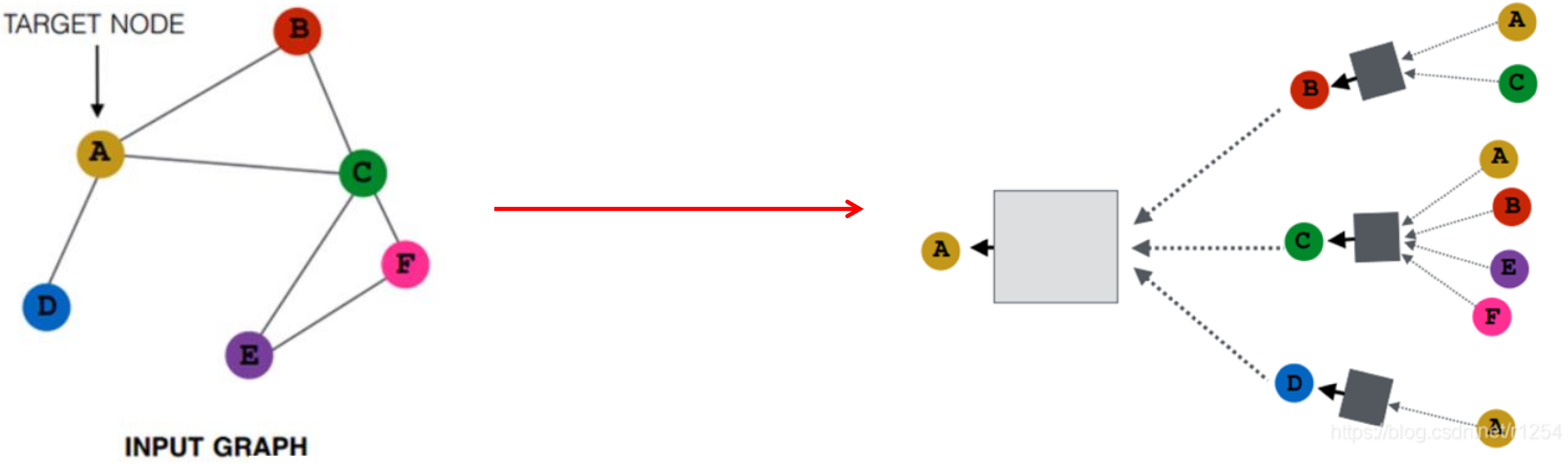
NRGNN: Learning a Label Noise-Resistant Graph Neural Network on Sparsely and Noisily Labeled Graphs

Enyan Dai Charu Aggarwal Suhang Wang

KDD 2021

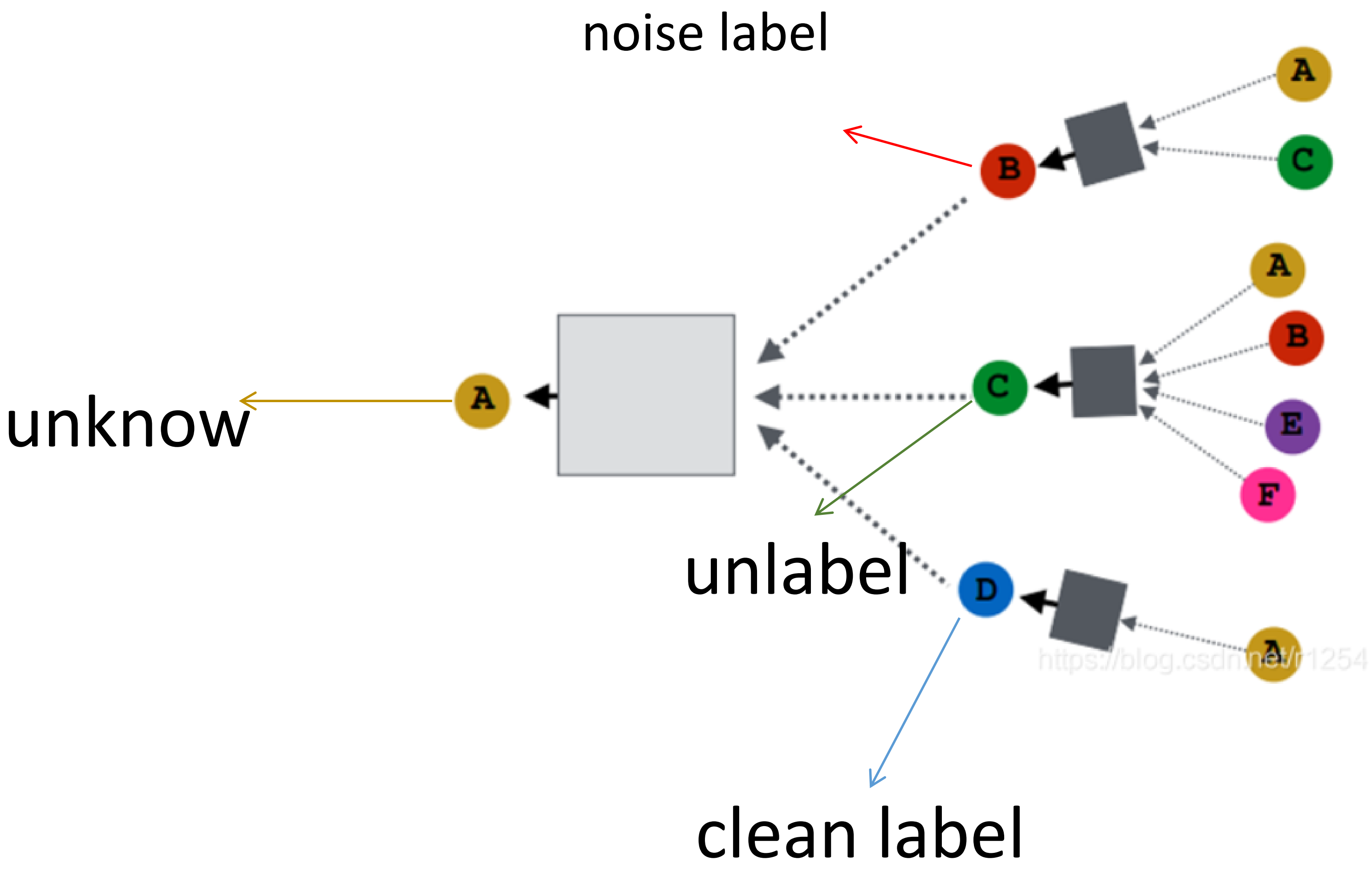
Graph Neural Network

node representations by aggregating the information from their neighbors



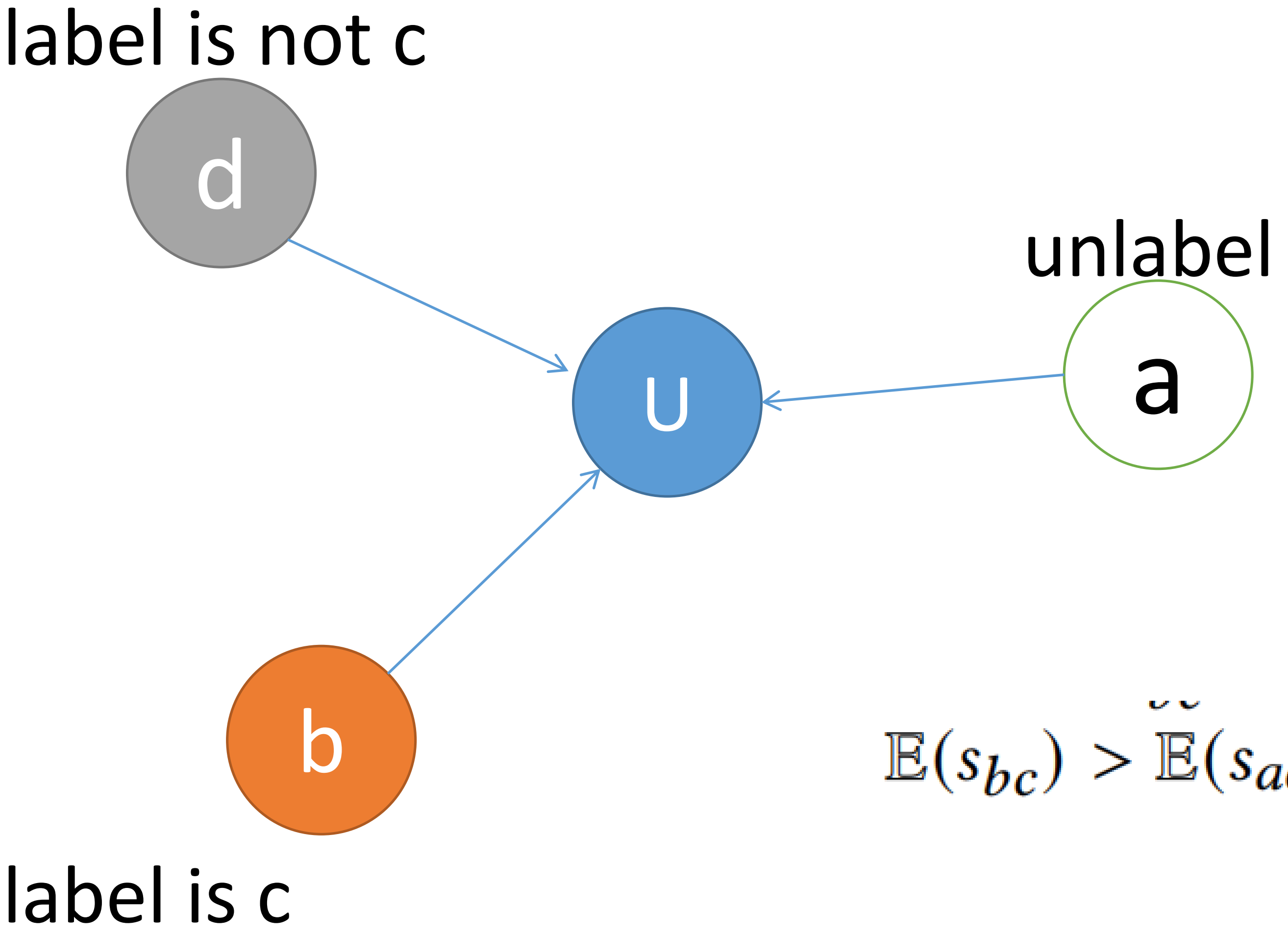
$$\mathbf{a}_v^{(k)} = \mathrm{AGGREGATE}^{(k-1)}(\{\mathbf{h}_u^{(k-1)} : u \in \mathcal{N}(v)\}),$$
$$\mathbf{h}_v^{(k)} = \mathrm{COMBINE}^{(k)}(\mathbf{h}_v^{(k-1)}, \mathbf{a}^{(k)}),$$

Noisy label and sparse label will affect Node Classification



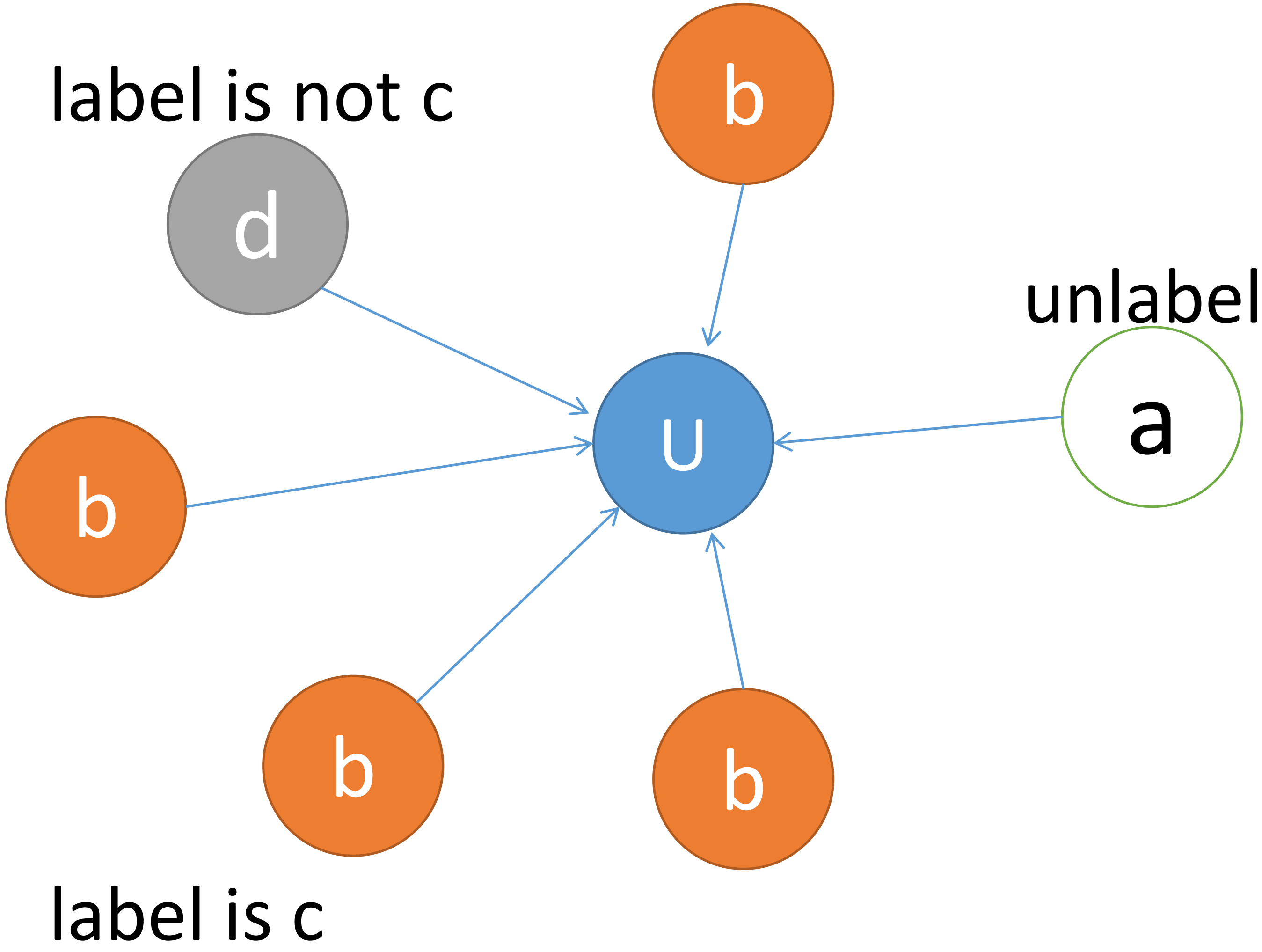
How the Size of Noisily Labeled Neighbors Affect the Node Classification?

unlabel node U ,label is c



$$\mathbb{E}(s_{bc}) > \mathbb{E}(s_{ac}) > \mathbb{E}(s_{dc}).$$

link with k labeled nodes which belong to c



$$y_{uc} = \frac{1}{m+n}\left(\sum_{v_a \in \mathcal{V}_a} s_{ac} + \sum_{v_l \in \mathcal{V}_n} s_{lc}\right),$$

$$\mathbb{E}(y_{uc}) = \frac{n}{m+n}\mathbb{E}(s_{ac}) + \frac{(hp_t + (1-h)p_f))m}{m+n}\mathbb{E}(s_{bc}) + \frac{(h(1-p_t) + (1-h)(1-p_f))m}{m+n}\mathbb{E}(s_{dc}),$$

$$\mathbb{E}(y_{uc}) = \frac{n\mathbb{E}(s_{ac}) + pm\mathbb{E}(s_{bc}) + (1-p)m\mathbb{E}(s_{dc})}{m+n}.$$

$$p_t\mathbb{E}(s_{bc}) + (1-p_t)\mathbb{E}(s_{dc}) > p\mathbb{E}(s_{bc}) + (1-p)\mathbb{E}(s_{dc}).$$

Assume: $p_t > \frac{\mathbb{E}(s_{ac}) - \mathbb{E}(s_{dc})}{\mathbb{E}(s_{bc}) - \mathbb{E}(s_{dc})}$

$$p_t\mathbb{E}(s_{bc}) + (1-p_t)\mathbb{E}(s_{dc}) > \mathbb{E}(s_{ac}).$$

$$p_t\mathbb{E}(s_{bc}) + (1-p_t)\mathbb{E}(s_{dc}) > \mathbb{E}(y_{uc}).$$

$$p = (hp_t + (1-h)p_f) < p_t.$$

link with k labeled nodes which belong to c

$$\mathbb{E}(y_{uc}^k) = \frac{m+n}{m+n+k}\mathbb{E}(y_{uc}) + \frac{kp_t\mathbb{E}(s_{bc}) + k(1-p_t)\mathbb{E}(s_{dc})}{m+n+k}.$$

$$\mathbb{E}(y_{uc}^k) > \mathbb{E}(y_{uc}).$$

Table 1: Accuracy(%) of node classification with noisy labels.

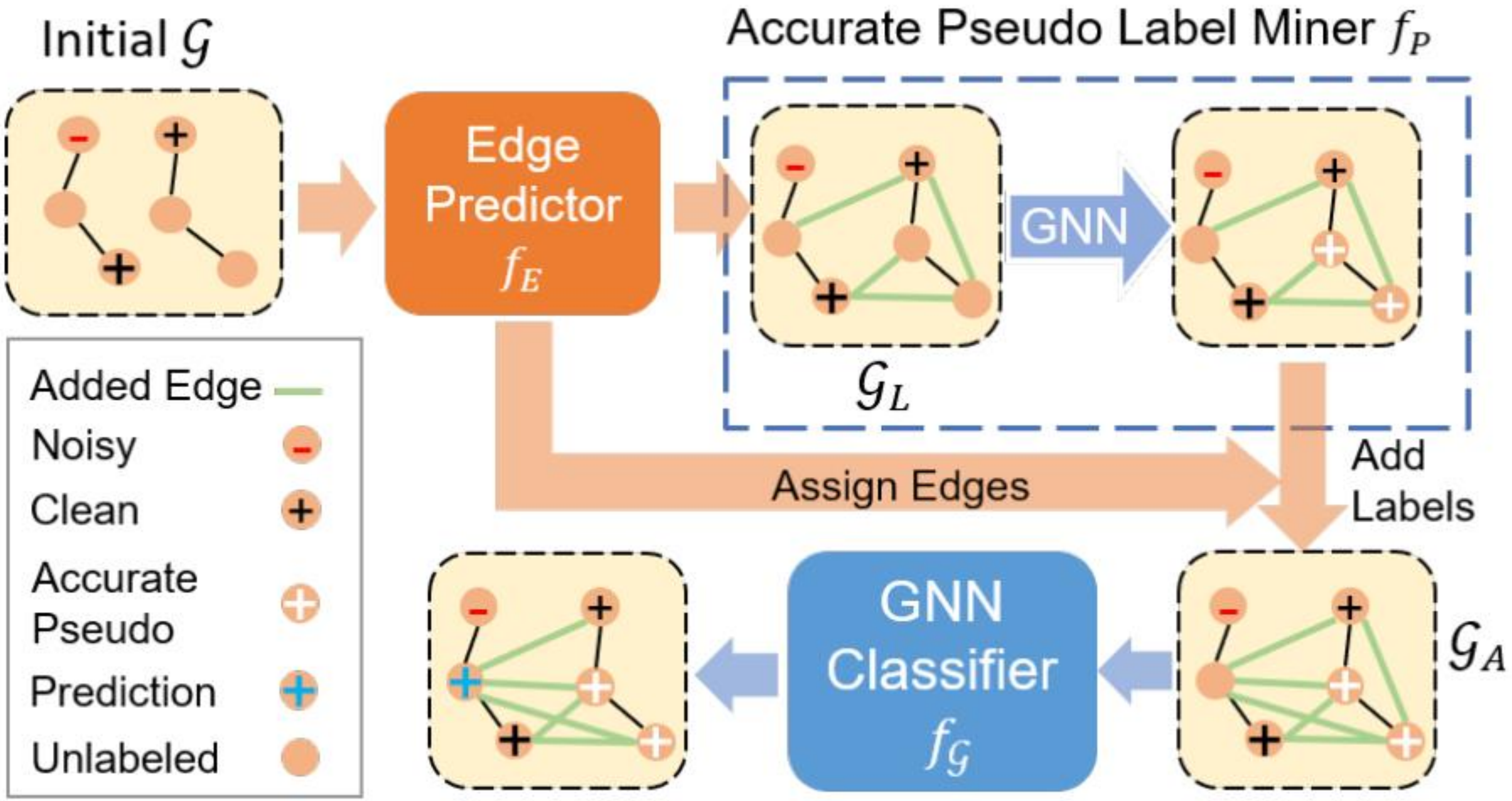
Dataset	Noise Rate	Initial $\mathcal{G}$	Link $\mathcal{V}_U$	Link $\mathcal{V}_L$
Cora	0.1	77.9 ±0.3	77.8 ±0.5	**78.7 ±0.4**
	0.2	72.8 ±1.8	72.8 ±1.0	**74.0 ±0.9**
	0.3	65.6 ±0.8	65.8 ±1.7	**68.5 ±1.4**
Citeseer	0.1	68.1 ±0.8	68.1 ±0.6	**69.0 ±1.0**
	0.2	64.9 ±1.7	65.2 ±0.8	**66.4 ±1.5**
	0.3	60.4 ±2.5	61.8 ±1.0	**62.7 ±1.0**

Table 2: Accuracy(%) of node classification with noisy labels.

Dataset	Initial $\mathcal{G}$	Link $\mathcal{V}_L$	Link $\mathcal{V}_L$ (Retrain)	Link $\mathcal{V}_A$
Cora	72.8 ±1.8	74.0 ±0.9	75.4 ±1.0	**77.1 ±1.3**
Citeseer	64.9 ±1.7	66.4 ±1.5	66.5 ±1.5	**68.0 ±1.4**

Even a simple strategy based on raw feature cosine similarity to link unlabeled nodes with labeled nodes could benefit node classification trained on noisy labels significantly.

Framework



Initial $\mathcal{G}$
Edge Predictor f_E
Accurate Pseudo Label Miner f_P
GNN
$\mathcal{G}_L$
Add Labels
Assign Edges
$\mathcal{G}_A$
GNN Classifier $f_\mathcal{G}$
Added Edge
Noisy
Clean
Accurate Pseudo
Prediction
Unlabeled

$$\min_{\theta_E} \mathcal{L}_E = \sum_{v_i \in \mathcal{V}} \sum_{v_j \in \mathcal{N}(v_i)} \left((\mathbf{S}_{ij} - 1)^2 + \sum_{n=1}^{K} \mathbb{E}_{v_n \sim P_n(v_i)} (\mathbf{S}_{in} - 0)^2 \right) \quad (6)$$

$$\mathbf{S}^L_{ij} = \begin{cases} 1 & \text{if } v_j \in \mathcal{N}(v_i); \\ \mathbf{S}_{ij} & \text{else if } \mathbf{S}_{ij} > t, v_i \in \mathcal{V}_U \text{ and } v_j \in \mathcal{V}_L; \\ 0 & \text{else,} \end{cases} \quad (7)$$

Algorithm 1: Training Algorithm of NRGNN.

Input: $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, $\mathcal{Y}$, K, t, T_p, α and β.
Output: $f_{\mathcal{G}}$, f_P and f_E

```
Pretrain f_P and f_E with Eq.(6) and Eq.(9)
repeat
    Obtain the graph S^L with f_E by Eq.(7).
    Feed S^L to f_P to obtain pseudo labels Y_P by Eq.(10)
    Generate the graph S^A for f_G with f_E by Eq.(11)
    Jointly optimize the parameters of f_G, f_P and f_E by Eq.(14)
until convergence
return f_G, f_P and f_E
```

$$\min_{\theta_P} \mathcal{L}_P = \sum_{v_i \in \mathcal{V}_L} l(\hat{y}^P_i, y_i), \quad (9)$$

$$\mathcal{Y}_P = \{\hat{y}^P_i \in \hat{\mathcal{Y}}^P_U ; \hat{y}^P_{ic} > T_p\}, \quad (10)$$

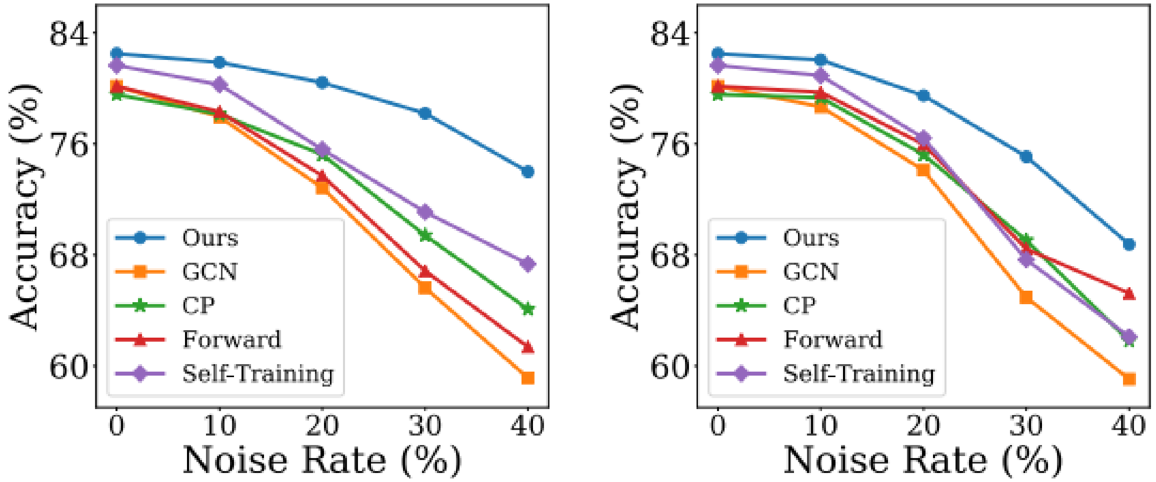
$$\mathbf{S}^A_{ij} = \begin{cases} 1 & \text{if } v_j \in \mathcal{N}(v_i); \\ \mathbf{S}_{ij} & \text{else if } \mathbf{S}_{ij} > t, v_i \in \mathcal{V}_U \text{ and } v_j \in \mathcal{V}_A; \\ 0 & \text{else,} \end{cases} \quad (11)$$

$$\arg\min_{\theta_E, \theta_P, \theta_{\mathcal{G}}} \mathcal{L}_{\mathcal{G}} + \alpha \mathcal{L}_E + \beta \mathcal{L}_P, \quad (14)$$

Experiments

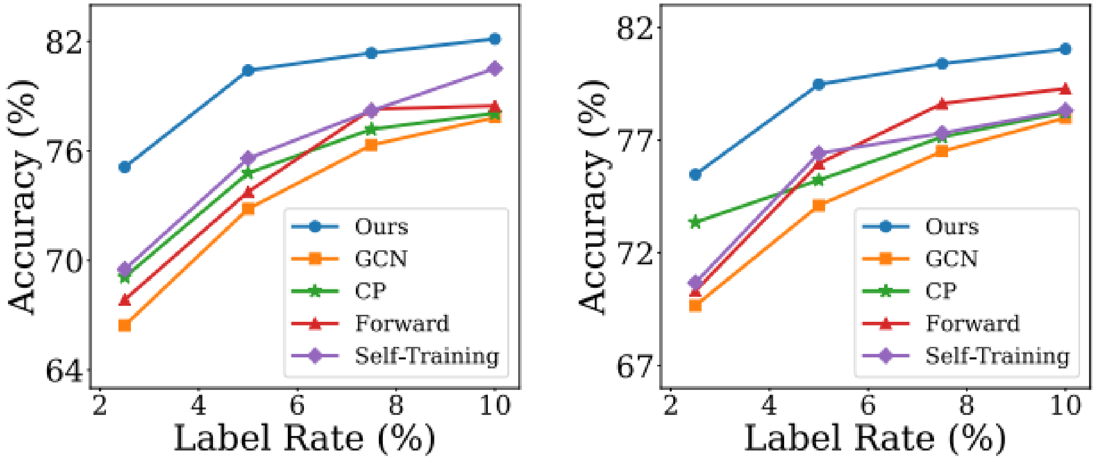
Table 3: Node classification performance (Accuracy (%)±Std) under various types of noise.

Dataset	Noise	GCN	GIN	Self-Training	Forward	Coteaching+	D-GNN	CP	Ours
Cora	Uniform	72.8 ±1.8	72.3 ±0.9	75.6 ±1.8	73.7 ±0.7	73.6 ±1.7	72.4 ±1.8	74.8 ±1.3	**80.4 ±0.5**
	Pair	74.1 ±0.7	74.7 ±1.4	76.4 ±1.4	76.0 ±0.7	73.8 ±1.4	73.5 ±1.6	75.2 ±1.4	**79.5 ±0.4**
Citeseer	Uniform	64.9 ±1.7	65.7 ±2.1	67.8 ±1.4	65.0 ±1.5	66.4 ±1.3	64.9 ±1.3	66.0 ±1.6	**70.1 ±1.3**
	Pair	60.3 ±1.0	61.6 ±1.0	62.0 ±1.6	61.6 ±0.4	65.1 ±2.1	62.3 ±1.2	62.0 ±1.0	**67.8 ±1.3**
Pubmed	Uniform	77.3 ±0.9	77.4 ±0.5	78.2 ±0.4	77.5 ±0.4	78.6 ±0.4	77.6 ±0.3	78.6 ±0.3	**80.0 ±0.2**
	Pair	78.0 ±0.4	78.1 ±0.6	78.9 ±0.8	79.6 ±0.2	78.5 ±0.1	79.4 ±0.4	77.9 ±0.3	**80.0 ±0.3**
DBLP	Uniform	71.0 ±1.5	72.4 ±0.7	74.9 ±0.7	73.1 ±0.3	73.5 ±1.3	72.8 ±1.2	74.2 ±0.5	**80.8 ±0.4**
	Pair	72.5 ±1.2	73.4 ±2.1	76.3 ±1.6	74.4 ±0.5	72.7 ±1.2	75.4 ±0.9	73.6 ±1.0	**81.1 ±0.3**



(a) Uniform Noise (b) Pair Noise

Figure 2: Accuracy on Cora with various levels of label noise.



(a) Uniform Noise (b) Pair Noise

Figure 3: Accuracy on Cora with various noisy label sizes.

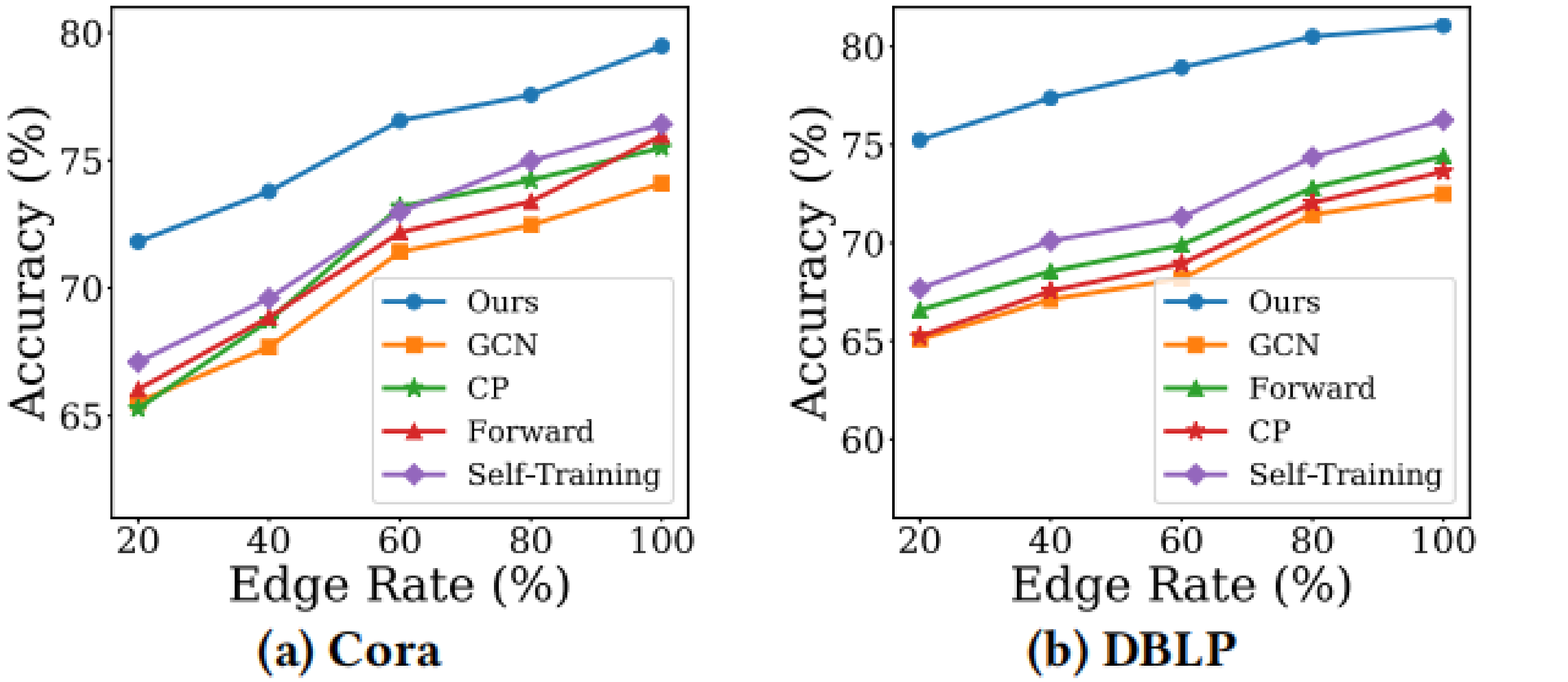


Figure 4: Performance on graphs with different densities.

Ablation Study

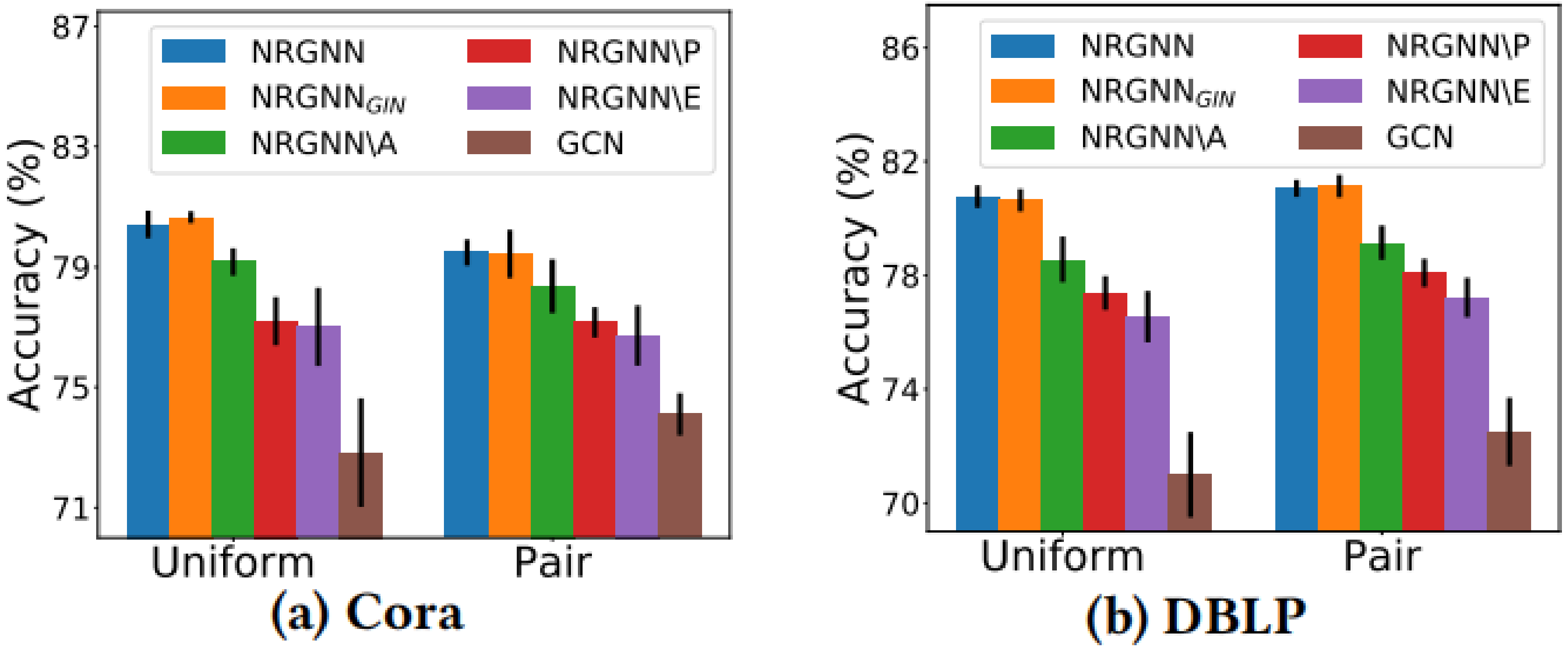


Figure 5: Comparisons between NRGNN and its variants.