TAM: Topology-Aware Margin Loss for Class-Imbalanced Node Classification

Jaeyun Song, Joonhyung Park* , Eunho Yang*

*(*equal contribution)*

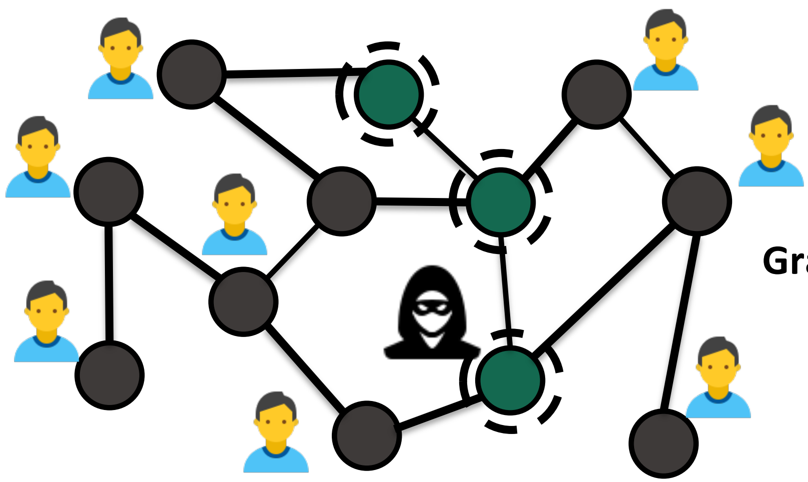
ICML 2022

Graduate School of AI, KAIST

Machine Learning & Intelligence Laboratory

Introduction

- Nodes in real-world graphs are inherently class-imbalanced
  - e.g. social networks, commercial graphs, chemical molecules
- Learning reliable node representations under class-imbalanced graphs is challenging due to the interactive nature of graph data
- Diverse strategies to handle imbalance in graphs have been proposed



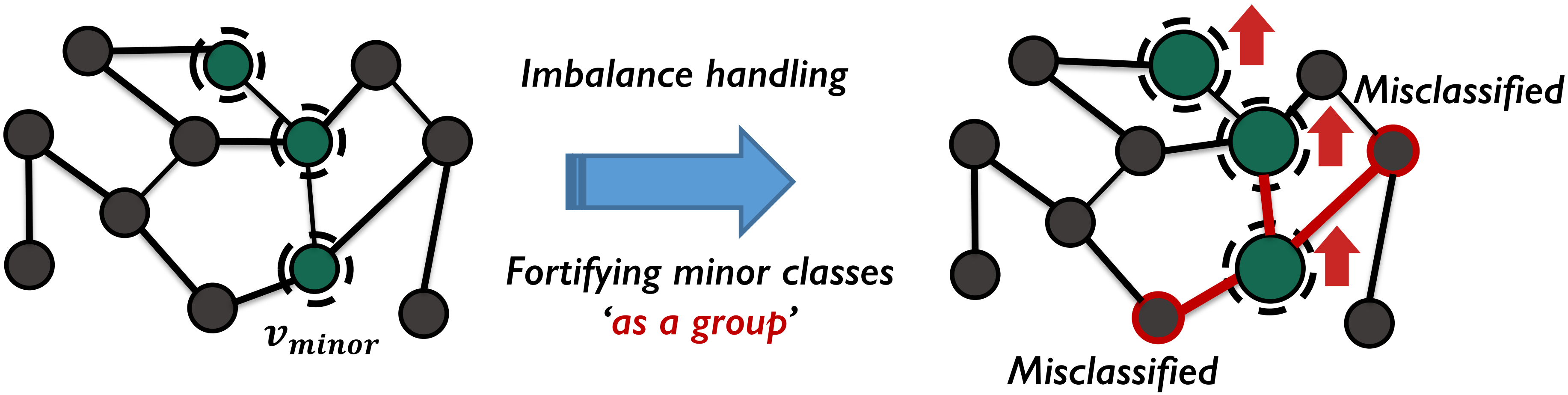
Social Networks

Graph Neural Networks can be biased toward major classes

Imbalance Handling in Graph-Structured Data

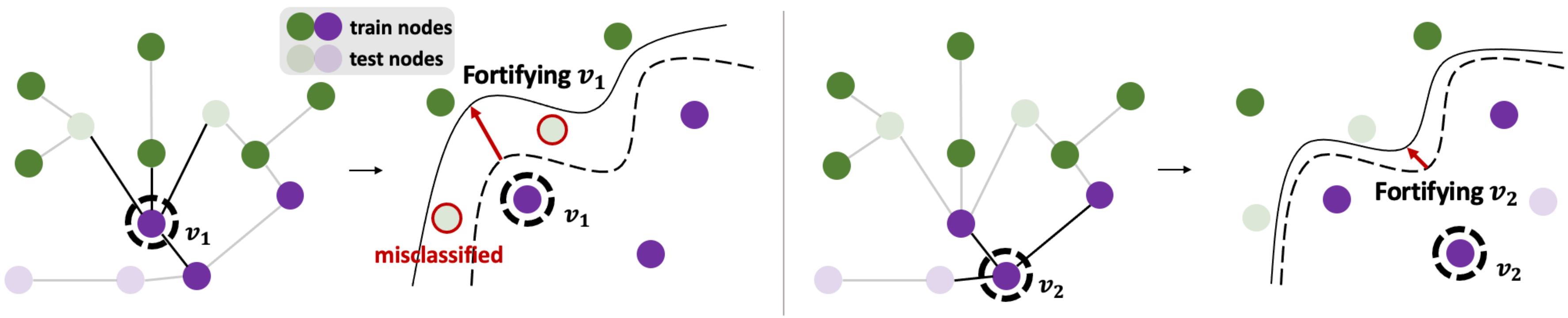
- Existing works have in common that they regard the minor class nodes '**as a group**' and **fortify** minor classes in their own way (*e.g.* SMOTE, re-weight, logit adjustment)
- These approaches effectively mitigate the model bias for major classes while unavoidably increasing ***false positives*** for major class nodes

  Misclassifying a major class node as a minor class
- Given the message interactions of GNNs, certain compensated minor nodes could significantly degrade the performance of other classes



Topological Positions of False Positives

- We hypothesize that weighted minor nodes having **high connectivity rates with other (major) classes** induce excessive false positives



(a) Reinforcing anomalously connected minor node v_1 **(b) Reinforcing normal minor node v_2**

- First we define **anomalously connected node set** V^* as:

Neighbor Label Distribution (Local Topology)

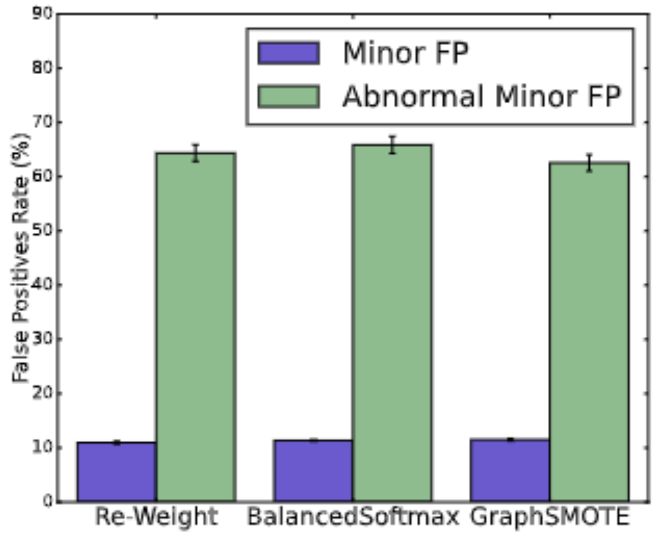
$$V^* = \{v \in V^L | \max_{t \in |\mathcal{Y}| \setminus \{y_v\}} \frac{\mathcal{D}_{v,t}}{\mathcal{C}_{y_v,t}} > 1\}$$

Class-wise Connectivity Matrix

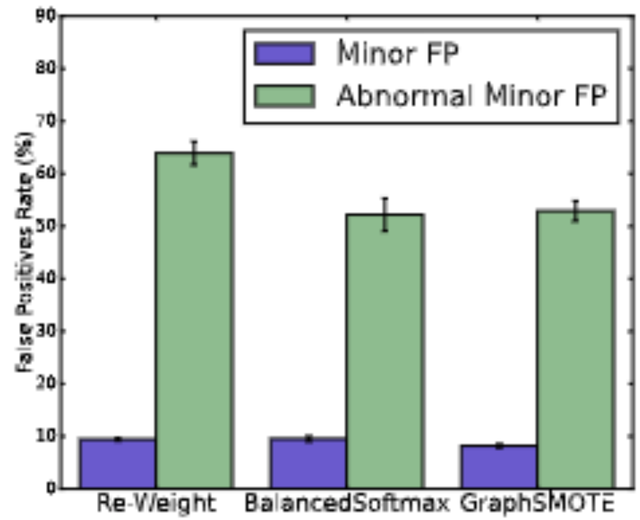
A set of nodes that has more connections with other classes compared to class-averaged level

Topological Positions of False Positives

- We compute the following two ratios
- $\frac{FP(\mathcal{N}(v) \cap V_{major} \mid v \in V^*_{minor})}{|\mathcal{N}(v) \cap V_{major} \mid v \in V^*_{minor}|}$: the probability of being false positives when major nodes are connected with anomalous minor nodes (**Abnormal Minor FP** in Figure)
- $\frac{FP(V_{major})}{|V_{major}|}$: the average probability of being false positives (**Minor FP** in Figure)



(b) CiteSeer (ρ=10)



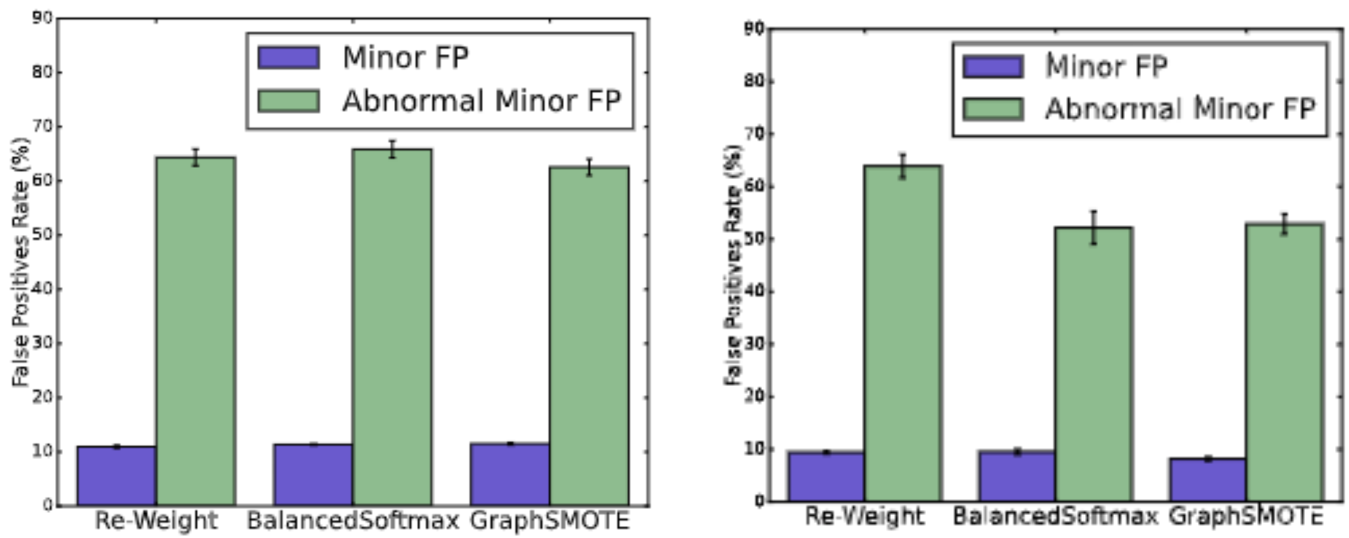
(d) Wisconsin (ρ=11.63)

** $FP(\cdot)$: a function that counts the number of false positives*

Topological Positions of False Positives

- We confirm that false positives on minor classes are intensively concentrated around minor nodes that have **higher connectivity with other classes**
- False positives due to fortifying minor nodes do NOT appear uniformly on graph

Minor nodes deviated from general connectivity patterns induce excessive false positive cases



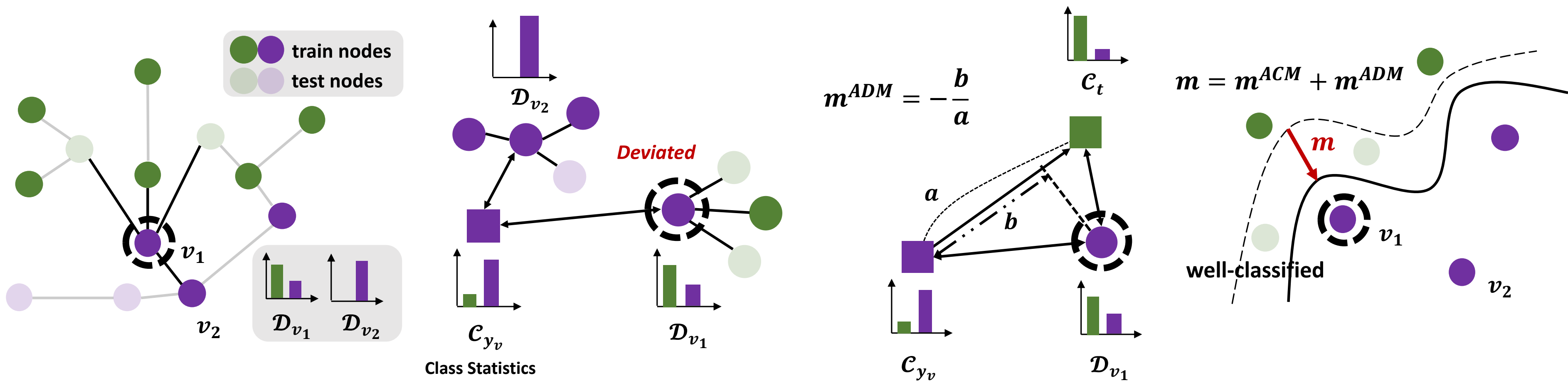
(b) CiteSeer (ρ=10) (d) Wisconsin (ρ=11.63)

** $FP(\cdot)$: a function that counts the number of false positives*

MLILAB Machine Learning & Intelligence

Method Overview

- To decrease the false positive cases, we propose an effective **margin adjustment**
- TAM determines the intensity of imbalance compensation based on **local topology**
- TAM consists of two core components: **ACM** and **ADM**



(a) Compute neighbor label distribution

(b) Identify pattern-deviated nodes (Determining the margin m^{ACM})

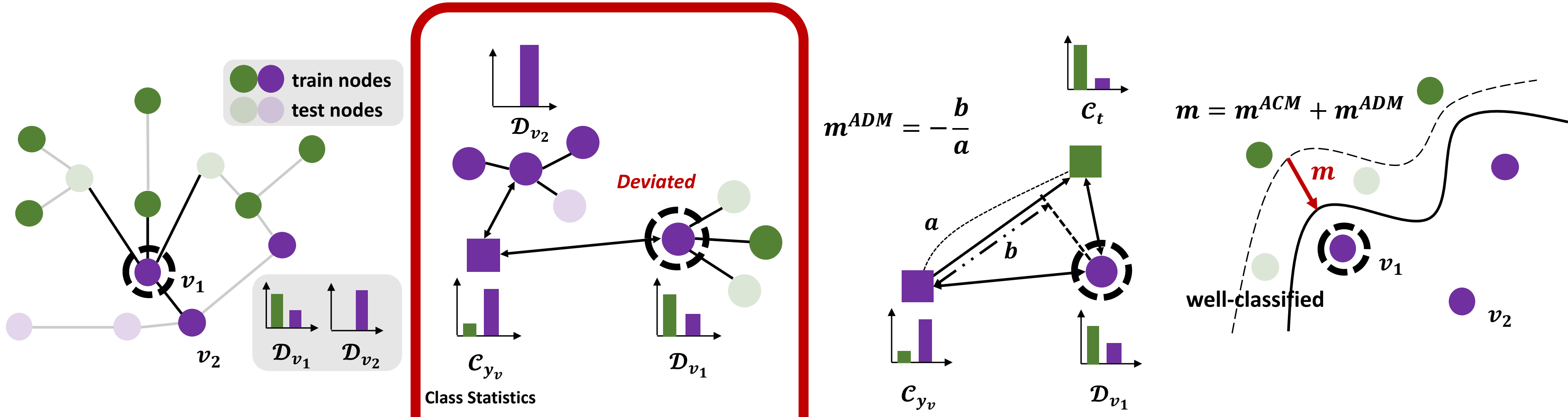
(c) Specify confusing classes (Determining the margin m^{ADM})

(d) Adjust margins

MLILAB Machine Learning & Intelligence

Method: Anomalous Connectivity Margin (ACM)

- **Deviated nodes from class-homophily tendency** would be risky in imbalance handling process
- ACM is designed to **reduce the learning signals** of deviated nodes
- ACM **decreases the margin** if a node is deviated from the connectivity pattern



(a) Compute neighbor label distribution

(b) Identify pattern-deviated nodes (Determining the margin m^{ACM})

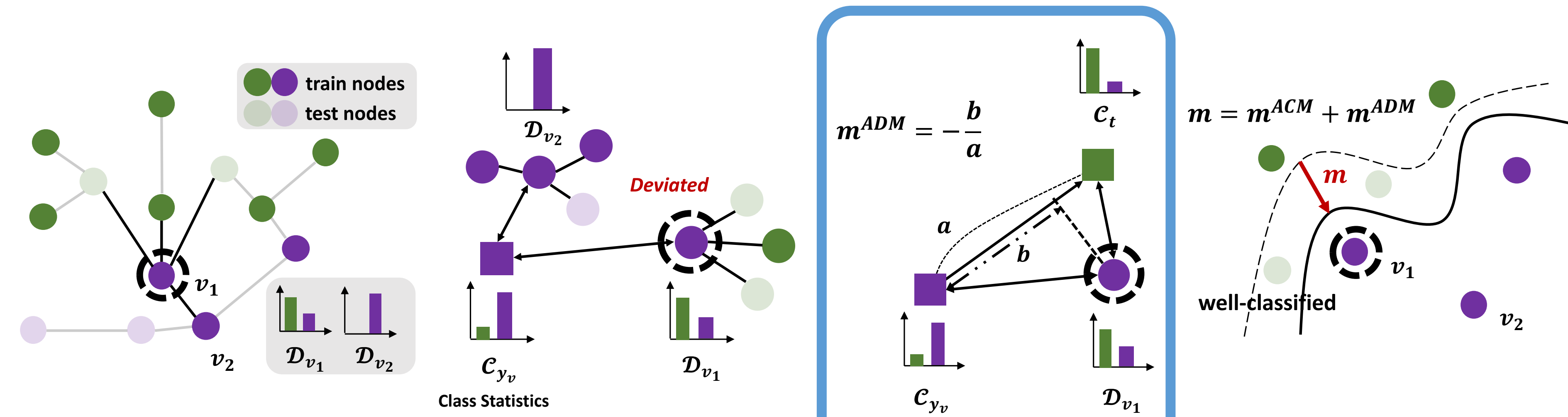
(c) Specify confusing classes (Determining the margin m^{ADM})

(d) Adjust margins

MLILAB Machine Learning & Intelligence

Method: Anomalous Distribution-Aware Margin (ADM)

- ACM does not recognize whether a deviated node is confused with other classes or simply an outlier
- ADM is devised to **identify indistinguishable nodes**
- ADM complementarily adjusts the margins according to **the relative closeness**



(a) Compute neighbor label distribution

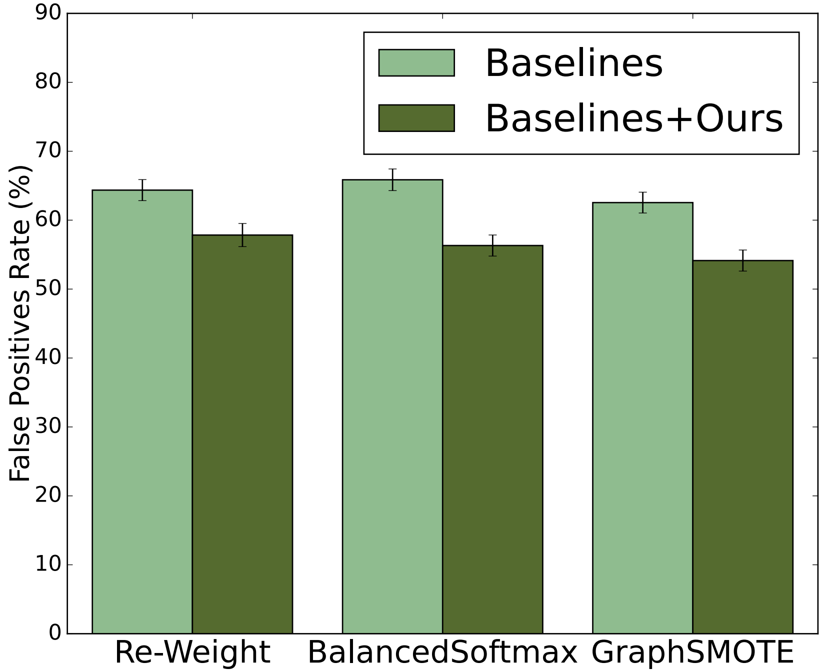
(b) Identify pattern-deviated nodes (Determining the margin m^{ACM})

(c) Specify confusing classes (Determining the margin m^{ADM})

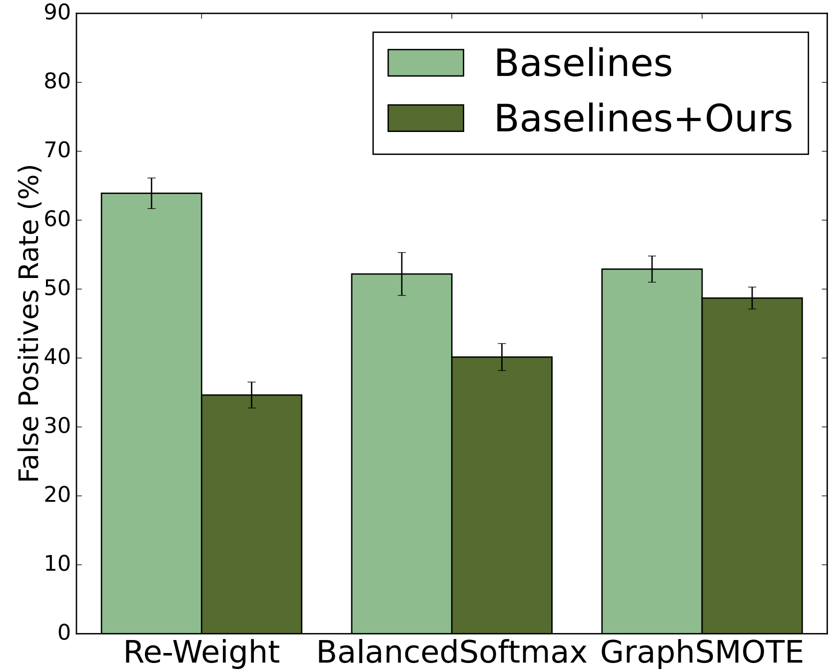
(d) Adjust margins

Experiment: Node & Neighbor Memorization

- Combining TAM decreases the false positives near anomalously connected nodes by adjusting margins of these nodes.



Baseline with ours in CiteSeer



Baselines with ours in Wisconsin

MLILAB Machine Learning & Intelligence

Experiment: Homophilous Graphs

- Imbalance handling methods combined with TAM show the best performance
- TAM improves the performance over various types of imbalance handling methods
- The rationale of these results is that TAM **identifies non-typically connected nodes** and adjust margins

	Dataset	Cora		CiteSeer		PubMed	
	Imbalance Ratio ($\rho = 10$)	bAcc.	F1	bAcc.	F1	bAcc.	F1
GCN	Cross Entropy	60.95 ± 1.22	59.30 ± 1.66	38.21 ± 1.12	29.40 ± 1.97	65.21 ± 1.40	55.43 ± 2.79
GCN	Re-Weight	65.52 ± 0.84	65.54 ± 1.20	44.52 ± 1.22	38.85 ± 1.62	70.17 ± 1.25	66.37 ± 1.73
GCN	PC Softmax	67.79 ± 0.92	67.39 ± 1.08	49.81 ± 1.12	45.55 ± 1.26	70.20 ± 0.60	68.83 ± 0.73
GCN	DR-GCN	60.17 ± 0.83	59.31 ± 0.97	42.64 ± 0.75	38.22 ± 1.22	65.51 ± 0.81	64.95 ± 0.53
GCN	GraphSMOTE	66.29 ± 0.93	66.30 ± 1.25	44.40 ± 1.27	39.10 ± 1.78	68.51 ± 1.14	62.63 ± 2.39
GCN	BalancedSoftmax	68.46 ± 0.67	68.41 ± 0.80	53.70 ± 1.40	50.73 ± 1.64	72.97 ± 0.80	70.80 ± 1.11
GCN	**+ TAM**	69.90 ± 0.73	69.89 ± 0.89	55.54 ± 1.40	54.18 ± 1.69	**74.13** ± 0.70	**73.27** ± 0.67
GCN	ReNode	67.61 ± 0.77	67.27 ± 0.91	47.78 ± 1.67	42.51 ± 2.30	71.59 ± 1.70	66.56 ± 2.90
GCN	**+ TAM**	67.18 ± 1.32	67.39 ± 1.62	48.36 ± 1.63	42.48 ± 2.10	71.00 ± 1.86	67.18 ± 3.42
GCN	GraphENS	70.31 ± 0.51	70.30 ± 0.65	55.42 ± 1.74	53.85 ± 2.00	71.89 ± 0.80	71.07 ± 0.66
GCN	**+ TAM**	**71.52** ± 0.30	**71.71** ± 0.45	**57.47** ± 1.56	**56.23** ± 1.87	74.01 ± 0.73	72.41 ± 0.94

Experiment: Heterophilous Graphs

- TAM also shows superior performance than baselines on heterophilous graphs
- TAM could **identify the outliers nodes** by using the class-wise connectivity pattern and **reduce the false positives** stemming from these nodes

	Dataset	Chameleon		Squirrel		Wisconsin	
	Imbalance Ratio	($\rho = 5$) bAcc.	F1	($\rho = 5$) bAcc.	F1	($\rho = 11.63$) bAcc.	F1
GAT	Cross Entropy	34.33 ± 0.74	31.54 ± 0.95	24.89 ± 0.37	21.33 ± 0.52	32.15 ± 2.72	30.92 ± 2.76
	Re-Weight	39.63 ± 0.49	39.08 ± 0.50	26.49 ± 0.41	25.92 ± 0.41	42.15 ± 2.33	37.66 ± 2.27
	PC Softmax	41.47 ± 0.78	40.51 ± 0.89	27.31 ± 0.51	26.74 ± 0.50	41.89 ± 3.95	38.03 ± 3.35
	DR-GCN	36.85 ± 0.77	34.61 ± 0.62	25.40 ± 0.43	22.83 ± 0.59	33.93 ± 2.34	31.75 ± 2.50
	GraphENS	40.66 ± 1.13	39.49 ± 1.10	26.87 ± 0.43	26.78 ± 0.41	40.93 ± 2.78	37.43 ± 2.74
	BalancedSoftmax	41.47 ± 0.71	40.52 ± 0.78	26.66 ± 0.39	25.97 ± 0.35	41.20 ± 3.08	37.93 ± 2.99
	+ TAM	42.56 ± 0.59	41.40 ± 0.74	27.75 ± 0.44	27.23 ± 0.45	**48.44** ± 3.32	**43.71** ± 2.91
	ReNode	40.41 ± 0.56	39.85 ± 0.60	26.89 ± 0.45	26.40 ± 0.46	40.88 ± 2.84	37.13 ± 2.74
	+ TAM	41.53 ± 0.35	40.76 ± 0.50	26.53 ± 0.40	26.00 ± 0.42	46.64 ± 3.35	41.60 ± 3.02
	GraphSMOTE	42.27 ± 0.51	41.43 ± 0.54	28.17 ± 0.56	27.38 ± 0.66	40.77 ± 2.24	38.96 ± 2.48
	+ TAM	**42.83** ± 0.82	**42.26** ± 0.83	**28.44** ± 0.33	**28.02** ± 0.37	41.82 ± 2.94	38.23 ± 3.13

Conclusion

- We found that the adjacent major nodes of anomalously connected minor nodes are prone to be misclassified as the minor class
- We propose TAM that adjusts margin according the extent of deviation from connectivity patterns and relative closeness to self class compared the target class
- We show that combing TAM improves the performance on both homophilous and heterophilous graphs