde{\mathbf{A}}_i,\widetilde{\mathbf{X}}_i), \tag{12}$$

$$\mathrm{MSE}(p_\theta(\mathbf{h}_i),\widetilde{\mathbf{h}}_i) = \left\|\frac{p_\theta(\mathbf{h}_i)}{\|p_\theta(\mathbf{h}_i)\|} - \frac{\widetilde{\mathbf{h}}_i}{\|\widetilde{\mathbf{h}}_i\|}\right\|^2 = 2 - 2\cdot\frac{p_\theta(\mathbf{h}_i)\cdot\widetilde{\mathbf{h}}_i}{\|p_\theta(\mathbf{h}_i)\|\|\widetilde{\mathbf{h}}_i\|}, \tag{13}$$

$$I(D;Z) \rightarrow \mathcal{L}_D = -\sum_{i=1}^{\widetilde{S}}\frac{p_\theta(\mathbf{h}_i)\cdot\widetilde{\mathbf{h}}_i}{\|p_\theta(\mathbf{h}_i)\|\|\widetilde{\mathbf{h}}_i\|}, \tag{14}$$

Meta Learning-based Optimization

$$\theta_t \leftarrow \theta_{t-1} - \alpha\nabla_{\theta_{t-1}}\mathcal{L}\left(\widetilde{S};\theta_{t-1}\right), \tag{15}$$

$$\theta =: \theta - \beta_1\nabla_\theta\mathcal{L}(Q;\theta_T), \quad \phi =: \phi - \beta_2\nabla_\phi\mathcal{L}_D(Q;\theta_T), \tag{16}$$

Experiment

Table 1: The overall few-shot node classification results (accuracy in %) of X-FNC and baselines under different settings.

Dataset	DBLP						Amazon-E					
Setting	5-way 3-shot			10-way 3-shot			5-way 3-shot			10-way 3-shot		
# Labels per Class	5	10	20	5	10	20	5	10	20	5	10	20
PN	49.4±3.2	51.9±3.1	53.3±3.9	36.3±3.8	38.5±2.8	40.2±3.9	51.6±2.3	52.2±2.3	53.8±2.3	36.7±3.0	38.2±2.0	41.3±3.9
MAML	50.9±3.1	51.8±1.8	56.1±2.1	39.4±2.3	44.3±2.0	45.4±3.1	48.8±2.4	49.4±3.3	53.9±2.7	39.0±3.2	40.3±3.2	41.5±3.2
G-Meta	59.8±3.3	61.8±3.5	63.3±4.1	44.9±2.9	51.0±3.4	52.9±3.6	53.4±2.2	55.7±3.6	56.6±3.2	39.6±4.1	41.9±3.0	45.6±4.3
GPN	58.6±3.8	62.5±2.8	66.9±4.3	50.6±3.9	52.7±2.4	54.6±3.4	56.0±4.1	60.7±4.7	63.0±2.3	42.1±4.8	45.8±3.3	52.1±4.8
RALE	64.7±4.1	66.9±4.7	67.9±4.0	51.3±4.2	55.0±3.2	56.9±4.0	60.4±4.5	64.0±4.8	66.1±4.5	47.8±4.4	48.6±4.8	52.4±3.3
X-FNC	**70.1**±4.0	**75.5**±3.5	**76.8**±3.3	**57.2**±3.4	**63.6**±3.3	**65.8**±3.1	**69.9**±3.9	**72.8**±3.4	**76.0**±4.8	**49.2**±4.1	**51.5**±2.8	**56.3**±3.4

Dataset	Cora-full						ogbn-arxiv					
Setting	5-way 3-shot			10-way 3-shot			5-way 3-shot			10-way 3-shot		
# Labels per Class	5	10	20	5	10	20	50	100	200	50	100	200
PN	45.5±2.7	48.1±3.6	48.9±3.8	28.2±3.8	31.6±3.3	34.4±2.5	39.1±2.5	40.8±3.7	42.6±3.1	23.1±3.6	24.4±3.1	27.7±3.5
MAML	46.9±2.6	48.6±3.0	49.2±2.7	32.7±2.5	33.2±2.3	35.8±1.9	41.0±2.4	41.9±1.9	43.1±3.4	23.2±2.1	25.4±3.1	28.0±3.2
G-Meta	57.7±3.9	58.7±3.6	59.8±2.6	41.7±3.3	42.0±3.0	43.8±2.7	43.5±3.6	44.7±2.9	46.5±4.4	27.4±4.4	29.0±2.8	29.9±2.5
GPN	54.6±2.8	55.2±3.6	57.7±4.2	38.4±2.8	40.2±2.9	42.0±4.5	46.6±3.4	47.1±3.9	48.4±2.9	26.1±2.6	30.9±3.6	33.5±3.5
RALE	58.2±2.8	59.3±4.1	63.1±3.9	38.1±4.2	43.4±2.8	44.0±4.5	49.3±3.0	51.4±3.9	52.5±4.6	30.4±2.5	31.7±3.3	33.9±4.8
X-FNC	**62.9**±4.5	**68.0**±3.7	**69.2**±4.6	**43.7**±4.8	**45.6**±4.4	**47.7**±4.5	**54.6**±2.6	**56.7**±4.0	**58.7**±4.1	**33.3**±3.9	**35.7**±4.4	**39.8**±2.4

Experiment

Table 2: Statistics of four node classification datasets.

Dataset	# Nodes	# Edges	# Features	Class Split
Amazon-E	42,318	43,556	8,669	90/37/40
DBLP	40,672	288,270	7,202	80/27/30
Cora-full	19,793	65,311	8,710	25/20/25
ogbn-arxiv	169,343	1,166,243	128	15/5/20

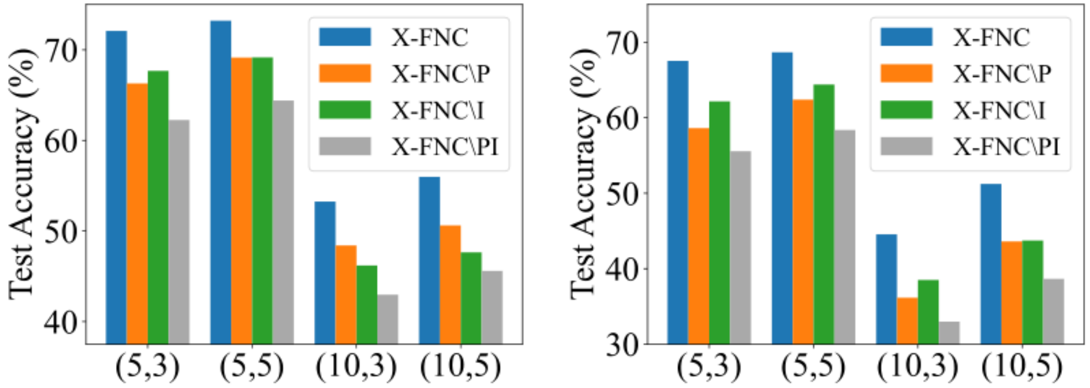


Figure 3: Ablation study on our framework on `Amazon-E` (left) and `Cora-full` (right) in the N-way K-shot setting (N, K).

Experiment

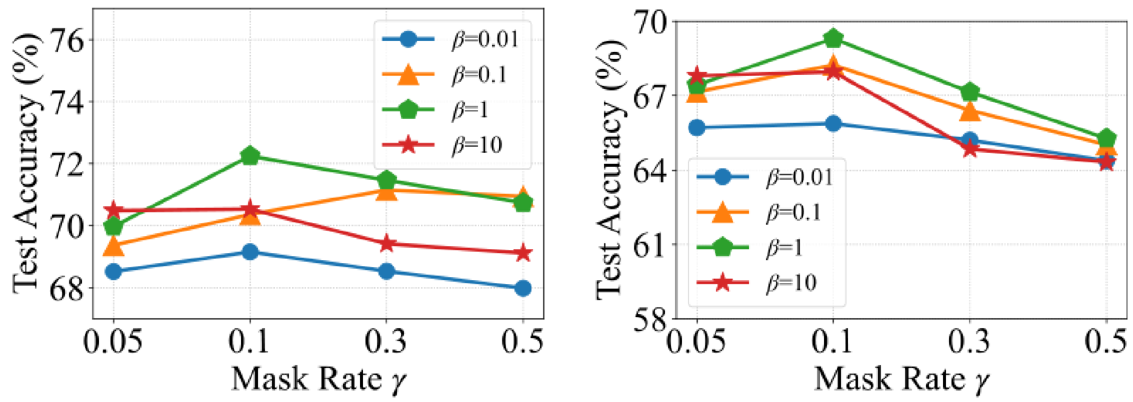


Figure 4: Results of our framework on `Amazon-E` (left) and `Cora-full` (right) with different mask rates.

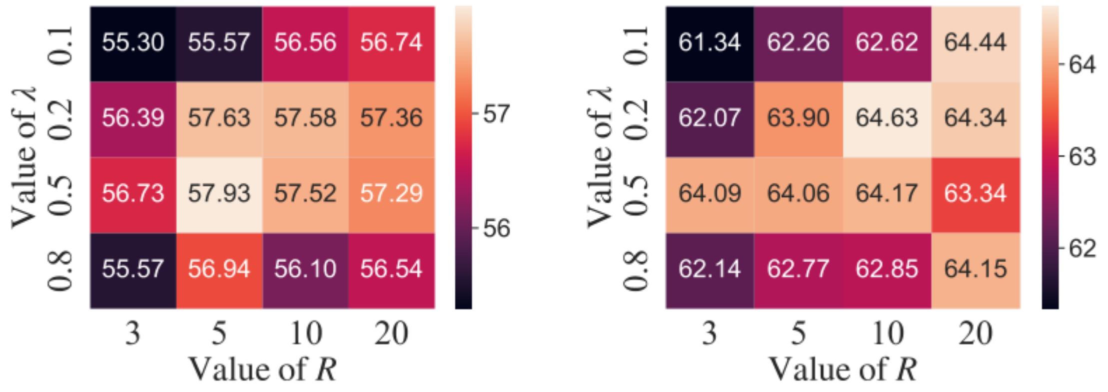


Figure 5: Results of pseudo-labeling accuracy (in %) on `Amazon-E` (left) and `Cora-full` (right) with different λ and R.

Thank you!

gesis
Leibniz-Institut
für Sozialwissenschaften