Rethinking Semi-Supervised Imbalanced Node Classification from Bias-Variance Decomposition

Divin Yan, Gengchen Wei, Chen Yang, Shengzhong Zhang, Zengfeng Huang

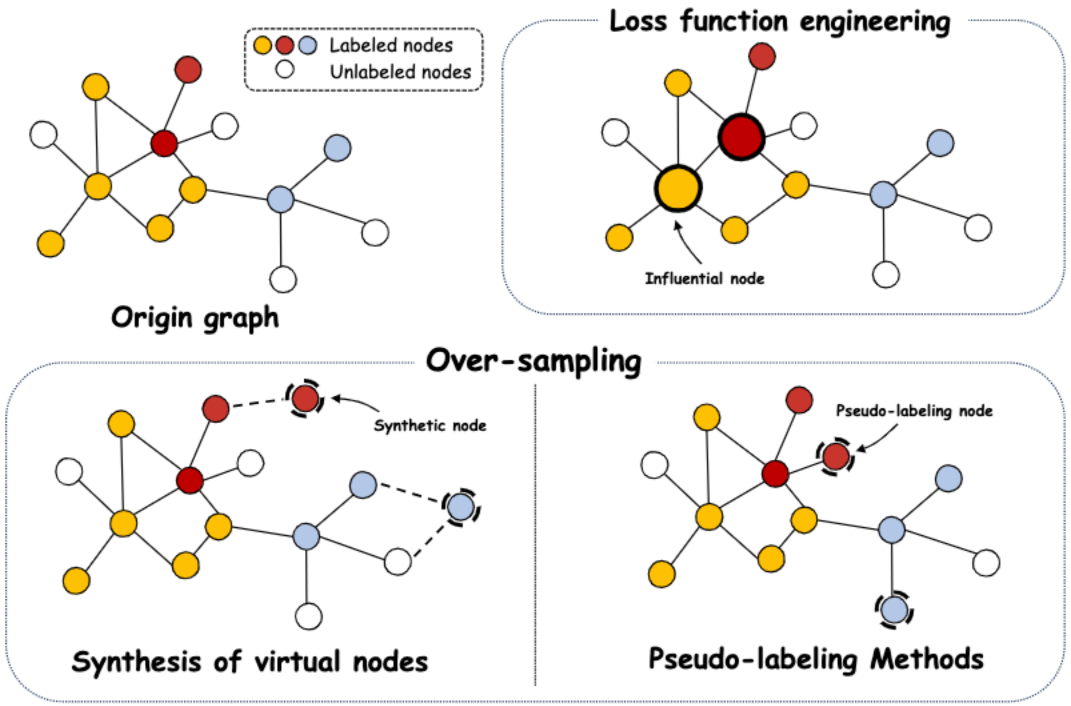
Fudan University

NEURAL INFORMATION PROCESSING SYSTEMS

FUDAN UNIVERSITY

Imbalanced Node Classification

- Learning on graphs commonly struggles with class imbalance issues.
- Due to topological asymmetries[1,2,3] impact model performance, conventional methods have been proved ineffective.
- Thus, a more fundamental and theoretical perspective is urgently needed.

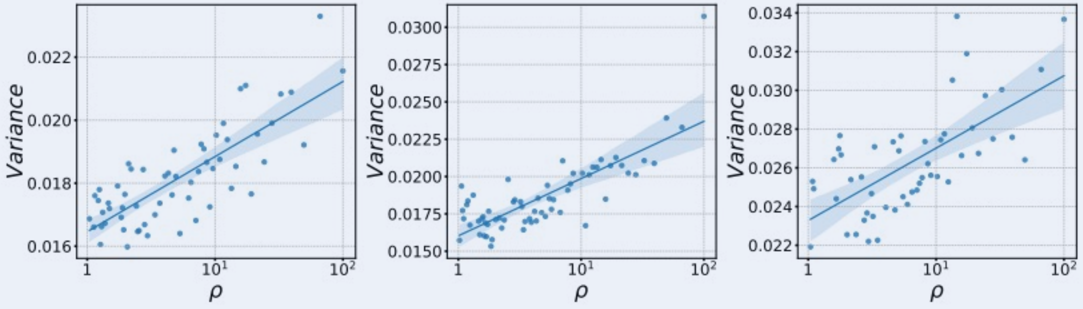


1. Topology-Imbalance Learning for Semi-Supervised Node Classification. NeurIPS 2021
2. GraphENS: Neighbor-Aware Ego Network Synthesis for Class-Imbalanced Node Classification. ICLR 2022
3. TAM: Topology-Aware Margin Loss for Class-Imbalanced Node Classification. ICML 2022

Graph Imbalance and Model Variance

Theorem 1 *Under the condition that $\sum_i n_i$ is a constant, the variance $\sum_{i=1}^{c} \mathbb{E}_x \left[\frac{1}{n_i} h^T(x) \Lambda^i h(x) \right]$ reach its minimum when all n_i equal.*

As the ratio of imbalance increases, the minority class exhibits a smaller sample size n_i, which consequently makes a greater contribution to the overall variance.



(a) *CiteSeer-GCN* (b) *CiteSeer-GAT* (c) *CiteSeer-SAGE*

Variance Regularization

Estimate the Expectation with Labeled Nodes: aswe lack access to other training sets. Therefore, we propose Lemma 1 to estimate the variance on training set with the variance on labeled data.

Lemma 1 *Under the above assumption for $h^i \sim N(\mu^i, \Lambda^i)$, $C^i \sim N\left(\mu^i, \frac{1}{n_i}\Lambda^i\right)$, minimizing the $\sum_{i=1}^{c} \mathbb{E}_x [\mathrm{Var}(x)]$ is equivalent to minimizing Equation 3:*

$$\frac{1}{N}\sum_{x \in G}\sum_{i=1}^{c}\left(h(x)^T \frac{1}{\sqrt{2n_i}}\left(h_1^i - h_2^i\right)\right)^2 = \frac{1}{N}\sum_{k=1}^{n_j}\sum_{j=1}^{c}\sum_{i=1}^{c}\left((\mu_k^j + \epsilon_k^j)^T \frac{1}{\sqrt{2n_i}}\left(\epsilon_1^i - \epsilon_2^i\right)\right)^2. \quad (3)$$

NEURAL INFORMATION PROCESSING SYSTEMS

FUDAN UNIVERSITY

Variance Regularization

Estimate the Expectation with Labeled Nodes: aswe lack access to other training sets. Therefore, we propose Lemma 1 to estimate the variance on training set with the variance on labeled data.

Lemma 1 *Under the above assumption for* $h^i \sim N(\mu^i, \Lambda^i)$, $C^i \sim N\left(\mu^i, \frac{1}{n_i}\Lambda^i\right)$, *minimizing the* $\sum_{i=1}^{c} \mathbb{E}_x\left[\mathrm{Var}(x)\right]$ *is equivalent to minimizing Equation 3:*

$$\frac{1}{N}\sum_{x\in G}\sum_{i=1}^{c}\left(h(x)^T\frac{1}{\sqrt{2n_i}}\left(h_1^i - h_2^i\right)\right)^2 = \frac{1}{N}\sum_{k=1}^{n_j}\sum_{j=1}^{c}\sum_{i=1}^{c}\left(\left(\mu_k^j + \epsilon_k^j\right)^T\frac{1}{\sqrt{2n_i}}\left(\epsilon_1^i - \epsilon_2^i\right)\right)^2. \quad (3)$$

Difficulties: Equation 3 required access to embedding pairs from the same class, which were difficult to obtain due to the lack of labels.

Variance Regularization

Estimate the Expectation with Unlabeled Nodes: To address this, we developed a two-step solution. The first step involved using graph augmentation to create pseudo embedding pairs (h_1, h_2).

Lemma 1 *Under the above assumption for $h^i \sim N(\mu^i, \Lambda^i)$, $C^i \sim N\left(\mu^i, \frac{1}{n_i}\Lambda^i\right)$, minimizing the $\sum_{i=1}^{c} \mathbb{E}_x\left[\mathrm{Var}(x)\right]$ is equivalent to minimizing Equation 3:*

$$\frac{1}{N}\sum_{x\in G}\sum_{i=1}^{c}\left(h(x)^T\frac{1}{\sqrt{2n_i}}\left(h_1^i - h_2^i\right)\right)^2 = \frac{1}{N}\sum_{k=1}^{n_j}\sum_{j=1}^{c}\sum_{i=1}^{c}\left(\left(\mu_k^j + \epsilon_k^j\right)^T\frac{1}{\sqrt{2n_i}}\left(\epsilon_1^i - \epsilon_2^i\right)\right)^2. \quad (3)$$

Variance Regularization

Estimate the Expectation with Unlabeled Nodes: To address this, we developed a two-step solution. The first step involved using graph augmentation to create pseudo embedding pairs (h_1, h_2).

Lemma 1 *Under the above assumption for* $h^i \sim N(\mu^i, \Lambda^i)$, $C^i \sim N\left(\mu^i, \frac{1}{n_i}\Lambda^i\right)$, *minimizing the* $\sum_{i=1}^{c} \mathbb{E}_x\left[\mathrm{Var}(x)\right]$ *is equivalent to minimizing Equation 3:*

$$\frac{1}{N}\sum_{x\in G}\sum_{i=1}^{c}\left(h(x)^T\frac{1}{\sqrt{2n_i}}\left(h_1^i - h_2^i\right)\right)^2 = \frac{1}{N}\sum_{k=1}^{n_j}\sum_{j=1}^{c}\sum_{i=1}^{c}\left(\left(\mu_k^j + \epsilon_k^j\right)^T\frac{1}{\sqrt{2n_i}}\left(\epsilon_1^i - \epsilon_2^i\right)\right)^2. \quad (3)$$

Difficulties: These pseudo node pairs was their lack of information about class label, preventing us from assigning the correct coefficients in Equation 3, which are vital for compensating minority classes.

Variance Regularization

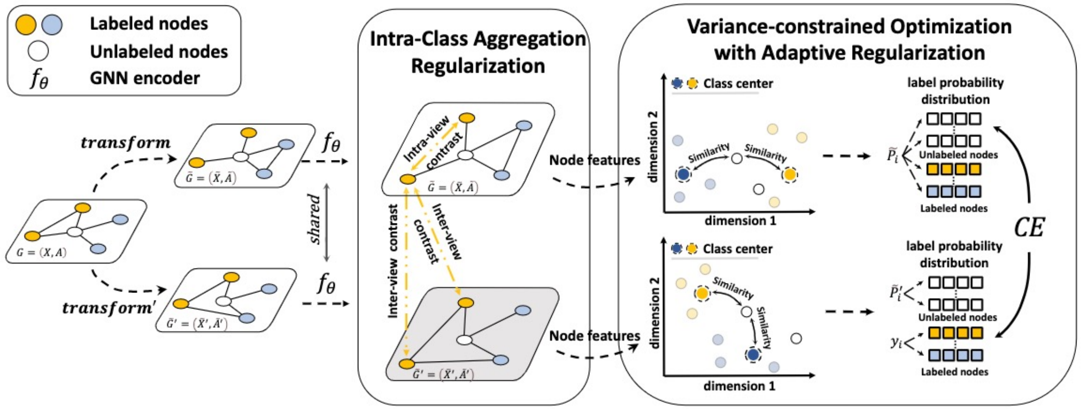
The Second Step: To overcome this, in our second step, we introduced the use of class centers , denoted as C^i to replace h in the equation.

$$
\begin{aligned}
&\frac{1}{N}\sum_{x\in G}\sum_{i=1}^{c}\Big((C^i)^T h(x) - (C^{i'})^T h'(x)\Big)^2 \\
&= \frac{1}{N}\sum_{k=1}^{n_j}\sum_{j=1}^{c}\sum_{i=1}^{c}\Big((\mu^i)^T(\epsilon_k^j - \epsilon_k^{j'}) + (e^i)^T\epsilon_k^j - (e^{i'})^T\epsilon_k^{j'}\Big)^2
\end{aligned} \tag{4}
$$

NEURAL INFORMATION PROCESSING SYSTEMS

FUDAN UNIVERSITY

The Final Algorithm: Revar



(a) Graph augmentations

(b) Intra-view and inter-view regularization (Corresponding to $\mathcal{L}_{IR}$)

(c) Variance regularization (Corresponding to $\mathcal{L}_{VR}$)

$$\mathcal{L}_{\mathrm{VR}} = \frac{1}{|V_{\mathrm{conf}}|}\sum_{i\in V_{\mathrm{conf}}} CE\left(\tilde{\pi}_i', \tilde{\pi}_i\right) + \frac{1}{|V_L|}\sum_{i\in V_L} CE\left(y_i, \tilde{\pi}_i\right) \tag{7}$$

$$\mathcal{L}_{\mathrm{IR}} = -\frac{1}{|V_U|}\sum_{\mathrm{h}_i,\mathrm{h}_i'\in V_U} \mathrm{sim}\left(\mathrm{h}_i\cdot \mathrm{h}_i'\right) - \frac{1}{N_{all}}\left(\sum_{l=1}^{k}\sum_{\mathrm{h}_i,\mathrm{h}_j'\in \mathbf{C}_l} \mathrm{sim}\left(\mathrm{h}_i\cdot \mathrm{h}_j'\right) + \sum_{l=1}^{k}\sum_{\substack{\mathrm{h}_i,\mathrm{h}_j\in \mathbf{C}_l\\ i\neq j}} \mathrm{sim}\left(\mathrm{h}_i\cdot \mathrm{h}_j\right)\right) \tag{8}$$

$$\mathcal{L}_{\mathrm{composite}} = \lambda_1\mathcal{L}_{\mathrm{VR}} + \lambda_2\mathcal{L}_{\mathrm{IR}} + \mathcal{L}_{\mathrm{sup}} \tag{9}$$

Performance of Revar

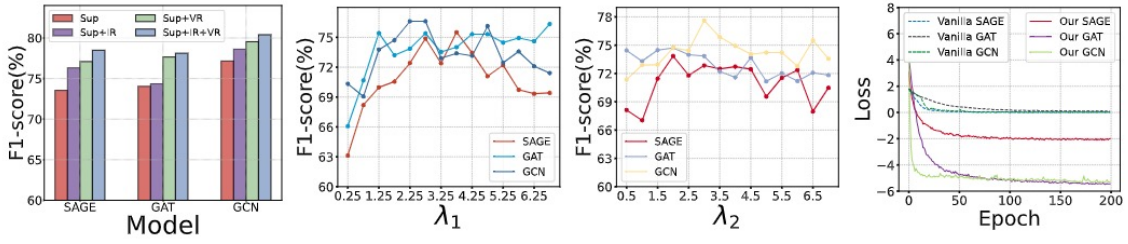
In all cases, ReVar achieves a decisive advantage that underscores its efficacy in addressing the challenge of class imbalance in node classification.

	Dataset	*CiteSeer-Semi*		*PubMed-Semi*		*Computers-Semi*	
	Imbalance Ratio ($\rho = 10$)	bAcc.	F1	bAcc.	F1	bAcc.	F1
GCN	Vanilla	38.72 ± 1.88	28.74 ± 3.21	65.64 ± 1.72	56.97 ± 3.17	80.01 ± 0.71	71.56 ± 0.81
GCN	Re-Weight	44.69 ± 1.78	38.61 ± 2.37	69.06 ± 1.84	64.08 ± 2.97	80.93 ± 1.30	73.99 ± 2.20
GCN	PC Softmax	50.18 ± 0.55	46.14 ± 0.14	72.46 ± 0.80	70.27 ± 0.94	81.54 ± 0.76	73.30 ± 0.51
GCN	GraphSMOTE	44.87 ± 1.12	39.20 ± 1.62	67.91 ± 0.64	62.68 ± 1.92	79.48 ± 0.47	72.63 ± 0.76
GCN	BalancedSoftmax	55.52 ± 0.97	53.74 ± 1.42	73.73 ± 0.89	71.53 ± 1.06	81.46 ± 0.74	**74.31** ± 0.51
GCN	+ TAM	56.73 ± 0.71	56.15 ± 0.78	74.62 ± 0.97	72.25 ± 1.30	82.36 ± 0.67	72.94 ± 1.43
GCN	Renode	43.47 ± 2.22	37.52 ± 3.10	71.40 ± 1.42	67.27 ± 2.96	81.89 ± 0.77	73.13 ± 1.60
GCN	+ TAM	46.20 ± 1.17	39.96 ± 2.76	72.63 ± 2.03	68.28 ± 3.30	80.36 ± 1.19	72.51 ± 0.68
GCN	GraphENS	56.57 ± 0.98	55.29 ± 1.33	72.13 ± 1.04	70.72 ± 1.07	**82.40** ± 0.39	74.26 ± 1.05
GCN	+ TAM	**58.01** ± 0.68	**56.32** ± 1.03	**74.14** ± 1.42	**72.42** ± 1.39	81.02 ± 0.99	70.78 ± 1.72
GCN	**ReVar**	**65.28** ± 0.51	**64.91** ± 0.51	**79.20** ± 0.72	**78.45** ± 0.46	**84.67** ± 0.17	**80.25** ± 0.87
GCN	**Δ**	**+ 7.27**(12.53%)	**+ 8.59**(15.25%)	**+ 5.06**(6.82%)	**+ 6.03**(8.33%)	**+ 2.27**(2.75%)	**+ 5.94**(7.99%)
GAT	Vanilla	38.84 ± 1.13	31.25 ± 1.64	64.60 ± 1.64	55.24 ± 2.80	79.04 ± 1.60	70.00 ± 2.50
GAT	Re-Weight	45.47 ± 2.35	40.60 ± 2.98	68.10 ± 2.85	63.76 ± 3.54	80.38 ± 0.66	69.99 ± 0.76
GAT	PC Softmax	50.78 ± 1.66	48.56 ± 2.08	72.88 ± 0.83	71.09 ± 0.89	79.43 ± 0.94	71.33 ± 0.86
GAT	GraphSMOTE	45.68 ± 0.93	38.96 ± 0.97	67.43 ± 1.23	61.97 ± 2.54	79.38 ± 1.97	69.76 ± 2.31
GAT	BalancedSoftmax	54.78 ± 1.25	51.83 ± 2.11	72.30 ± 1.20	69.30 ± 1.79	**82.02** ± 1.19	**72.94** ± 1.54
GAT	+ TAM	56.30 ± 1.25	53.87 ± 1.14	73.50 ± 1.24	71.36 ± 1.99	75.54 ± 2.09	66.69 ± 1.44
GAT	Renode	44.48 ± 2.06	37.93 ± 2.87	69.93 ± 2.10	65.27 ± 2.90	76.01 ± 1.08	66.72 ± 1.42
GAT	+ TAM	45.12 ± 1.41	39.29 ± 1.79	70.66 ± 2.13	66.94 ± 3.54	74.30 ± 1.13	66.13 ± 1.75
GAT	GraphENS	51.45 ± 1.28	47.98 ± 2.08	73.15 ± 1.24	71.90 ± 1.03	81.23 ± 0.74	71.23 ± 0.42
GAT	+ TAM	**56.15** ± 1.13	**54.31** ± 1.68	**73.45** ± 1.07	**72.10** ± 0.36	81.07 ± 1.03	71.27 ± 1.98
GAT	**ReVar**	**66.04** ± 0.66	**65.70** ± 0.69	**77.85** ± 0.76	**77.08** ± 0.69	**86.37** ± 0.02	**82.35** ± 0.02
GAT	**Δ**	**+ 9.89**(17.61%)	**+ 11.39**(20.97%)	**+ 4.40**(5.99%)	**+ 4.98**(6.91%)	**+ 4.35**(5.30%)	**+ 9.41**(12.90%)
SAGE	Vanilla	43.18 ± 0.52	36.66 ± 1.25	68.68 ± 1.51	64.16 ± 2.38	72.36 ± 2.39	64.32 ± 2.21
SAGE	Re-Weight	46.17 ± 1.32	40.13 ± 1.68	69.89 ± 1.60	65.71 ± 2.31	76.08 ± 1.14	65.76 ± 1.40
SAGE	PC Softmax	50.66 ± 0.99	47.48 ± 1.66	71.49 ± 0.94	70.23 ± 0.67	74.63 ± 3.01	66.44 ± 4.04
SAGE	GraphSMOTE	42.73 ± 2.87	35.18 ± 1.75	66.63 ± 0.65	61.97 ± 2.54	71.85 ± 0.98	68.92 ± 0.73
SAGE	BalancedSoftmax	51.74 ± 2.32	49.01 ± 3.16	71.36 ± 1.37	69.66 ± 1.81	73.67 ± 1.11	65.23 ± 2.44
SAGE	+ TAM	51.93 ± 2.19	48.67 ± 3.25	72.28 ± 1.47	71.02 ± 1.31	77.00 ± 2.93	70.85 ± 2.28
SAGE	Renode	48.65 ± 1.37	44.25 ± 2.20	71.37 ± 1.33	67.78 ± 1.38	77.37 ± 0.74	68.42 ± 1.81
SAGE	+ TAM	48.39 ± 1.76	43.56 ± 2.31	71.25 ± 1.07	68.69 ± 0.98	74.87 ± 2.25	66.87 ± 2.52
SAGE	GraphENS	53.51 ± 0.78	51.42 ± 1.19	70.97 ± 0.78	70.00 ± 1.22	**82.57** ± 0.50	71.95 ± 0.51
SAGE	+ TAM	**54.69** ± 1.12	**53.56** ± 1.86	**73.61** ± 1.35	**72.50** ± 1.58	82.17 ± 0.93	**72.46** ± 1.00
SAGE	**ReVar**	**60.48** ± 0.88	**57.99** ± 1.54	**77.72** ± 1.06	**76.01** ± 1.20	**83.50** ± 0.02	**76.48** ± 0.05
SAGE	**Δ**	**+ 5.79**(10.59%)	**+ 4.53**(8.46%)	**+ 4.11**(5.58%)	**+ 3.51**(4.84%)	**+ 0.93**(1.13%)	**+ 4.02**(5.55%)

Performance of Revar

We also conduct further analysis to verify the superiority of Revar.

Dataset(*CS-Random*)	GCN		GAT		SAGE	
Imbalance Ratio($\rho = 41.00$)	bAcc.	F1	bAcc.	F1	bAcc.	F1
Vanilla	84.85 $\pm$ 0.16	87.12 $\pm$ 0.14	82.47 $\pm$ 0.36	84.21 $\pm$ 0.31	83.76 $\pm$ 0.27	86.22 $\pm$ 0.19
Re-Weight	87.42 $\pm$ 0.17	88.70 $\pm$ 0.10	83.55 $\pm$ 0.39	84.73 $\pm$ 0.32	85.76 $\pm$ 0.24	87.32 $\pm$ 0.16
PC Softmax	**88.36** $\pm$ 0.12	88.94 $\pm$ 0.04	85.22 $\pm$ 0.31	85.54 $\pm$ 0.33	87.18 $\pm$ 0.14	88.00 $\pm$ 0.19
GraphSMOTE	85.76 $\pm$ 1.73	87.31 $\pm$ 1.32	84.65 $\pm$ 1.32	85.63 $\pm$ 1.01	85.76 $\pm$ 1.98	87.34 $\pm$ 0.98
BalancedSoftmax	87.72 $\pm$ 0.07	88.67 $\pm$ 0.07	84.38 $\pm$ 0.20	84.53 $\pm$ 0.41	86.78 $\pm$ 0.10	88.05 $\pm$ 0.09
+ TAM	88.22 $\pm$ 0.11	**89.22** $\pm$ 0.08	85.48 $\pm$ 0.24	85.77 $\pm$ 0.50	**87.83** $\pm$ 0.13	**88.77** $\pm$ 0.07
Renode	87.53 $\pm$ 0.11	88.91 $\pm$ 0.06	85.98 $\pm$ 0.19	86.97 $\pm$ 0.09	86.13 $\pm$ 0.10	87.89 $\pm$ 0.09
+ TAM	87.55 $\pm$ 0.06	89.03 $\pm$ 0.05	**86.61** $\pm$ 0.30	**87.42** $\pm$ 0.24	85.21 $\pm$ 0.33	87.01 $\pm$ 0.31
GraphENS	85.97 $\pm$ 0.29	86.68 $\pm$ 0.20	85.86 $\pm$ 0.19	86.51 $\pm$ 0.32	85.39 $\pm$ 0.26	86.41 $\pm$ 0.24
+ TAM	86.34 $\pm$ 0.12	87.36 $\pm$ 0.08	86.29 $\pm$ 0.20	87.28 $\pm$ 0.13	85.99 $\pm$ 0.13	87.25 $\pm$ 0.07
ReVar	**88.44** $\pm$ 0.16	**89.54** $\pm$ 0.11	**87.33** $\pm$ 0.04	**88.33** $\pm$ 0.06	**90.11** $\pm$ 0.11	**91.18** $\pm$ 0.11
Δ	**+ 0.08**(0.09%)	**+ 0.32**(0.36%)	**+ 0.72**(0.83%)	**+ 0.91**(1.04%)	**+ 2.28**(2.60%)	**+ 2.41**(2.71%)



(a) *Computers-Semi* (b) *Pubmed-*λ_1 (c) *PubMed-*λ_2 (d) *CiteSeer-Loss*

NEURAL INFORMATION PROCESSING SYSTEMS

FUDAN UNIVERSITY

Thanks