Chongqing University of Technology

ATAI
Advanced Technique of Artificial Intelligence

UNREAL:Unlabeled Nodes Retrieval and Labeling for Heavily-imbalanced Node Classification

Liang Yan * Shengzhong Zhang * Bisheng Li Min Zhou Zengfeng Huang
Fudan University
huangzf@fudan.edu.cn

Code： https://github.com/yanliang/unreal_demo.

2023. 4. 27 • ChongQing

—— ICLR 2023

Reported by Renhui Luo

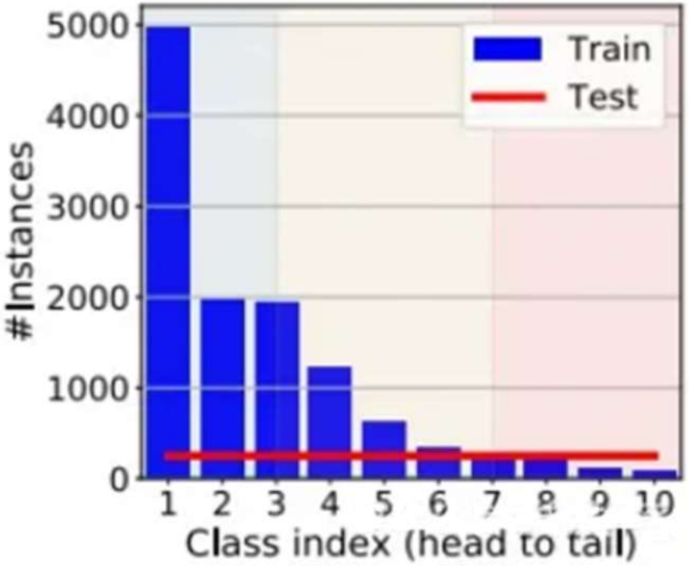
1.Introduction

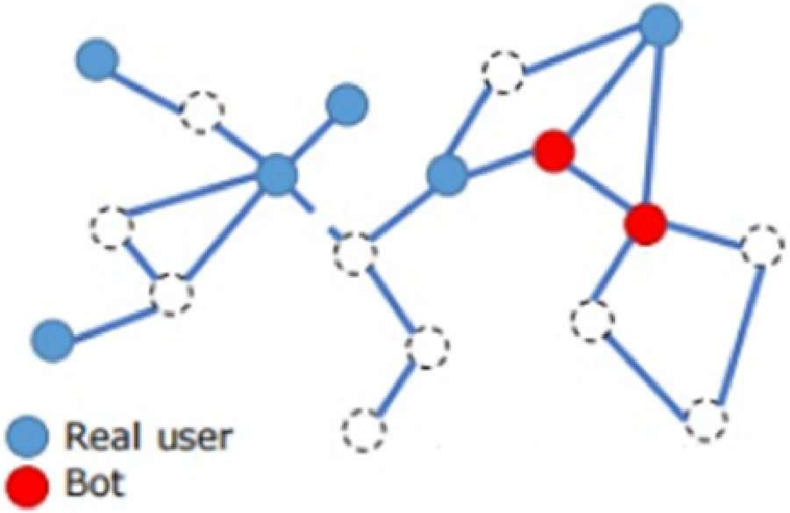
2.Overview

3.Methods

4.Experiments

Introduction





skewed label distribution

Introduction

QUESTION:

1. over-sampling used to graph data. However, it needs to additionally generate topological information for newly synthesized nodes.

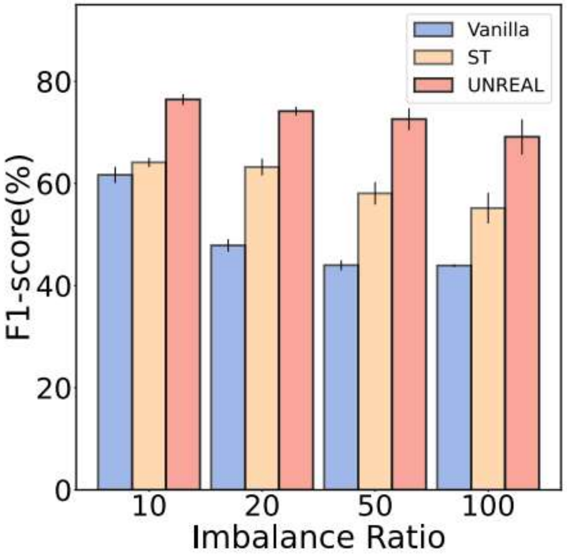
2. Self-training fails to achieve satisfactory performance in heavily-imbalanced scenarios because the bias in the original training set results in unreliable predictions, which makes the pseudo-labels used in ST highly noisy.

WORK:

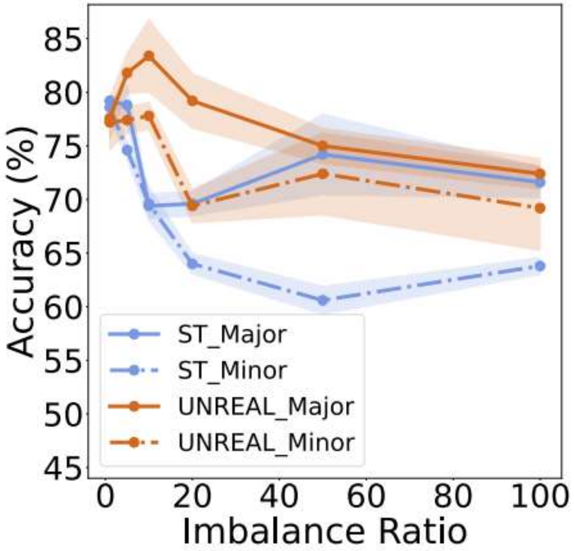
1. UNREAL adds unlabeled nodes together with their pseudo-labels to the training set. Since there is no need for syn-thesizing node features and topology, it overcomes critical shortcomings of existing oversampling approaches.

2. Geometric Imbalance (GI) issue in the embedding space and propose a metric to measure GI and discard imbalanced nodes

Preliminaries



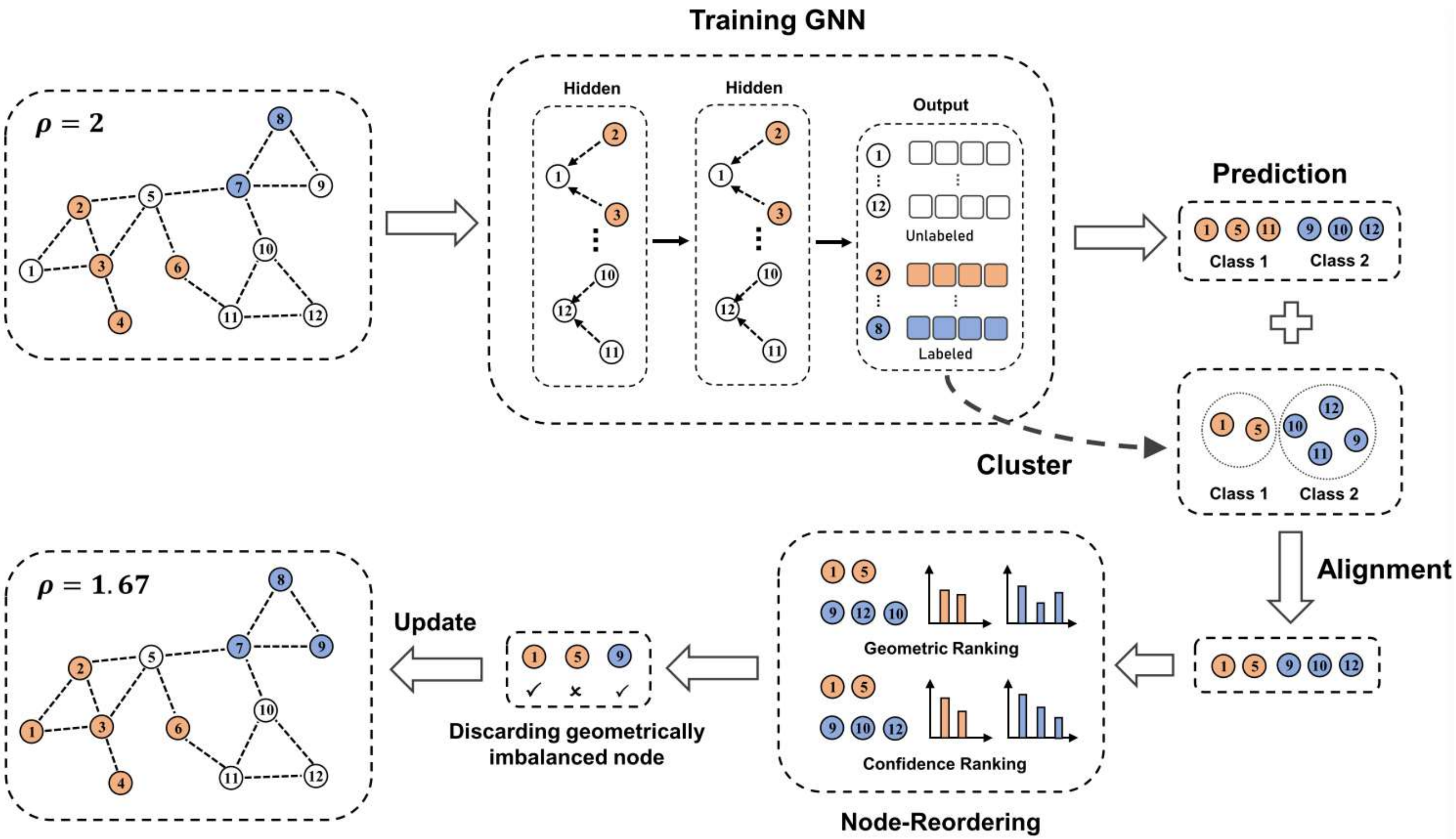
(a) Cora-GCN



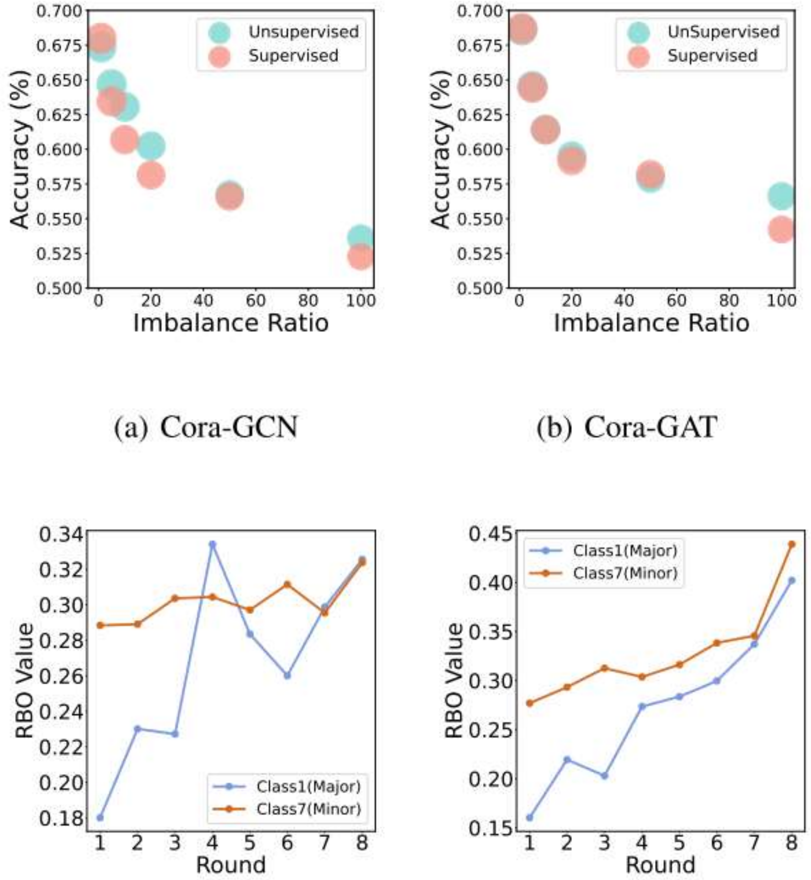
(b) Cora-GCN

Across different ratios, ST consistently outperforms vanilla model by a large margin, which verifies the positive value of the unlabeled samples of graph-structured data. As imbalance ratio increases, the performance of ST degrades rapidly, which renders that ST is insufficient for high imbalance ratios.

Overview


Training GNN
$\rho = 2$
Hidden
Hidden
Output
Unlabeled
Labeled
Prediction
Class 1
Class 2
Cluster
Class 1
Class 2
Alignment
Node-Reordering
Geometric Ranking
Confidence Ranking
Discarding geometrically imbalanced node
Update
$\rho = 1.67$

Overview



(a) Cora-GCN

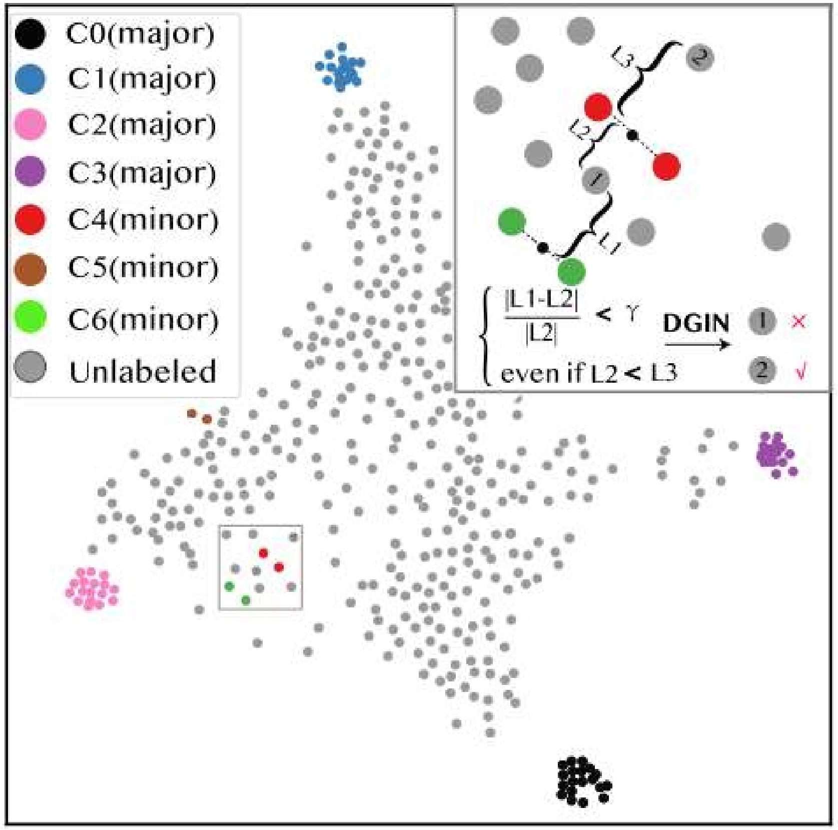
(b) Cora-GAT

(c) Cora-GCN

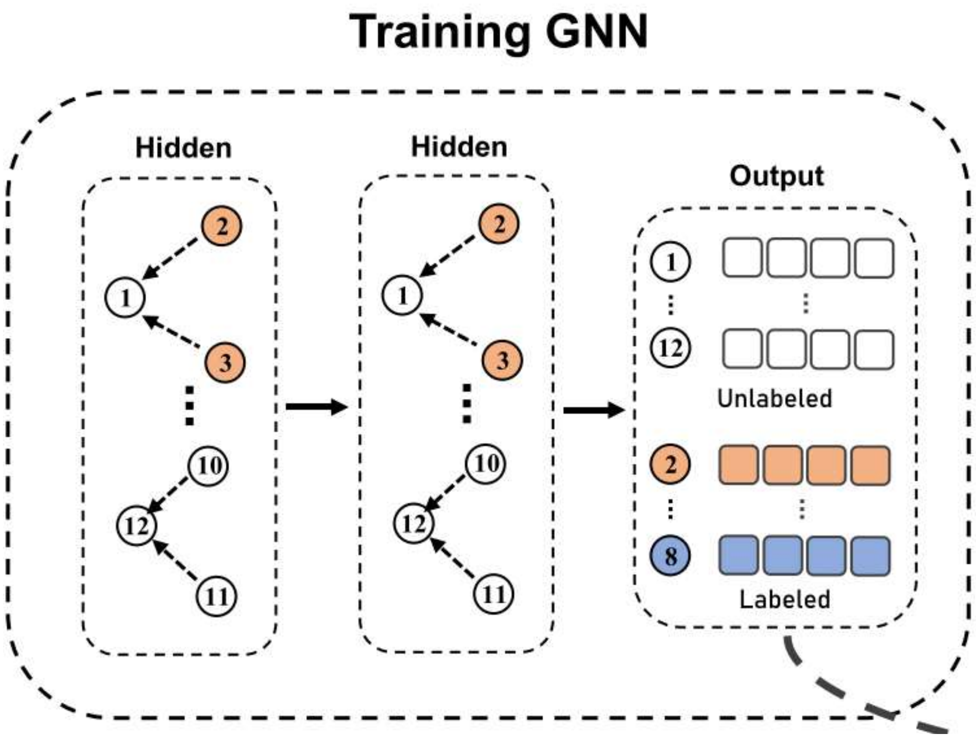
(d) Cora-GAT

Overview

Geometric Imbalance



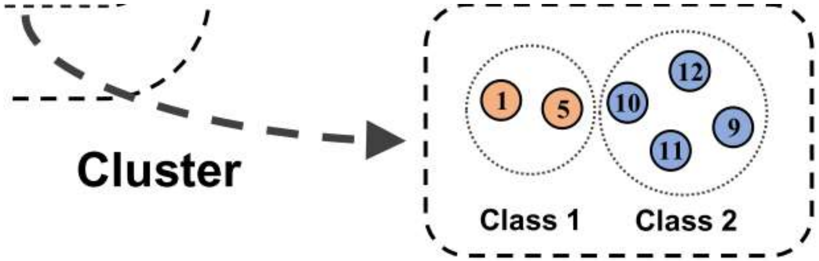
Method



Message Passing Neural Networks

$$h_v^{(l+1)} = \psi_l\left(h_v^{(l)}, \theta_l\left(\left\{m_l\left(h_v^{(l)}, h_u^{(l)}, e_{v,u}\right) \mid u \in \mathcal{N}(v)\right\}\right)\right) \tag{1}$$

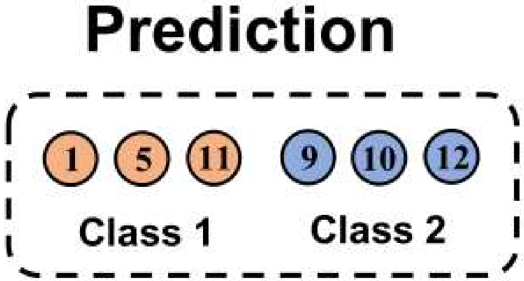
Method



$$f_{\text{cluster}}(H^U) \Longrightarrow \{\mathcal{K}_1, c_1, \mathcal{K}_2, c_2, \cdots, \mathcal{K}_{k'}, c_{k'}\} \qquad (2)$$

$$c_i^{\text{train}} = M(\{h_u^L \mid y_u \in \mathcal{C}_i\}). \qquad (3)$$

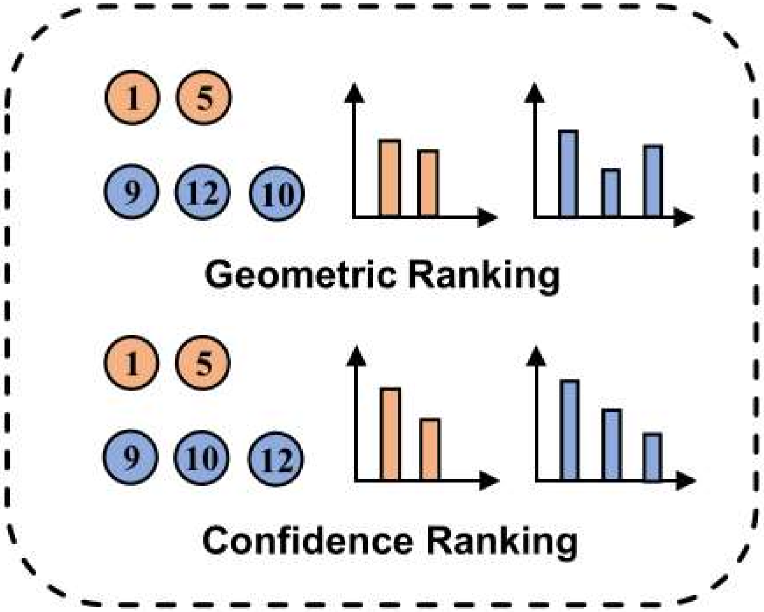
$$\tilde{y}_i = \arg\min_j \mathsf{distance}(c_j^{\text{train}}, c_i). \qquad (4)$$



$$\mathcal{U} = \bigcup_{m=1}^{k} \tilde{\mathcal{U}}_m.$$

$$\mathcal{U} = \bigcup_{m=1}^{k} \mathcal{U}_m.$$

Method



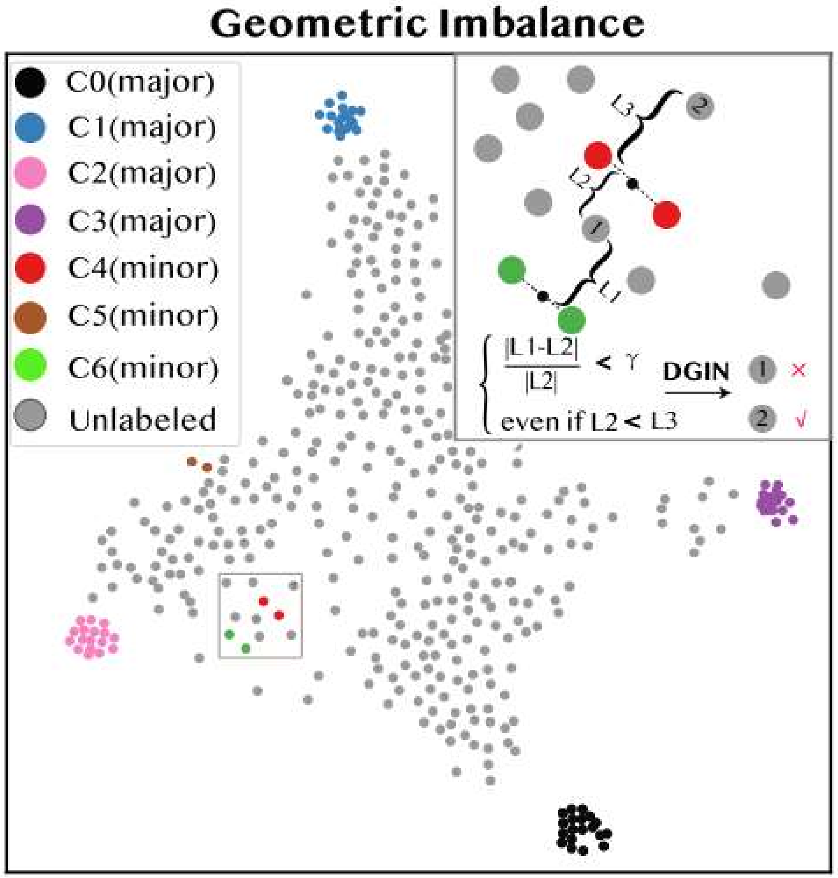
Node-Reordering

$$\delta_u = \text{distance}\left(h_u^U, c_m^{\text{train}}\right) \tag{5}$$

$$\text{confidence} = \max\left(\text{softmax}\left(logits\right)\right), \tag{6}$$

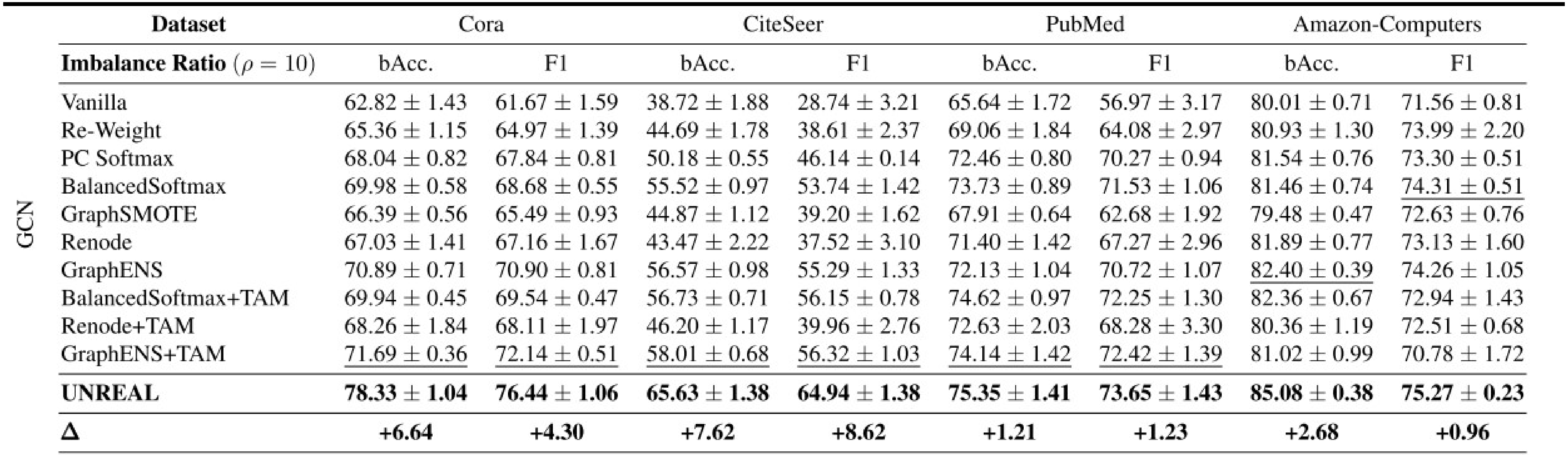
$$\mathcal{N}_m^{New} = \max\{r_m, 1-r_m\} \cdot \mathcal{S}_m + \min\{r_m, 1-r_m\} \cdot \mathcal{T}_m, \tag{7}$$

Method



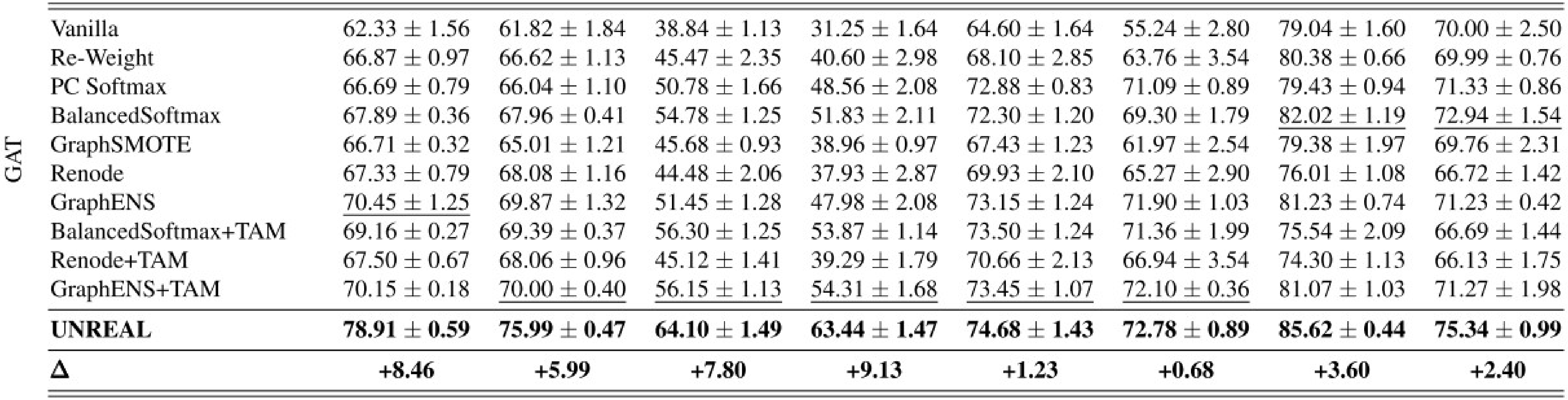
$$\mathrm{GI}_u = \frac{\beta_u - \delta_u}{\delta_u}. \qquad (8)$$

Experiments



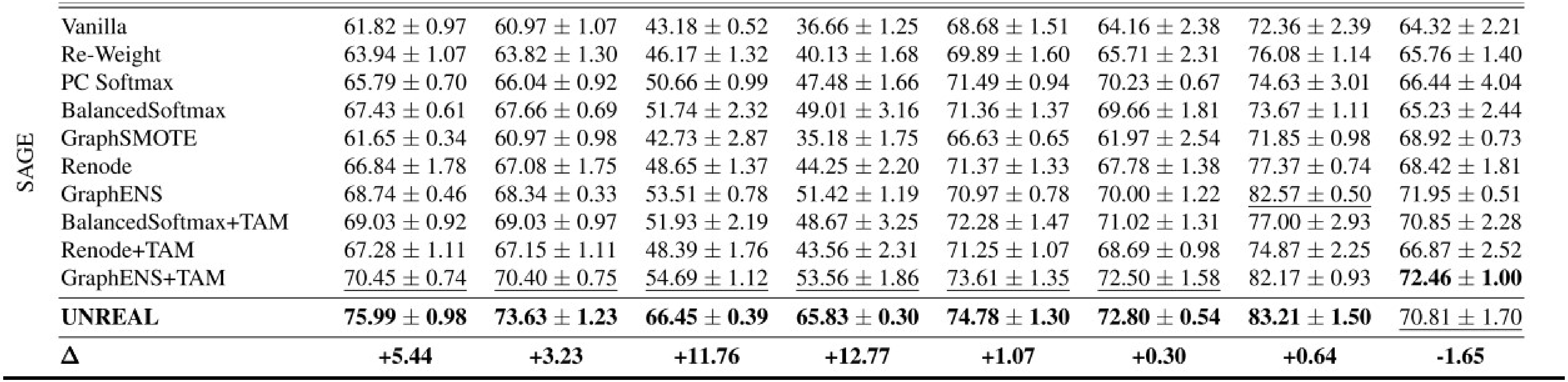
	Dataset	Cora		CiteSeer		PubMed		Amazon-Computers	
	Imbalance Ratio ($\rho = 10$)	bAcc.	F1	bAcc.	F1	bAcc.	F1	bAcc.	F1
GCN	Vanilla	62.82 ± 1.43	61.67 ± 1.59	38.72 ± 1.88	28.74 ± 3.21	65.64 ± 1.72	56.97 ± 3.17	80.01 ± 0.71	71.56 ± 0.81
	Re-Weight	65.36 ± 1.15	64.97 ± 1.39	44.69 ± 1.78	38.61 ± 2.37	69.06 ± 1.84	64.08 ± 2.97	80.93 ± 1.30	73.99 ± 2.20
	PC Softmax	68.04 ± 0.82	67.84 ± 0.81	50.18 ± 0.55	46.14 ± 0.14	72.46 ± 0.80	70.27 ± 0.94	81.54 ± 0.76	73.30 ± 0.51
	BalancedSoftmax	69.98 ± 0.58	68.68 ± 0.55	55.52 ± 0.97	53.74 ± 1.42	73.73 ± 0.89	71.53 ± 1.06	81.46 ± 0.74	74.31 ± 0.51
	GraphSMOTE	66.39 ± 0.56	65.49 ± 0.93	44.87 ± 1.12	39.20 ± 1.62	67.91 ± 0.64	62.68 ± 1.92	79.48 ± 0.47	72.63 ± 0.76
	Renode	67.03 ± 1.41	67.16 ± 1.67	43.47 ± 2.22	37.52 ± 3.10	71.40 ± 1.42	67.27 ± 2.96	81.89 ± 0.77	73.13 ± 1.60
	GraphENS	70.89 ± 0.71	70.90 ± 0.81	56.57 ± 0.98	55.29 ± 1.33	72.13 ± 1.04	70.72 ± 1.07	82.40 ± 0.39	74.26 ± 1.05
	BalancedSoftmax+TAM	69.94 ± 0.45	69.54 ± 0.47	56.73 ± 0.71	56.15 ± 0.78	74.62 ± 0.97	72.25 ± 1.30	82.36 ± 0.67	72.94 ± 1.43
	Renode+TAM	68.26 ± 1.84	68.11 ± 1.97	46.20 ± 1.17	39.96 ± 2.76	72.63 ± 2.03	68.28 ± 3.30	80.36 ± 1.19	72.51 ± 0.68
	GraphENS+TAM	71.69 ± 0.36	72.14 ± 0.51	58.01 ± 0.68	56.32 ± 1.03	74.14 ± 1.42	72.42 ± 1.39	81.02 ± 0.99	70.78 ± 1.72
	UNREAL	**78.33 ± 1.04**	**76.44 ± 1.06**	**65.63 ± 1.38**	**64.94 ± 1.38**	**75.35 ± 1.41**	**73.65 ± 1.43**	**85.08 ± 0.38**	**75.27 ± 0.23**
	Δ	**+6.64**	**+4.30**	**+7.62**	**+8.62**	**+1.21**	**+1.23**	**+2.68**	**+0.96**

Experiments



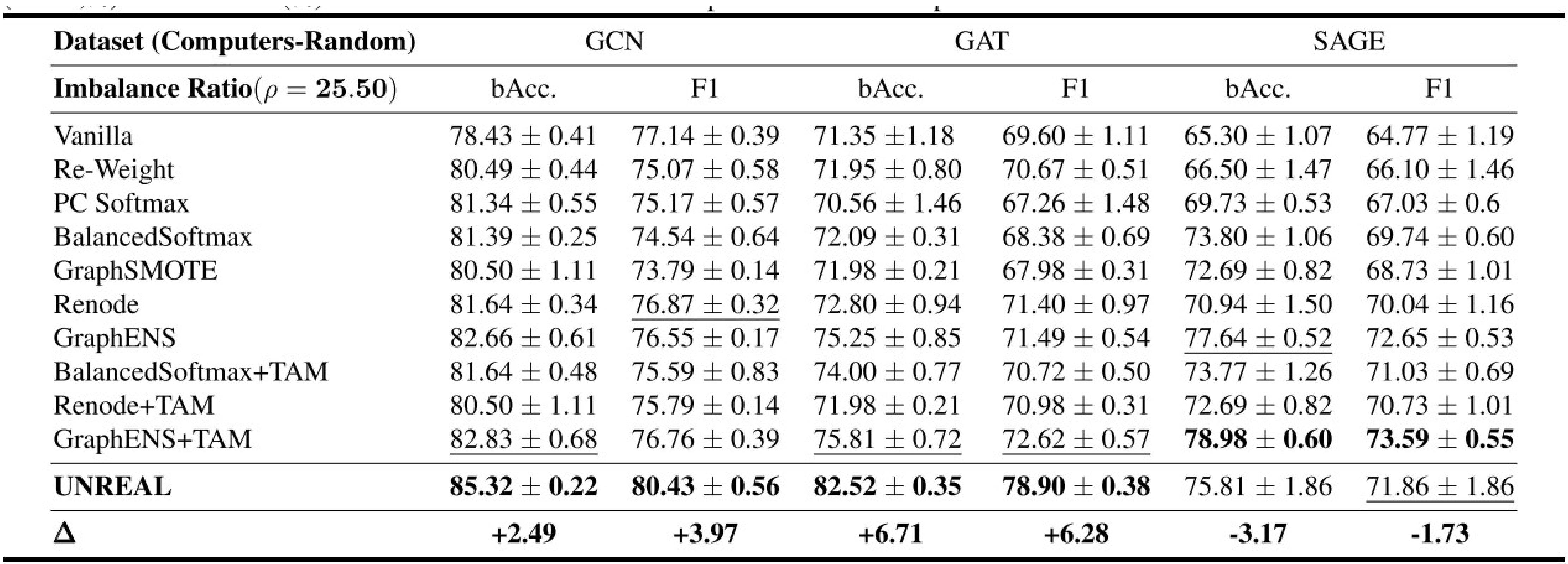
GAT	Vanilla	62.33 ± 1.56	61.82 ± 1.84	38.84 ± 1.13	31.25 ± 1.64	64.60 ± 1.64	55.24 ± 2.80	79.04 ± 1.60	70.00 ± 2.50
	Re-Weight	66.87 ± 0.97	66.62 ± 1.13	45.47 ± 2.35	40.60 ± 2.98	68.10 ± 2.85	63.76 ± 3.54	80.38 ± 0.66	69.99 ± 0.76
	PC Softmax	66.69 ± 0.79	66.04 ± 1.10	50.78 ± 1.66	48.56 ± 2.08	72.88 ± 0.83	71.09 ± 0.89	79.43 ± 0.94	71.33 ± 0.86
	BalancedSoftmax	67.89 ± 0.36	67.96 ± 0.41	54.78 ± 1.25	51.83 ± 2.11	72.30 ± 1.20	69.30 ± 1.79	82.02 ± 1.19	72.94 ± 1.54
	GraphSMOTE	66.71 ± 0.32	65.01 ± 1.21	45.68 ± 0.93	38.96 ± 0.97	67.43 ± 1.23	61.97 ± 2.54	79.38 ± 1.97	69.76 ± 2.31
	Renode	67.33 ± 0.79	68.08 ± 1.16	44.48 ± 2.06	37.93 ± 2.87	69.93 ± 2.10	65.27 ± 2.90	76.01 ± 1.08	66.72 ± 1.42
	GraphENS	70.45 ± 1.25	69.87 ± 1.32	51.45 ± 1.28	47.98 ± 2.08	73.15 ± 1.24	71.90 ± 1.03	81.23 ± 0.74	71.23 ± 0.42
	BalancedSoftmax+TAM	69.16 ± 0.27	69.39 ± 0.37	56.30 ± 1.25	53.87 ± 1.14	73.50 ± 1.24	71.36 ± 1.99	75.54 ± 2.09	66.69 ± 1.44
	Renode+TAM	67.50 ± 0.67	68.06 ± 0.96	45.12 ± 1.41	39.29 ± 1.79	70.66 ± 2.13	66.94 ± 3.54	74.30 ± 1.13	66.13 ± 1.75
	GraphENS+TAM	70.15 ± 0.18	70.00 ± 0.40	56.15 ± 1.13	54.31 ± 1.68	73.45 ± 1.07	72.10 ± 0.36	81.07 ± 1.03	71.27 ± 1.98
	UNREAL	**78.91 ± 0.59**	**75.99 ± 0.47**	**64.10 ± 1.49**	**63.44 ± 1.47**	**74.68 ± 1.43**	**72.78 ± 0.89**	**85.62 ± 0.44**	**75.34 ± 0.99**
	Δ	**+8.46**	**+5.99**	**+7.80**	**+9.13**	**+1.23**	**+0.68**	**+3.60**	**+2.40**

Experiments



SAGE	Vanilla	61.82 ± 0.97	60.97 ± 1.07	43.18 ± 0.52	36.66 ± 1.25	68.68 ± 1.51	64.16 ± 2.38	72.36 ± 2.39	64.32 ± 2.21
	Re-Weight	63.94 ± 1.07	63.82 ± 1.30	46.17 ± 1.32	40.13 ± 1.68	69.89 ± 1.60	65.71 ± 2.31	76.08 ± 1.14	65.76 ± 1.40
	PC Softmax	65.79 ± 0.70	66.04 ± 0.92	50.66 ± 0.99	47.48 ± 1.66	71.49 ± 0.94	70.23 ± 0.67	74.63 ± 3.01	66.44 ± 4.04
	BalancedSoftmax	67.43 ± 0.61	67.66 ± 0.69	51.74 ± 2.32	49.01 ± 3.16	71.36 ± 1.37	69.66 ± 1.81	73.67 ± 1.11	65.23 ± 2.44
	GraphSMOTE	61.65 ± 0.34	60.97 ± 0.98	42.73 ± 2.87	35.18 ± 1.75	66.63 ± 0.65	61.97 ± 2.54	71.85 ± 0.98	68.92 ± 0.73
	Renode	66.84 ± 1.78	67.08 ± 1.75	48.65 ± 1.37	44.25 ± 2.20	71.37 ± 1.33	67.78 ± 1.38	77.37 ± 0.74	68.42 ± 1.81
	GraphENS	68.74 ± 0.46	68.34 ± 0.33	53.51 ± 0.78	51.42 ± 1.19	70.97 ± 0.78	70.00 ± 1.22	82.57 ± 0.50	71.95 ± 0.51
	BalancedSoftmax+TAM	69.03 ± 0.92	69.03 ± 0.97	51.93 ± 2.19	48.67 ± 3.25	72.28 ± 1.47	71.02 ± 1.31	77.00 ± 2.93	70.85 ± 2.28
	Renode+TAM	67.28 ± 1.11	67.15 ± 1.11	48.39 ± 1.76	43.56 ± 2.31	71.25 ± 1.07	68.69 ± 0.98	74.87 ± 2.25	66.87 ± 2.52
	GraphENS+TAM	70.45 ± 0.74	70.40 ± 0.75	54.69 ± 1.12	53.56 ± 1.86	73.61 ± 1.35	72.50 ± 1.58	82.17 ± 0.93	$\mathbf{72.46 \pm 1.00}$
	UNREAL	$\mathbf{75.99 \pm 0.98}$	$\mathbf{73.63 \pm 1.23}$	$\mathbf{66.45 \pm 0.39}$	$\mathbf{65.83 \pm 0.30}$	$\mathbf{74.78 \pm 1.30}$	$\mathbf{72.80 \pm 0.54}$	$\mathbf{83.21 \pm 1.50}$	70.81 ± 1.70
	Δ	**+5.44**	**+3.23**	**+11.76**	**+12.77**	**+1.07**	**+0.30**	**+0.64**	**-1.65**

Experiments



Dataset (Computers-Random)	GCN		GAT		SAGE	
Imbalance Ratio$(\rho = \mathbf{25.50})$	bAcc.	F1	bAcc.	F1	bAcc.	F1
Vanilla	78.43 ± 0.41	77.14 ± 0.39	71.35 ±1.18	69.60 ± 1.11	65.30 ± 1.07	64.77 ± 1.19
Re-Weight	80.49 ± 0.44	75.07 ± 0.58	71.95 ± 0.80	70.67 ± 0.51	66.50 ± 1.47	66.10 ± 1.46
PC Softmax	81.34 ± 0.55	75.17 ± 0.57	70.56 ± 1.46	67.26 ± 1.48	69.73 ± 0.53	67.03 ± 0.6
BalancedSoftmax	81.39 ± 0.25	74.54 ± 0.64	72.09 ± 0.31	68.38 ± 0.69	73.80 ± 1.06	69.74 ± 0.60
GraphSMOTE	80.50 ± 1.11	73.79 ± 0.14	71.98 ± 0.21	67.98 ± 0.31	72.69 ± 0.82	68.73 ± 1.01
Renode	81.64 ± 0.34	76.87 ± 0.32	72.80 ± 0.94	71.40 ± 0.97	70.94 ± 1.50	70.04 ± 1.16
GraphENS	82.66 ± 0.61	76.55 ± 0.17	75.25 ± 0.85	71.49 ± 0.54	77.64 ± 0.52	72.65 ± 0.53
BalancedSoftmax+TAM	81.64 ± 0.48	75.59 ± 0.83	74.00 ± 0.77	70.72 ± 0.50	73.77 ± 1.26	71.03 ± 0.69
Renode+TAM	80.50 ± 1.11	75.79 ± 0.14	71.98 ± 0.21	70.98 ± 0.31	72.69 ± 0.82	70.73 ± 1.01
GraphENS+TAM	82.83 ± 0.68	76.76 ± 0.39	75.81 ± 0.72	72.62 ± 0.57	**78.98 ± 0.60**	**73.59 ± 0.55**
UNREAL	**85.32 ± 0.22**	**80.43 ± 0.56**	**82.52 ± 0.35**	**78.90 ± 0.38**	75.81 ± 1.86	71.86 ± 1.86
Δ	**+2.49**	**+3.97**	**+6.71**	**+6.28**	**-3.17**	**-1.73**

THANKS