

# GIER: Addressing Class Imbalance in GNNs Through Experience Replay

Liu Yang<sup>1</sup>, Chuyao Liu<sup>1</sup>, Zidong Wang<sup>1</sup>, Tingxuan Chen<sup>1\*</sup>, Mengni Chen<sup>1</sup>, Hongyu Zhang<sup>2</sup>

<sup>1</sup>School of Computer Science and Engineering, Central South University, Changsha, China

<sup>2</sup>Management Science and Engineering, Central South University, Changsha, China  
{yangliu, 244711047, zdwang, chentingxuan, chenmengni, hyzhang}@csu.edu.cn

## Abstract

The prevalent class imbalance in real-world graphs significantly affects the performance of Graph Neural Networks (GNNs). Existing methods for analyzing graph imbalance ignore the influence of minority nodes during the dynamic model training process, resulting in performance limitations. In this paper, we focus on minority class information during model training, identifying and defining the minority class forgetting phenomenon that exists in graph imbalanced method training processes. To address this issue, we propose Graph Imbalance Experience Replay (GIER) framework. On one hand, the method enhances the model’s ability to mine minority node information in historical data, thereby achieving feature completion for minority class nodes. On the other hand, the proposed short-term confidence mechanism allows the model to adaptively calibrate the topological relationships in high-confidence nodes, thereby mitigating the model’s tendency to propagate erroneous information about minority classes during training. GIER is a unified framework consisting of two synergistic components: Long-term Subgraph Memory (LSM) constructs multi-period feature-representative subgraphs to address distribution imbalance, and Short-term Confidence Calibration (SCC) dynamically reconstructs graph topology through degree-aware node selection and confidence-based filtering to address topological imbalance. The extensive experimental results demonstrate that GIER effectively improves the classification performance of GNNs on imbalanced graphs, achieving up to a 3.44% improvement in BA<sub>cc</sub> over the state-of-the-art, and is particularly effective in extreme scenarios with very small minority classes.

## Introduction

Graph is a powerful and prevalent data structure for representing complex relational systems, such as social networks (Ga et al. 2025), citation networks (Lachaud, Conde-Cespedes, and Trocan 2022), biomolecular networks (Burkhart et al. 2023), and knowledge graphs (Liu et al. 2025). Graph Neural Networks (GNNs) achieve remarkable progress in node classification across diverse graph-structured applications due to their strong ability to model on graphs (Fang et al. 2022; Zheng et al. 2024; Pham

et al. 2025; Zheng et al. 2025). Imbalance is a pervasive issue, and many real-world graphs display class imbalance, in which certain classes contain significantly more nodes than others (Pu et al. 2025; Ma et al. 2025).

Existing approaches to address graph imbalance problems can be categorized into two main aspects. One is **distribution imbalance** commonly encountered, and the other is **topological imbalance** which exists uniquely in graph-structured data. Distribution imbalance refers to the unequal numbers of different classes within the dataset (Zhao, Zhang, and Wang 2021; Park, Song, and Yang 2021). Topological imbalance refers to the unequal distribution of connections and relationships between nodes in an imbalanced graph (Liu et al. 2024; Yang et al. 2025).

Intuitively, during the training process, models tend to focus more attention on the majority class nodes while neglecting information from minority class nodes. Prior research has demonstrated that sample forgetting phenomena are closely correlated with model performance in neural networks (Toneva et al. 2019; Stern, Yaacoby, and Weinshall 2025). Specifically, samples with high forgetting frequencies often correspond to critical support regions of decision boundaries, which learn quality directly impacts generalization performance (Zhao et al. 2025; Rohlf’s 2025). Inspired by this finding, we conduct a deep investigation into node forgetting phenomena in graphs, revealing that forgetting behaviors exist not only in traditional neural networks but also exhibit significant class-biased characteristics when GNNs process imbalanced graph data.

As illustrated in Figure 1, our empirical analysis on the Cora dataset reveals that existing methods for addressing both distribution imbalance and topological imbalance suffer from excessive forgetting in minority class nodes. While majority class nodes maintain a stable forgetting frequency of 4-5 times on average, the minority class nodes exhibit significantly higher forgetting rates: distribution imbalance methods such as GraphENS (Park, Song, and Yang 2021) result in an average of 9.84 forgetting events, and topological imbalance methods such as BAT (Liu et al. 2024) result in an average of 7.21 forgetting events. This indicates a systematic learning bias against minority nodes despite data or structural optimizations. These high-frequency forgotten minority nodes, which serve as critical support points for decision boundaries, directly constrain model performance

\*Corresponding author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Forget Count Distribution for GNN models  
(imbalance-ratio=10)

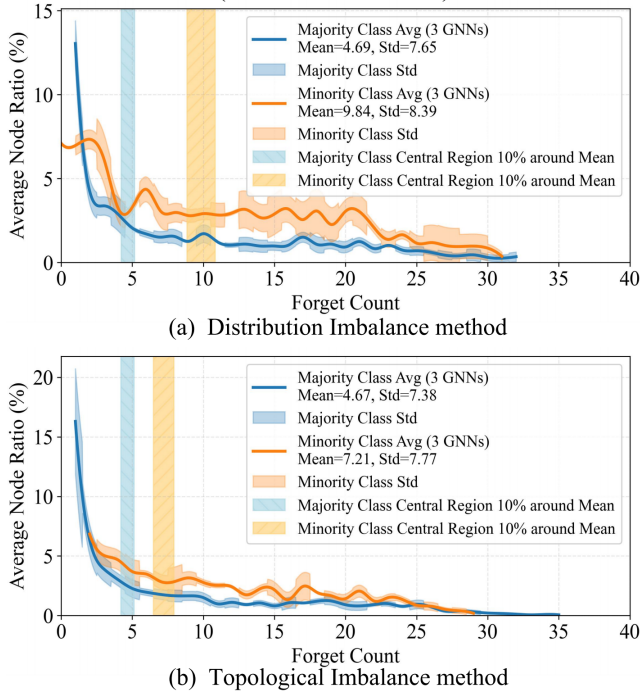


Figure 1: **Comparative analysis of node forgetting phenomena in imbalanced graph learning.** We recorded the total number of “sample forgetting events” over 300 epochs for nodes that have already been learned during the training process of three GNN models (GCN, GAT, and SAGE) on Cora dataset.

through unstable learning states. We address node class imbalance through explicit forgetting balance mechanisms to stabilize decision boundaries and improve classification performance.

In this paper, we theoretically analyze of forgetting-performance relationship and systematically investigate node forgetting phenomena in GNN training dynamics, and we propose Graph Imbalance Experience Replay (GIER), a dynamic framework that simultaneously addresses both distribution and topological imbalances by explicitly managing forgetting dynamics. Aiming to minimize inter-class forgetting disparities to stabilize decision boundary formation, we introduce two complementary mechanisms: 1) **Long-term Subgraph Memory (LSM)** mechanism that reduces minority class node forgetting through multi-epoch memory subgraphs by experience replay, effectively alleviating frequent forgetting during training, 2) **Short-term Confidence Calibration (SCC)** mechanism that dynamically adjusts topological structures through confidence-based neighborhood reconstruction and weighted message passing, mitigating error signal amplification caused by topological imbalance. Our memory-enhanced framework reduces the probability of forgetting for the minority class in imbalanced graphs, decreasing from 1.5 times that of the majority class to 1.1

times. This reduction alleviates the forgetting imbalance between the two classes and achieves an improvement of up to 3.44% in BAcc for classification tasks.

Our main contributions are summarized as follows:

- We are the first to reveal that GNNs exhibit systematic forgetting tendencies toward minority classes across both distribution balance methods and topological adjustment methods. To the best of our knowledge, GIER is the first method to simultaneously address both distribution imbalance and topological imbalance through explicit forgetting management.
- We design two complementary mechanisms: one is the LSM mechanism, which constructs minority class subgraphs using representative minority nodes from multiple training epochs to alleviate forgetting phenomena for minority classes; the other is the SCC mechanism, which reconstructs topological relationships for high-confidence nodes within individual epochs to prevent the propagation of erroneous information to minority classes in GNNs.
- Empirical results from three imbalanced semi-supervised benchmark datasets demonstrate that our model significantly outperforms seven baseline methods, particularly excelling in scenarios of extreme imbalance.

## Related Work

### Class Imbalance in Graph Neural Networks

Existing methods for addressing the performance degradation of GNNs on imbalanced graphs fall into two perspectives: distribution imbalance and topological imbalance.

**Distribution Imbalance.** Distribution Imbalance primarily focuses on balancing sample quantities. GraphSMOTE (Zhao, Zhang, and Wang 2021), GraphENS (Park, Song, and Yang 2021), and GraphSR (Zhou and Gong 2023) improve imbalanced sample quantities through methods such as generating synthetic nodes and selecting unlabeled nodes.

**Topological Imbalance.** Topological Imbalance targets the significant differences in graph structures, connectivity patterns, and degree distributions between minority and majority classes. ReNode (Chen et al. 2021) method adjusts node training weights based on topological boundaries. BAT (Liu et al. 2024) identifies two fundamental local topological phenomena: ambivalent message-passing (AMP) and distant message-passing (DMP), and modifies the topological structure.

The above methods address graph imbalance either in terms of distribution or topology. However, no method has simultaneously resolved both.

### Forgetting Phenomena & Experience Replay

Forgetting phenomena exist at different scales in deep learning (Wei et al. 2024; Zhang et al. 2024): sample-level forgetting, where model predictions change from correct to incorrect during training (Toneva et al. 2019), and catastrophic forgetting, which represents knowledge loss when learning new tasks (Cheng et al. 2025). Experience replay (Hassani,

Nikan, and Shami 2025) has been used to mitigate forgetting phenomena in neural networks and has also been adapted for various graph learning scenarios. As for experience replay in graph-structured data, the most basic method is to select representative nodes (Zhou and Cao 2021). The sparsified subgraph memory chooses subgraphs for storage based on node degree (Zhang, Song, and Tao 2022). Recent research, FTF-ER (Pang et al. 2024), combines feature and global topological information to select and preserve important nodes.

Inspired by methodologies for preserving critical information using experience replay to address the forgetting phenomenon, we mitigate the tendency of sample-level forgetting in minority classes of GNNs by preserving information during the training process.

### Problem Definition

In this section, we introduce the definition of class-imbalanced graphs learning of GNNs on imbalanced graphs, focusing on the performance of minority and majority classes in forgetting phenomenon. We are the first to analyze the imbalanced class distribution problem from the perspective of forgetting, thereby utilizing experience replay methods to improve imbalanced learning.

### Imbalanced Graphs

Consider a graph  $G = (V, E)$  with  $n$  nodes and  $m$  edges, node feature matrix  $X \in \mathbb{R}^{n \times d}$  represents node features. Adjacency matrix  $A \in \mathbb{R}^{n \times n}$  indicates edge connections, where  $A_{ij} = 1$  if edge  $(v_i, v_j)$  exists.

**Imbalance Ratio.** In a semi-supervised node classification setting, we consider a graph with  $n$  nodes. The node set  $\mathcal{V}$  is partitioned into labeled and unlabeled subsets  $\mathcal{V} = \mathcal{V}_L \cup \mathcal{V}_U$ , where  $\mathcal{V}_L \cap \mathcal{V}_U = \emptyset$ . Setting  $\mathcal{V}_L$  with labels  $\mathcal{Y}_L$ , leaving  $\mathcal{V}_U$  unlabeled, the labeled data exhibits class-wise organization as  $\{C_1, \dots, C_m\}$  where  $C_i$  denotes the  $i$ -th class membership. Class imbalance severity is measured by  $\rho = \frac{\max_i |C_i|}{\min_i |C_i|}$ .

### GNNs Classifier

We define the GNNs class output at the  $t$ -epoch as

$$\mathbf{Z}_{\text{out}}^t = \text{GNN}(X, A, \theta^t) \quad (1)$$

where  $\mathbf{Z}_{\text{out}}^t \in \mathbb{R}^{N \times C}$  is the predicted output for all nodes.  $\mathbf{z}_i \in \mathbb{R}^C$  is the predicted output for node  $i$ , i.e.,  $[z_{i,1}, z_{i,2}, \dots, z_{i,C}]$ . For each class  $c \in \{1, 2, \dots, C\}$ , softmax function calculates the probability of node  $i$  belongs to class  $c$  by

$$\hat{p}_{i,c} = \frac{e^{z_{i,c}}}{\sum_{j=1}^C e^{z_{i,j}}} \quad (2)$$

where  $\hat{\mathbf{p}}_i = [\hat{p}_{i,1}, \hat{p}_{i,2}, \dots, \hat{p}_{i,C}]$  is the predicted probability distribution for node  $i$ .  $\hat{\mathbf{P}}^t$  is the predicted probability distribution for all nodes in epoch  $t$ .

### Forgetting Phenomenon in Imbalanced Graphs

**Definition 1.1 Node Forgetting Event:** For node  $v_i$ , if it is predicted correctly in the  $t$ -th training round but predicted incorrectly in the  $(t + 1)$ -th round, a forgetting event is said

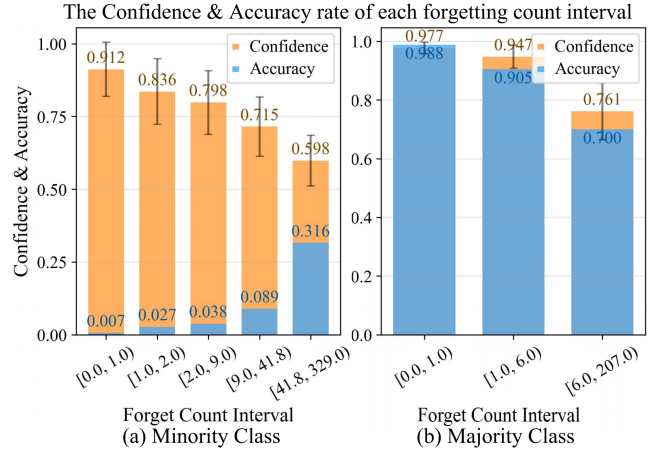


Figure 2: Forgetting Phenomenon In Imbalanced Graphs

to occur (Toneva et al. 2019). The forgetting count of node  $v_i$  throughout the training process is defined as:

$$\text{Forget\_Count}(v_i) = \sum_{t=1}^{T-1} \mathbb{I}[\hat{y}_i^{(t)} = y_i \wedge \hat{y}_i^{(t+1)} \neq y_i] \quad (3)$$

**Definition 1.2 Confidence:** We define the confidence of GNNs for node  $i$  at time  $t$  as the maximum probability of predicting different categories:

$$\text{confidence}_i^{(t)} = \max \hat{\mathbf{P}}_i^t \quad (4)$$

In general, the larger the value, the smaller the possibility that GNNs determine presently that the node belongs to another category.

### High Confidence Bias in Minority Class

We analyze the relationship between forgetting count, confidence, and accuracy for both majority and minority classes using GCN, a classical GNN model, on the PubMed dataset.

As shown in Figure 2, both classes exhibit a negative correlation between prediction confidence and forgetting count, indicating that low-confidence nodes are more frequently forgotten, and high-confidence nodes achieve a lower degree of forgetting frequency.

However, the confidence-accuracy relationship reveals a critical asymmetry. For majority classes, higher confidence correlates with higher accuracy, reflecting normal learning patterns due to abundant training samples. Conversely, minority classes show an inverse relationship that higher confidence often corresponds to lower accuracy. This counter-intuitive phenomenon stems from the model’s over-reliance on limited minority samples, leading to overconfident but incorrect predictions on high-confidence biased nodes. Meanwhile, easily-forgotten and low-confidence nodes from the minority class are typically located near decision boundaries; and these boundary nodes are essential for accurate minority classification but suffer from unstable learning due to insufficient representation during training.

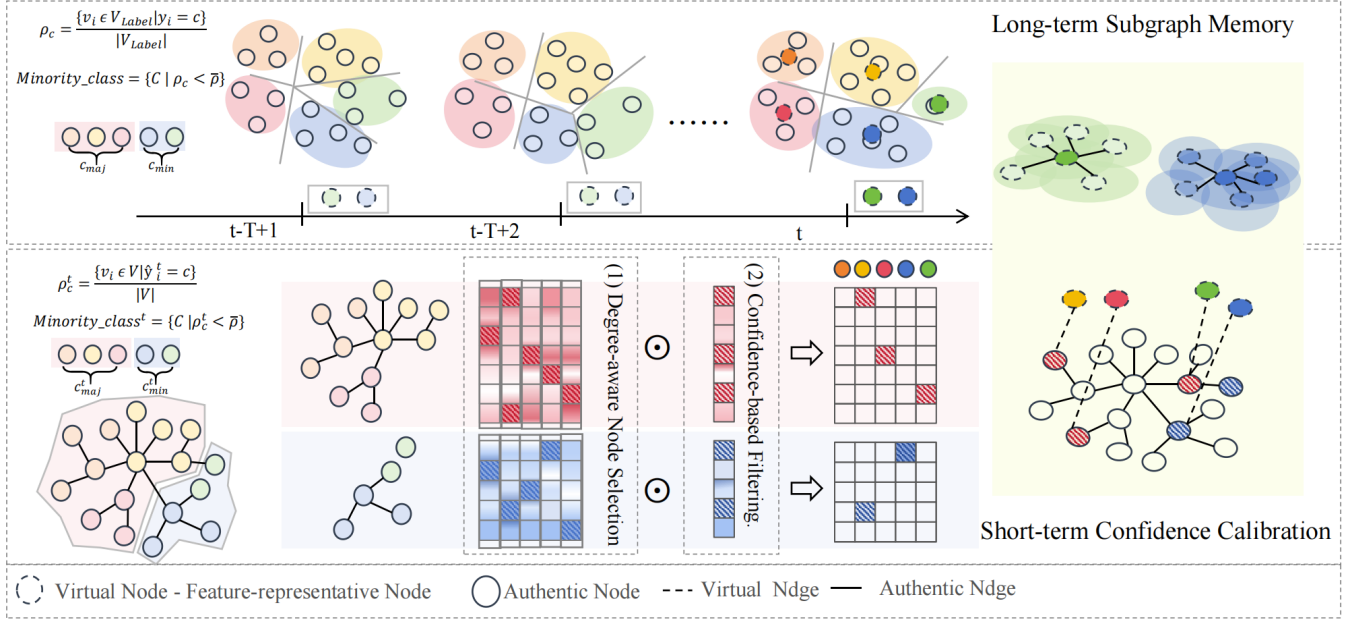


Figure 3: Graph Imbalance Experience Replay (GIER) Framework. (a) Long-term Subgraph Memory (LSM) records multiple-epochs minority nodes feature representation. (b) Short-term Confidence Calibration (SCC) dynamically updates the topological selection of high-confidence nodes.

## Methodology

As shown in Figure 3, we propose **Graph Imbalance Experience Replay (GIER)**. We design two mechanisms for distribution imbalance and topological imbalance on the graph, named Long-term Subgraph Memory Mechanism and Short-term Confidence Calibration Mechanism.

**Long-term Subgraph Memory (LSM)** implements a multiple-epochs memory mechanism that strategically retains historical representations of minority class nodes, effectively addressing distribution imbalance challenges inherent in graph-based learning scenarios. **Short-term Confidence Calibration (SCC)** incorporates a dynamic confidence calibration framework that adaptively modifies graph structural properties, thereby mitigating the propensity of GNNs for error propagation on minority classes while simultaneously enhancing the local connectivity of minority class nodes to remediate structural imbalance issues. Both mechanisms synergistically operate through shared feature-representative nodes, enabling efficient information transfer and knowledge integration across GIER framework.

**Feature-representative Nodes** are defined as those selected to represent class centroids in long-term memory mechanism and connected to nodes chosen for confidence calibration in short-term mechanism. These nodes are pivotal to preserving memory integrity and calibration efficacy, ensuring both mechanisms synergistically enhance GNN performance on imbalanced graphs. At epoch  $t$ , first predict the class labels for each node, obtaining  $\hat{y}_i^{(t)}$ . Based on these predictions, we partition the nodes into sets corresponding to each class, denoted as  $\mathcal{V}_c^{(t)} = \{v_i \in \mathcal{V} : \hat{y}_i^{(t)} = c\}$ , and

compute the class centroid for class  $c$  as:

$$\mathbf{RF}_c^{(t)} = \frac{1}{|\mathcal{V}_c^{(t)}|} \sum_{v_i \in \mathcal{V}_c^{(t)}} \mathbf{x}_i^{(t)} \quad (5)$$

where  $\mathcal{V}_c^{(t)}$  denotes nodes predicted as class  $c$  at epoch  $t$ , and  $\mathbf{x}_i^{(t)}$  represents the input feature of node  $v_i$ .

### Long-term Subgraph Memory Mechanism

Long-term Subgraph Memory Mechanism is a multiple-round memory mechanism specifically designed to address the issue of data imbalance. Subsequently, we will present a thorough review of LSM.

On imbalanced graphs, significant sample quantity disparities between majority and minority classes cause GNNs to suffer from insufficient minority sampling, leading to overfitting and higher-frequency forgetting of minority classes. Our experiments reveal that minority class training exhibits the forgetting phenomenon across multiple epochs, making sampling from previously trained minority representations feasible. However, individual forgotten nodes are difficult to locate and exhibit significant bias.

Therefore, LSM is proposed to select the classes in the original labeled nodes whose quantities are below the average class quantity as the original minority classes and store feature-representative nodes from each epoch as current minority class distribution representatives. Utilizing a sliding window mechanism to maintain nodes from multiple previous epochs, LSM ensures representative coverage of minority class distributions over extended periods, constructing a stable memory subgraph. This approach prevents excessive

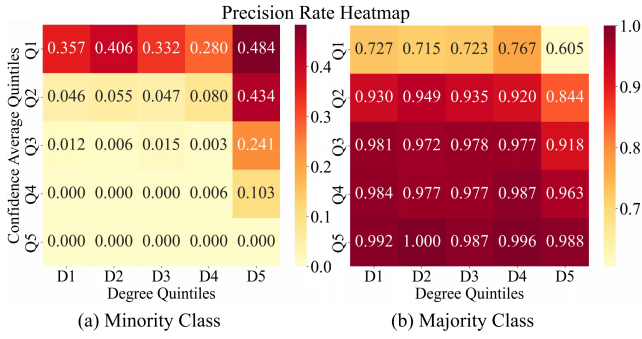


Figure 4: Degree & Confidence Influence on Precision

bias and high variance while maintaining balanced sampling across different classes.

**Sliding Window Memory.** To maintain computational efficiency while incorporating historical information, we maintain a sliding window containing the most recent  $T$  feature-representative nodes, where  $T$  is designed based on the sample difference between minority and majority classes:

$$\mathcal{H}_c^{(t)} = \{\mathbf{RF}_c^{(t-T+1)}, \mathbf{RF}_c^{(t-T+2)}, \dots, \mathbf{RF}_c^{(t)}\} \quad (6)$$

where  $\mathbf{RF}_c^{(t)} \in \mathbb{R}^d$  is the feature-representative nodes of class  $c$  at epoch  $t$ , and  $d$  is the feature dimension. The sliding window ensures that we only retain the most recent information, preventing memory overflow while still capturing the dynamic changes in minority class distributions.

**Historical Centroid Graph Construction.** Building upon the homophily principle in graphs, we construct connections between similar historical feature-representative nodes. The similarity matrix  $S_c^{(t)} \in \mathbb{R}^{T \times T}$  is defined as:

$$S_{ij}^{(c,t)} = \cos(\mathbf{RF}_c^{(t-T+i)}, \mathbf{RF}_c^{(t-T+j)}) \quad (7)$$

We perform a random selection based on a Bernoulli distribution, retaining the edges with higher probabilities from  $S_{ij}^{(c,t)}$  to form the edge set of the subgraph, denoted as  $\mathcal{E}_c^{(t)}$ .

This creates a temporal subgraph  $\mathcal{G}_c^{(t)} = (\mathcal{H}_c^{(t)}, \mathcal{E}_c^{(t)})$  for each minority class  $c$ , where current feature-representative nodes connect to all historical ones, while historical nodes connect based on similarity.

### Short-term Confidence Calibration Mechanism

SCC mechanism is a global dynamic confidence calibration framework specifically designed to address topological imbalance problems. Subsequently, we will present a thorough review of SCC.

As for minority classes, high-confidence, low-forgetting tendency and low recall rate may lead to significant bias in dynamic learning processes. Therefore, SCC is proposed to improve the predictions of GNNs on imbalanced graphs which dynamically adjusts graph structural attributes. It connects feature-representative nodes of each class to nodes with high confidence to mitigate the high-confidence in the classification results of misclassified minority class nodes in GNNs which can lead to the continued propagation of misinformation. This mechanism mitigates the tendency of GNNs

to propagate incorrect information with high confidence for minority class nodes, while simultaneously enhance the local connectivity of minority class nodes to address structural imbalance. The mechanism achieves efficient information transmission and knowledge integration through shared feature representative nodes in a dual-mechanism architecture. During the training phase, nodes are first divided into majority and minority sets based on the prediction results of GNNs. High-confidence nodes in both sets are then corrected separately.

**Dynamic Class Identification.** Unlike static approaches that rely on fixed label distributions, we dynamically identify minority classes based on current model predictions:

$$\rho_c^{(t)} = \frac{|\{v_i \in \mathcal{V} : \hat{y}_i^{(t)} = c\}|}{|\mathcal{V}|} \quad \mathcal{C}_{\min}^{(t)} = \{c : \rho_c^{(t)} < \bar{\rho}^{(t)}\} \quad (8)$$

**Degree-aware Node Selection.** To address structural bias in minority class selection, we employ degree-weighted probability for minority classes:

$$\hat{\mathbf{P}}_{\text{adj}}^{(t)} = \begin{cases} \text{softmax}_{\text{col}}(\hat{\mathbf{P}}^{(t)}) & \text{for } c \in \mathcal{C}_{\text{maj}}^{(t)} \\ \text{softmax}_{\text{col}}(\hat{\mathbf{P}}^{(t)} \odot \mathbf{D}) & \text{for } c \in \mathcal{C}_{\text{min}}^{(t)} \end{cases} \quad (9)$$

Where  $\mathbf{D} = \text{diag}(\deg(v_1), \dots, \deg(v_n))$  represents the degree of the nodes, and enhances selection probability for well-connected minority nodes.

**Experiment Justification.** The degree of a node can significantly affect its topological position in a graph. Figure 4 presents a probability heatmap of correctly predicted minority and majority class nodes under different confidence levels and degrees. We found that among nodes predicted as minority class, a higher degree corresponds to a higher correct prediction probability, indicating a higher true positive (TP) rate. For majority class nodes, the degree often tends to reduce prediction accuracy across various confidence levels. Therefore, when selecting TP nodes for minority class enhancement, we introduce degree correction, which helps increase the probability of selecting TP nodes while affording higher selection opportunities for low-confidence, high-degree minority class nodes.

**Confidence-based Filtering.** According to the high confidence bias, we choose high-confidence predictions to carry out deviation correction:

$$w_i^{(t)} = (\text{confidence}_i^{(t)} - \text{confidence}^{(t)})_+ \quad (10)$$

where  $w_i^{(t)}$  retains nodes with confidence levels above the average. We perform Bernoulli mask on both  $\hat{\mathbf{P}}_{\text{adj}}^{(t)}$  and  $\mathbf{w}^{(t)}$  to prevent overfitting.

$$\tilde{\mathbf{P}}_{\text{final}}^{(t)} = \hat{\mathbf{P}}_{\text{adj}}^{(t)} \odot \mathbf{w}^{(t)} \quad (11)$$

$\tilde{\mathbf{P}}_{\text{final}}^{(t)}$  indicates whether a connection is formed between the node and each feature-representative node. Selected high-confidence nodes form connections with corresponding class centroids, creating local topology enhancements that facilitate minority class learning. To ensure efficiency, we ultimately retain only the top  $\alpha$  percent of

Method		Cora			CiteSeer			PubMed		
		BAcc.	F1	Acc.	BAcc.	F1	Acc.	BAcc.	F1	Acc.
GCN	Vanilla	58.77	56.57	64.52	40.86	33.17	38.74	65.61	55.14	59.82
	reweight	63.44	62.52	68.62	46.92	41.34	45.54	71.81	68.15	69.28
	renode	64.19	63.13	68.90	47.56	43.54	46.80	71.55	68.84	69.44
	smote	59.12	57.37	65.50	38.71	30.83	37.42	70.63	64.77	67.40
	graphsmote	67.48	67.55	71.84	46.18	42.03	45.26	74.13	72.47	73.48
	graphens	69.91	69.20	71.90	58.04	56.77	60.60	73.61	72.54	73.52
	BAT(g-ens)	73.18	71.58	73.60	63.81	63.46	66.82	75.38	74.31	<b>76.06</b>
	GIER (LS)	72.87	72.69	<b>76.56</b>	65.63	65.31	67.90	<b>75.46</b>	<b>74.54</b>	74.98
GIER (Full)	<b>74.44</b>	<b>73.54</b>	75.62	<b>66.97</b>	<b>66.78</b>	<b>70.10</b>	74.10	74.03	74.68	
GAT	Vanilla	58.29	55.80	64.52	42.96	37.12	41.24	65.61	57.90	61.54
	reweight	64.02	63.46	68.00	45.24	39.10	46.08	70.09	66.45	68.58
	renode	62.79	62.28	67.90	46.39	41.48	45.42	71.43	67.71	69.78
	smote	56.83	53.45	64.24	37.75	28.51	35.40	30.78	58.18	61.82
	graphsmote	62.84	60.66	68.48	44.12	38.36	43.10	71.06	67.29	68.98
	graphens	69.02	68.75	71.94	53.02	51.18	54.26	72.64	71.32	72.22
	BAT(g-ens)	73.95	72.51	73.92	64.33	63.69	66.56	73.75	73.27	74.38
	GIER (LS)	70.41	70.95	75.22	65.53	65.27	68.36	75.06	72.94	74.01
GIER (Full)	<b>74.43</b>	<b>73.30</b>	<b>75.28</b>	<b>66.98</b>	<b>66.56</b>	<b>69.72</b>	<b>76.43</b>	<b>75.07</b>	<b>76.14</b>	
SAGE	Vanilla	57.64	55.21	64.18	43.93	38.69	43.80	65.23	56.53	60.8
	reweight	62.96	61.89	67.34	45.48	46.77	50.10	69.46	65.98	67.52
	renode	60.75	59.81	66.32	48.01	42.90	46.96	71.21	69.12	70.32
	smote	53.98	50.29	62.44	39.59	30.79	37.46	67.58	62.66	64.50
	graphsmote	57.33	55.02	64.32	40.96	34.87	40.62	72.06	67.78	69.12
	graphens	65.04	64.74	69.02	54.60	52.97	55.36	71.7	71.18	71.80
	BAT(g-ens)	72.09	70.78	73.20	64.95	64.40	68.52	76.13	74.57	75.42
	GIER (LS)	71.07	70.53	74.08	68.23	68.07	71.62	<b>77.24</b>	<b>76.08</b>	<b>75.98</b>
GIER (Full)	<b>73.51</b>	<b>71.73</b>	<b>73.96</b>	<b>68.39</b>	<b>68.03</b>	<b>71.10</b>	76.38	75.78	75.84	

Table 1: Performance comparison of various methods on different datasets. Comparisons of GIER with other baselines when imbalance ratio ( $\rho$ ) is 10.

<sup>LS</sup> Where  $\alpha$  represents selecting only the top  $\alpha = 10$  percent of nodes for confidence calibration during the SCC stage

Method		Cora			CiteSeer			PubMed		
		BAcc.	F1	Acc.	BAcc.	F1	Acc.	BAcc.	F1	Acc.
GCN	-LSM	63.16	61.62	68.08	39.90	34.71	37.60	62.76	56.03	60.14
	-SCC	73.12	73.05	76.52	57.15	52.52	57.16	75.39	74.39	74.96
GAT	-LSM	64.67	64.39	69.88	43.36	38.72	41.74	70.66	66.00	67.96
	-SCC	72.23	72.62	76.32	58.99	55.64	60.80	75.07	72.77	73.76
SAGE	-LSM	65.98	65.98	71.10	57.17	55.86	58.58	76.37	73.87	74.02
	-SCC	68.43	67.77	71.98	66.02	65.59	69.26	74.74	74.01	74.60

Table 2: Ablation studies of two mechanisms on GIER when imbalance ratio ( $\rho$ ) is 10.

nodes based on their confidence when constructing the final  $\tilde{\mathbf{P}}_{final}^{(t)}$ . When processing large-scale graphs ( $n \geq 10,000$ ), we first filter nodes with confidence levels in the top  $\alpha$  percent before SCC mechanism to minimize computational expenses.

## Experiments

### Experimental Setup

**Dataset** We validate our proposed GIER on three classic benchmark datasets: the citation networks Cora, CiteSeer, and PubMed (Sen et al. 2008). Following the same settings

as previous studies (Park, Song, and Yang 2021), we select half of the classes as minority classes. The imbalance ratio  $\rho = n_{max}/n_{min}$  refers to the proportion between the number of samples in the majority classes and those in the minority classes. To remain consistent with semi-supervised learning protocols, we set the number of samples in the majority classes to 20.

**Baselines** We evaluate our method on three architectures: GCN (Kipf and Welling 2017), GAT (Velickovic et al. 2017), and GraphSAGE (Hamilton, Ying, and Leskovec 2017). Our experiments include the following seven class imbalance adaptation techniques:

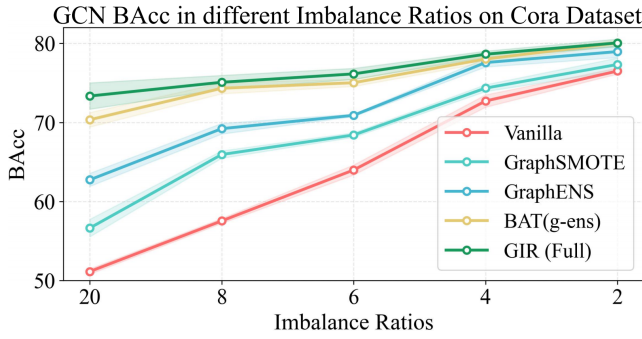


Figure 5: Different Imbalance Ratios in GIER.

“vanilla”, “reweight” (Japkowicz and Stephen 2002), “renode” (Chen et al. 2021), “smote” (Chawla et al. 2002), “graphsmote” (Zhao, Zhang, and Wang 2021), “graphens” (Park, Song, and Yang 2021), and “BAT” (Liu et al. 2024). Among them, “reweight”, “resample”, and “smote” are the most fundamental methods used for all imbalanced datasets. “Graphsmote” extends smote to graph structures. “Renode” improves weights by considering the structural differences between majority and minority class node positions. “Graphens” addresses the problem of minority node neighbor overfitting. “BAT” is a lightweight CR framework from the perspective of graph topology paradigms. We use three evaluation metrics to assess model performance: Balanced Accuracy, Macro-F1, and Accuracy for classification performance. Balanced Accuracy and Macro-F1 are evaluation metrics specifically designed for handling imbalanced datasets. Our experiments are performed on the RTX 3090(24GB), and the average value is taken 5 times. The value of  $T$  is  $(n_{max} - n_{min})2/3$ .

## Experiment Results

**Class-Imbalanced Node Classification.** Our experiments on graphs with an imbalance ratio of  $\rho = 10$  demonstrate that our method outperforms other baselines in terms of imbalance handling performance in most scenarios, as shown in Table 1. Our method particularly achieves significant improvements when applied to the GAT. The LS method filters nodes before SCC, enhancing performance on large graphs.

**Ablation Study.** To validate the effectiveness of each mechanism in GIER, we conduct an ablation study. The results are shown in Table 2. When we drop LSM module, the performance of GIER drops significantly, indicating that the LSM module is crucial for maintaining the stability of minority class representations. Similarly, when we remove SCC, the performance of GIER also decreases. SCC enables GIER to adaptively adjust the graph structure for better correction of deviations, thereby improving classification performance.

**Different Imbalance Ratios.** To further validate the effectiveness of GIER, we conduct experiments on graphs with varying imbalance ratios. We test GIER on imbalance ratios  $\rho$  of 2,4,6, 8, and 20, and compare its performance with other methods. As shown in Figure 5, the results indicate that GIER consistently outperforms other methods across differ-

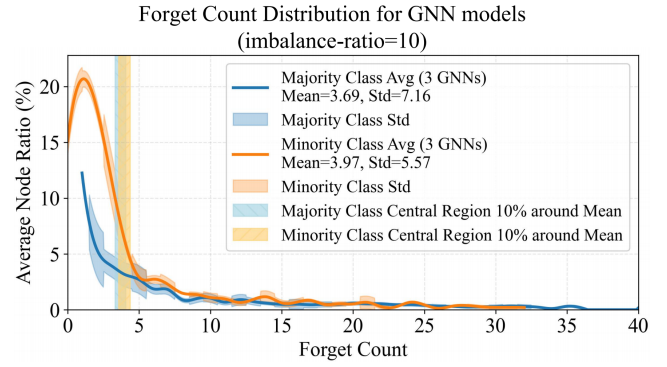


Figure 6: Forgetting Phenomenon in GIER.

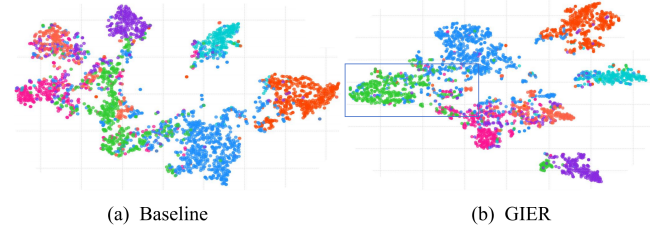


Figure 7: Visualization of Class Centroids

ent imbalance ratios, demonstrating its robustness and effectiveness in handling imbalanced graphs.

**Forgetting Phenomenon.** GIER primarily addresses class imbalance by improving the forgetting phenomenon of minority classes in imbalanced graphs using GNNs. We analyze the density distribution of “sample forgetting event” counts for majority and minority classes on the Cora dataset. As shown in Figure 6, the forgetting events of minority classes are concentrated in lower-frequency spaces, while their average forgetting counts are closer to those of the majority class. GIER can effectively alleviate minority class forgetting and simultaneously improve the performance of GNNs on imbalanced graphs.

**Visualization of Class Centroids.** To visualize node representations learned by GIER, we employ t-SNE(Maaten and Hinton 2008) for dimensionality reduction of centroid embeddings. As presented in Figure 7, minority classes in GIER form more compact and distinct clusters, indicating clearer class boundaries and preserving intra-class node similarity.

## Conclusion and Future Work

We introduce Graph Imbalance Experience Replay (GIER), a novel memory-augmented framework that addresses class imbalance in GNNs through forgetting dynamics. By combining long-term centroid-based memory with short-term topology enhancement, GIER effectively addresses minority-class forgetting bias and enhances the performance of GNNs on imbalanced graphs without compromising computational efficiency. In future work, we will explore broader heterogeneous graphs and investigate extensions to lifelong learning scenarios and their interrelations.

## Acknowledgments

We appreciate constructive feedback from anonymous reviewers and meta-reviewers. This work is being supported by the National Natural Science Foundation of China under the Grant No. 62172451 and Key Program of the National Natural Science Foundation of China under the Grant No. 62536009.

## References

- Burkhardt, J. G.; Wu, G.; Song, X.; Raimondi, F.; McWeeney, S.; Wong, M. H.; and Deng, Y. 2023. Biology-inspired graph neural network encodes reactome and reveals biochemical reactions of disease. *Patterns*, 4(7).
- Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; and Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16: 321–357.
- Chen, D.; Lin, Y.; Zhao, G.; Ren, X.; Li, P.; Zhou, J.; and Sun, X. 2021. Topology-imbalance learning for semi-supervised node classification. *Advances in Neural Information Processing Systems*, 34: 29885–29897.
- Cheng, Z.; Li, Z.; Li, Y.; Song, Y.; Zhao, K.; Cheng, D.; Li, J.; and Yu, J. X. 2025. Can LLMs alleviate catastrophic forgetting in graph continual learning? A systematic study. *arXiv preprint arXiv:2505.18697*.
- Fang, R.; Wen, L.; Kang, Z.; and Liu, J. 2022. Structure-preserving graph representation learning. In *2022 IEEE International Conference on Data Mining (ICDM)*, 927–932. IEEE.
- Ga, S.; Cho, P. H.; Moon, G. E.; and Jung, S. 2025. Efficient GNN-based social recommender systems through social graph refinement. *The Journal of Supercomputing*, 81(1): 215.
- Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.
- Hassani, H.; Nikan, S.; and Shami, A. 2025. Improved exploration–exploitation trade-off through adaptive prioritized experience replay. *Neurocomputing*, 614: 128836.
- Japkowicz, N.; and Stephen, S. 2002. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5): 429–449.
- Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.
- Lachaud, G.; Conde-Cespedes, P.; and Trocan, M. 2022. Graph neural networks-based multilabel classification of citation network. In *Asian Conference on Intelligent Information and Database Systems*, 128–140. Springer.
- Liu, Y.; Zheng, X.; Li, Y.; and Guo, Y. 2025. Test-Time Adaptation on Recommender System with Data-Centric Graph Transformation. *International Joint Conference on Artificial Intelligence (IJCAI)*.
- Liu, Z.; Qiu, R.; Zeng, Z.; Yoo, H.; Zhou, D.; Xu, Z.; Zhu, Y.; Weldemariam, K.; He, J.; and Tong, H. 2024. Class-imbalanced graph learning without class rebalancing. In *Proceedings of the 41st International Conference on Machine Learning, ICML’24*. JMLR.org.
- Ma, Y.; Tian, Y.; Moniz, N.; and Chawla, N. V. 2025. Class-imbalanced learning on graphs: A survey. *ACM Computing Surveys*, 57(8): 1–16.
- Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov): 2579–2605.
- Pang, J.; Lin, C.; Hao, X.; Yin, R.; Wang, Z.; Zhang, Z.; He, J.; and Tai Sheng, H. 2024. FTF-ER: Feature-Topology Fusion-Based Experience Replay Method for Continual Graph Learning. In *Proceedings of the 32nd ACM International Conference on Multimedia, MM’24*, 8336–8344. ACM.
- Park, J.; Song, J.; and Yang, E. 2021. Graphens: Neighbor-aware ego network synthesis for class-imbalanced node classification. In *International conference on learning representations*.
- Pham, P.; Bui, Q.-T.; Nguyen, N. T.; Kozma, R.; Yu, P. S.; and Vo, B. 2025. Topological data analysis in graph neural networks: Surveys and perspectives. *IEEE Transactions on Neural Networks and Learning Systems*.
- Pu, R.; Xu, G.; Fang, R.; Bao, B.-K.; Ling, C.; and Wang, B. 2025. Leveraging group classification with descending soft labeling for deep imbalanced regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 19978–19985.
- Rohlf, C. 2025. Generalization in neural networks: A broad survey. *Neurocomputing*, 611: 128701.
- Stern, U.; Yaacoby, T.; and Weinshall, D. 2025. On local overfitting and forgetting in deep neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 20592–20600.
- Toneva, M.; Sordani, A.; des Combes, R. T.; Trischler, A.; Bengio, Y.; and Gordon, G. J. 2019. An Empirical Study of Example Forgetting during Deep Neural Network Learning. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.
- Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y.; et al. 2017. Graph attention networks. *stat*, 1050(20): 10–48550.
- Wei, J.; Zhang, Y.; Zhang, L. Y.; Ding, M.; Chen, C.; Ong, K.-L.; Zhang, J.; and Xiang, Y. 2024. Memorization in deep learning: A survey. *arXiv preprint arXiv:2406.03880*.
- Yang, L.; Chen, M.; Chen, T.; Hu, J.; and Wang, Z. 2025. GraphDEH: Graph diffusion enhanced hypergraph method for class-imbalanced node classification. *2025 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.
- Zhang, C.; Wang, W.; Tian, Z.; and Yu, S. 2024. Forgetting and remembering are both you need: Balanced graph structure unlearning. *IEEE Transactions on Information Forensics and Security*, 19: 6751–6763.

- Zhang, X.; Song, D.; and Tao, D. 2022. Sparsified subgraph memory for continual graph representation learning. In *2022 IEEE International Conference on Data Mining (ICDM)*, 1335–1340. IEEE.
- Zhao, C.; Qian, Y.; Wang, B.; Gu, Z.; Ji, S.; Wang, W.; and Zhang, Y. 2025. Adversarial training via multi-guidance and historical memory enhancement. *Neurocomputing*, 619: 129124.
- Zhao, T.; Zhang, X.; and Wang, S. 2021. Graphsmote: Imbalanced node classification on graphs with graph neural networks. In *Proceedings of the 14th ACM international conference on web search and data mining*, 833–841.
- Zheng, X.; Huang, W.; Zhou, C.; Li, M.; and Pan, S. 2025. Test-time graph neural dataset search with generative projection. In *Forty-second International Conference on Machine Learning*.
- Zheng, X.; Zhang, M.; Chen, C.; Nguyen, Q. V. H.; Zhu, X.; and Pan, S. 2024. Structure-free graph condensation: From large-scale graphs to condensed graph-free data. *Advances in Neural Information Processing Systems*, 36.
- Zhou, F.; and Cao, C. 2021. Overcoming catastrophic forgetting in graph neural networks with experience replay. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 4714–4722.
- Zhou, M.; and Gong, Z. 2023. GraphSR: A data augmentation algorithm for imbalanced node classification. In *Proceedings of the AAAI Conference on artificial intelligence*, volume 37, 4954–4962.