SeSA: Synergizing Semantic and Structural Awareness in LLMs for Imbalanced Node Classification on Text-Attributed Graphs

Anonymous ACL submission

Abstract

Text-Attributed Graphs (TAGs) have emerged as a prominent paradigm for modeling complex relational data, such as academic citation networks and social media platforms. Concurrently, Large Language Models (LLMs) have achieved remarkable success in graph-centric tasks by serving as powerful tools for text attribute processing. Nevertheless, the ubiquitous real-world class imbalance problem severely hinders the robustness and generalization capability of existing models. Although various LLM-based approaches have been proposed, they generally overlook class imbalance within TAGs. To address this gap, this paper comprehensively reveals that such imbalance extends beyond mere disparities in sample size, presenting a multi-dimensional challenge characterized by label distributions, structural biases, and textual attribute heterogeneity. Driven by this holistic understanding, we propose SeSA: A Semantic-Structure-Aware Framework for Imbalanced Node Classification on Text-Attributed Graphs. Specifically, SeSA leverages LLMs to generate highly discriminative node embeddings and pseudo-labels, thereby mitigating structural bias inherently present in the message-passing process. Extensive experiments on five benchmark class-imbalanced datasets demonstrate that SeSA significantly outperforms mainstream baselines across multiple evaluation metrics.

1 Introduction

In real-world graph-structured data, nodes are frequently characterized by rich textual attributes, defining the prevalent structure of Text-Attributed Graphs (TAGs) (Yan et al., 2023). Spanning from citation networks to social media, TAGs serve as a fundamental architecture for modeling complex relational data. While the integration of Large Language Models (LLMs) (He et al., 2024; Chien et al., 2021; Jin et al., 2024; Li et al., 2023b) and Graph Neural Networks (GNNs) (Chen et al., 2018; Kipf and Welling, 2016; Veličković et al., 2017, 2018; Hamilton et al., 2017) has demonstrated advanced capabilities in node classification on TAGs, leveraging the synergy between high-dimensional semantics and topological fusion, most existing methodologies rely on the idealized assumption of uniform label distributions. Consequently, they overlook the pervasive class imbalance problem inherent in practical scenarios (Liu et al., 2025b). This skewness transcends simple numerical disparity; it is intrinsically intertwined with the graph topology and high-dimensional textual space, posing significant challenges to the robustness of representation learning. We illustrate these multifaceted imbalance phenomena in Figure 1.

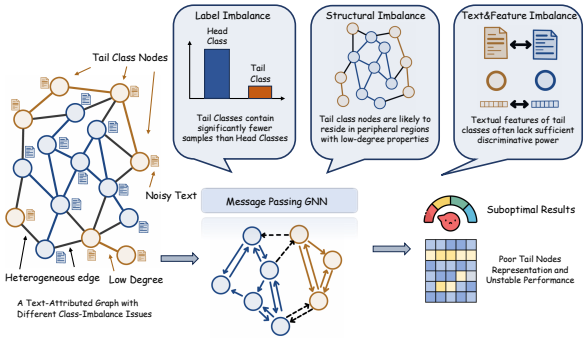

Figure 1: Intertwined Imbalance Phenomena Inherent in Long-Tail Text-Attributed Graphs

In TAGs, class imbalance is not merely a one-dimensional issue of numerical scarcity, but rather a complex confluence of several distinct phenomena (Liu et al., 2025b). Firstly, there is the label long-tail distribution, whereby the sample sizes of minority classes are significantly dwarfed by those of majority classes. This imbalance frequently results in a prediction bias biased toward the majority classes during model training. Secondly, and more critically, is structural imbalance, a factor that is often overlooked. Empirical evidence suggests that minority class nodes are typically situated at the

graph's periphery or exhibit low-degree characteristics; their local neighborhoods are often sparse and dominated by heterophilous neighbors. Within the Message Passing Neural Networks (MPNNs), this structural disadvantage severely hinders minority nodes from aggregating discriminative information through feature propagation. Lastly, textual-feature imbalance is equally non-negligible. Substantial variations exist across classes in terms of semantic length, lexical composition, and the distribution of domain-specific terminology (Wang et al., 2025a). Consequently, the textual features associated with minority classes often lack sufficient discriminative power, further exacerbating the difficulty of classifier decision-making. Despite their promise, existing LLM-based approaches for TAGs still yield sub-optimal performance when confronted with severe and intricate class imbalance. To rectify these deficiencies, it is imperative for LLM-driven frameworks to capture the complex, reciprocal interplay between node textual features and topological structures. This necessitates leveraging the profound semantic comprehension of LLMs to actively guide graph structure learning, while simultaneously unlocking their latent potential in modeling graph topology.

To address the aforementioned challenges and gaps in prior work, we propose a unified framework for node classification in class-imbalanced Text-Attributed Graphs, aiming to systematically tackle the multiple, intertwined imbalance factors inherent in TAGs. The main contributions of this work are summarized as follows:

- Collaborative Modeling of Multi-dimensional Class Imbalance: This work systematically investigates and mitigates the coupled effects of class-level imbalance, structural imbalance, and textual attribute heterogeneity in TAGs.

- Semantic–Structure Aware Prompt Engineering: We design a class-imbalance-aware prompting strategy for language models, leveraging LLMs to enhance the semantic discriminability of minority-class nodes.

- Structural Imbalance Enhancement Module: We propose a structure enhancement module that explicitly accounts for the interplay between graph structure and node features, effectively alleviates structural bias during the MPNN process.

- Rigorous Experimental Validation: Extensive experiments are conducted on five widely used class-imbalanced TAG datasets to highlight the proposed model's effectiveness in handling complex imbalance scenarios.

2 Methodology

2.1 Preliminary

Class-imbalance. Let $\mathcal{G} = (\mathcal{V}, \mathcal{A}, \mathcal{T})$ denote a graph with N nodes, where $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ is the set of nodes, $\mathcal{A} \in \{0,1\}^{N\times N}$ is the adjacency matrix, and $T_i \in \mathcal{T}$ is the textual-feature of node v_i. Considering node classification, each node in the graph corresponds to a class $C_i \in \{1, \cdots, C\}$ where C is the number of classes and $|C_i|$ denote the number of the nodes in class C_i. In the class-imbalanced scenario, the labeled nodes in training set $\mathcal{V}^L \subset \mathcal{V}$ are imbalanced, the imbalance ratio can be expressed as

$$\mathcal{IR} = \frac{\max_i(|C_i|)}{\min_j(|C_j|)}. \tag{1}$$

Foundational Characterization of Multi-dimensional Imbalances on TAGs. To quantify the intrinsic impact of graph structure on imbalanced learning, we propose the *Comprehensive Imbalance Ratio*. This metric evaluates the differential representation-learning capacity of various classes, providing a robust measure for the structural disparities inherent.

We first define Local Structural Purity, given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, the topological influence of a node $u \in \mathcal{V}$ is recalibrated by a structural saliency function $S(u)$:

$$S(u) = \ln(1 + d_u), \tag{2}$$

where d_u denotes the degree of node u. Then for any node $v \in C_i$, its local structural purity $\rho(v)$ is defined as the weighted ratio of effective signals to topological noise within its neighborhood:

$$\rho(v) = \frac{\sum_{u\in N(v)} f_i(u) \cdot S(u)}{\sum_{u\in N(v)} S(u) + \epsilon}, \tag{3}$$

where ϵ is a small constant for numerical stability, and $f_i(u)$ is a class indicator function. The value $\rho(v) \in [-1, 1]$ characterizes the signal purity of a node within its local neighborhood.

The global topological potential of class C_i is defined as the expected Structural Information Purity (SIP) of its constituent nodes, denoted as

$$\Phi_i = \mathbb{E}_{v\in C_i}[\rho(v)]. \tag{4}$$

Finally, the Comprehensive Imbalance Ratio (CIR) for the entire graph is formulated as a non-linear mapping of the potential range:

$$CIR = 1.0 + \eta \cdot \left| \max_i \Phi_i \cdot \Sigma_i - \min_j \Phi_j \cdot \Sigma_j \right|^{\theta}, \tag{5}$$

where Σ_i is the intra-class textual semantic homogeneity of class C_i, η is a contrast normalization constant, and θ is a power factor reflecting the topological sensitivity across different datasets.

2.2 Semantic-Structure-Aware Framework

In this section, we present a novel framework tailored for node classification on class-imbalanced TAGs, designed to comprehensively address the intertwined imbalance phenomena inherent in such data. As illustrated in Figure 2, the proposed architecture primarily comprises two stages: an **Dual-stage LM Feature Extraction Module**, and a **Semantic-aware Graph Structure Augmentation Module**. In the following subsections, we elaborate on how the feature extraction module generates refined embeddings and pseudo-labels by capturing information through two decoupled stages. Furthermore, we discuss how the structure augmentation module leverages these semantically rich upstream embeddings to optimize the graph topology for subsequent GNN training.

2.3 Dual-stage LM Feature Extraction Module

Within the LLM-as-Enhancer paradigm for TAGs, synergizing semantics with topology is critical. However, existing class-imbalance methods suffer from suboptimal structural handling and architectural over-coupling. Current approaches typically employ homogeneous transformations that neglect the inherent sparsity and heterophily prevalent in minority nodes. Moreover, the rigid coupling between semantic extraction and structural modeling inhibits adaptive adjustments across diverse imbalance sources. To address these limitations, we propose a decoupled, dual-stage LM feature extraction module. This approach facilitates phase-specific adaptations, thereby generating high-fidelity representations optimized for imbalanced classification.

Structural-Semantic Distillation. In the initial stage, raw textual features $T_i \in \mathcal{T}$ are compressed into a compact representation sensitized to the underlying graph topology. A pre-trained LLM functions as a structural-semantic distiller to produce the optimal messages that a node generates for subsequent neighborhood aggregation in the classification task. During the distillation process, the LLM integrates pivotal structural and distributional metrics, specifically node degree, degree centrality, and local homophily (the proportion of intra-class neighbors). While degree and centrality capture fundamental structural properties, the local homophily reflects the attributes that manifest distinct patterns across majority and minority classes.

For each node $v_i \in \mathcal{V}$, we first extract a structural attribute vector $s_i = [d_i, c_i, \psi_i]$, where d_i denotes the degree, c_i represents the degree centrality, and ψ_i signifies the local homophily estimation. The LLM is employed as a structural-semantic distiller for the raw text T_i. Under the guidance of s_i, the LLM generates a simplified text $\tau_{i,\text{sim}}$, which can be viewed as the most discriminative information node v_i intends to propagate to its neighbors. Formally, this process is defined as:

$$\tau_{i,\text{sim}} = \text{Prompt}(T_i, s_i). \tag{6}$$

In Equation 6, $\tau_{i,\text{sim}}$ explicitly encodes the node's role within the imbalanced structure, such as its membership in a minority community or its potential as a bridge node, without requiring explicit neighborhood text in this stage. In addition to the distilled semantics, the LLM generates two pivotal inferential indicators: the predicted class label and the probability of the node's membership in a minority category. Comprehensive prompt templates and representative input-output exemplars are provided in the Appendix .

Dual-Context Augmented Encoding. In the second stage, a Small Language Model (SLM) is employed to generate high-fidelity node embeddings by synthesizing multifaceted semantic contexts. Consistent with established feature extraction paradigms, the proposed Semantic Learning Module integrates raw textual attributes with the previously distilled summaries to produce both node representations and class predictions. The predicted labels serve as auxiliary supervision for fine-tuning the SLM, while the optimized embeddings are subsequently utilized by downstream GNN. For a target node v_i, we construct an augmented textual input sequence $\hat{T}_i$ by concatenating its original text, its own semantic kernel, and the aggregated kernels from its immediate neighborhood:

$$\hat{T}_i = \text{Concat}\left(T_i, \text{Aggr}(\{\tau_{j,\text{sim}} \mid u_j \in \mathcal{N}(v_i)\})\right), \tag{7}$$

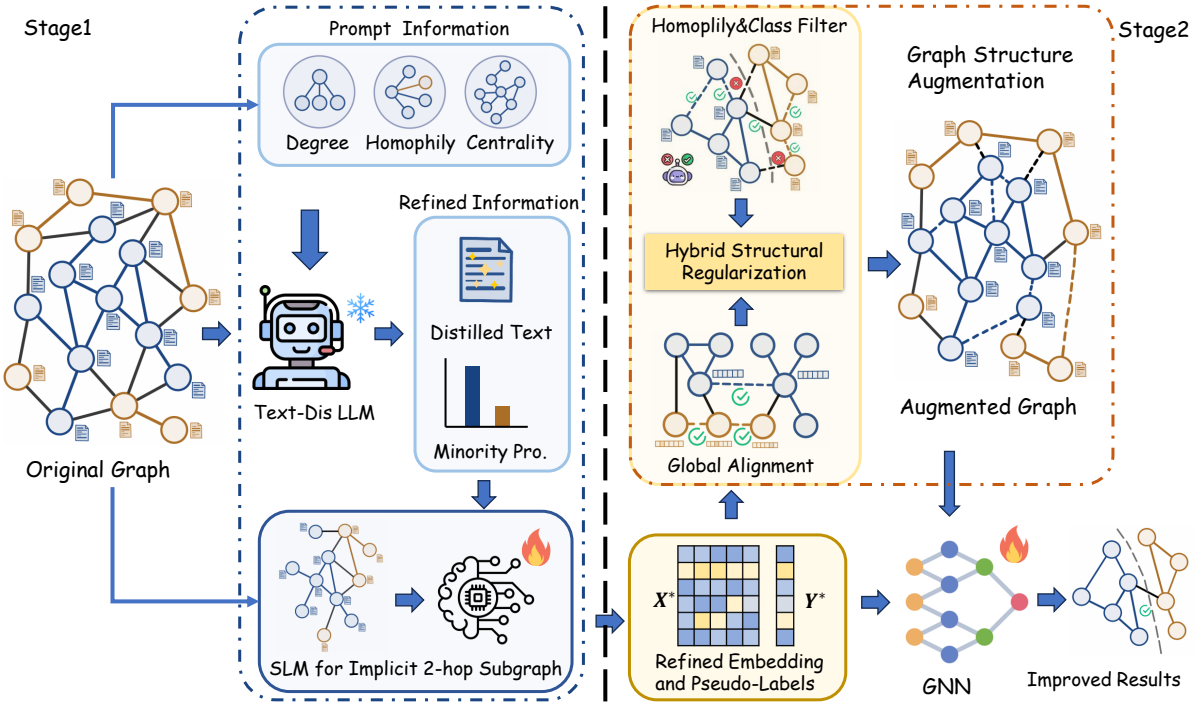

Figure 2: The Overall Framework of The Proposed SeSA.

where $\mathrm{Aggr}(\cdot)$ denotes a symmetric aggregation operator. Subsequently, the SLM maps the augmented sequence $\hat{T}_i$ into a low-dimensional latent space to produce the final node embedding z_i and a preliminary class prediction $\hat{y}_i$:

$$z_i, \hat{y}_i = \mathrm{SLM}_\theta(\hat{T}_i). \tag{8}$$

The resulting embedding z_i encapsulates both the local original semantics and the distilled context of its 1-hop neighbors. Using pre-distilled $\tau_{j,\mathrm{sim}}$, the SLM effectively perceives the categorical and homophily distribution within the local subgraph, providing discriminative signals for the classification of minority nodes.

2.4 Semantic-aware Graph Structure Augmentation Module

Despite the efficacy of the dual-stage LM feature extraction in capturing deep semantics and implicit structural features, message passing over raw topology remains constrained by inherent structural biases. In class-imbalanced TAGs, structural imbalance and semantic bias are intrinsically intertwined, characterized by the topological isolation of minority nodes and heterophily-induced representation collapse. Consequently, a Semantic-aware Graph Structure Augmentation module is proposed to dynamically refine the graph topology using embeddings from the preceding stage as informative priors.Through a Multi-view Topology Fusion Mechanism, dual complementary components and a hybrid regularization strategy are integrated to generate homophily-enhanced edges. This mechanism bolsters the sparse connectivity of minority classes, thereby comprehensively mitigating structural imbalance and enhancing the robustness of downstream representation learning.

Global Semantic Alignment. The raw adjacency matrix $\mathcal{A}$ often confines the receptive field of minority nodes to sparse, local neighborhoods. To circumvent the constraints of topological proximity, we construct a global semantic affinity view $\mathcal{A}^{(S)}$ in the latent space. Using LLM-enhanced embeddings, the semantic similarity S_{ij} between a node pair (v_i, v_j) is computed as:

$$S_{ij} = \frac{\exp(\mathrm{sim}(z_i, z_j)/\delta)}{\sum_{k\in\mathcal{V}} \exp(\mathrm{sim}(z_i, z_k)/\delta)}, \tag{9}$$

where $\mathrm{sim}(\cdot,\cdot)$ denotes cosine similarity and τ is a temperature hyperparameter. To maintain computational efficiency on large-scale graphs, we employ a block-wise retrieval strategy to extract the top-k nearest neighbors for each node:

$$\mathcal{A}^{(S)}_{ij} = \begin{cases} 1, & \text{if } v_j \in \text{Top-}k(v_i, \mathbf{S}) \\ 0, & \text{otherwise} \end{cases}. \tag{10}$$

This mechanism establishes semantic bridges to connect topologically isolated minority nodes with semantically related clusters. By capturing long-range semantic dependencies, this approach mitigates supervision isolation and enriches the informative signals available to minority classes.

Confidence-aware Homophily Enforcement. Feature-based similarity alone often induces spurious edges at semantic boundaries. To mitigate this, predictive confidence is utilized as high-level supervision to enforce a strict homophily constraint. Specifically, the class posterior $\hat{y}_i$ from the preceding stage defines the prediction confidence $\phi_i = \max(\hat{y}_i)$. Given a dynamic threshold γ, high-confidence nodes are identified as "pseudo-anchors" to guide the construction of the homophily-enhanced view $\mathcal{A}^{(H)}$:

$$\begin{aligned}\mathcal{A}_{ij}^{(H)} = \mathbb{I}(\arg\max \hat{y}_i = \arg\max \hat{y}_j) \\ \cdot \mathbb{I}(\phi_i > \gamma) \cdot \mathbb{I}(\phi_j > \gamma),\end{aligned} \tag{11}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. By necessitating both label consistency and high confidence, $\mathcal{A}^{(H)}$ ensures that augmented edges predominantly link nodes within the same semantic manifold. This mechanism effectively prevents majority-class noise from contaminating minority-class representations, thereby preserving the discriminative integrity of the feature space.

Hybrid Structural Regularization. To achieve a balance between manifold smoothness and inter-class separability, we preserve the intrinsic skeletal structure while selectively integrating complementary insights to form the augmented graph $\tilde{\mathcal{A}}$ via the hybrid refinement strategy:

$$\tilde{\mathcal{A}} = \text{Normalize}\left(\mathcal{A} + \lambda_1 \mathcal{A}^{(S)} \odot \lambda_2 \mathcal{A}^{(H)}\right), \tag{12}$$

where $\odot$ denotes the element-wise refinement operator, and λ_1, λ_2 serve as scaling factors to ensure that long-range connections conform to high-level homophily constraints. The resulting augmented topology $\tilde{\mathcal{A}}$ harmonizes local structural dependencies with LLM-derived global semantic consistency. This produces a refined computational graph with distinct class boundaries and enhanced homophily, optimized for downstream GNN classification in imbalanced scenarios.

3 Experiments

In this section, we conduct extensive experiments to evaluate the effectiveness of our proposed framework, underscoring its capability to mitigate multiple intertwined imbalance challenges simultaneously. Specifically, we investigate the performance advantages of SeSA by addressing the following three key research questions (**RQs**) :

RQ1: How does SeSA perform in comparison with state-of-the-art baselines from both graph imbalanced learning and LLM-driven graph paradigms?

RQ2: Do the Dual-stage LM Feature Extraction Module and the Semantic-aware Graph Structural Augmentation Module each contribute significant performance gains, and do they exhibit complementary effects within the overall framework?

RQ3: Can our proposed method consistently enhance the classification performance of minority classes across datasets exhibiting diverse manifestations and varying intensities of imbalance?

3.1 Experimental Setup

The proposed framework is evaluated on five benchmark datasets: three citation networks: Cora, CiteSeer, and PubMed (Sen et al., 2008) and two co-purchase graphs: Amazon-Computers and Amazon-Photo (Yan et al., 2023). Data partitioning follows the supervised node classification protocol established by Chen et al. (2018). To simulate real-world class skewness, a Long-Tail (LT) distributed training set is constructed following Cui et al. (2019), with the Imbalance Ratio (IR) as the primary control variable. Model performance is quantified using Accuracy (Acc), Balanced Accuracy (BAcc) and the Macro-F1 score (F1). All metrics are reported in the range $[0, 1]$, where higher values denote superior efficacy. Detailed empirical analyses and additional implementation specifics are provided in Appendix.

3.2 Model Performance (RQ1)

3.2.1 Comparison with Graph Imbalance Learning Baselines

As summarized in Table 1, the proposed framework is evaluated against 14 representative baselines across five benchmark datasets under a long-tailed setting (IR=100), consistently achieving state-of-the-art (SOTA) performance. On citation networks, our model exhibits substantial superiority; notably, it reaches 89.83% accuracy on Pubmed, outperforming the strongest baseline, NodeImport, by 5.36%, while surpassing leading competitors by approximately 3.3% in BAcc and 4.7% in F1-score. This performance advantage extends to co-

Table 1: Performance comparison against Specialized Graph Imbalance Baselines on five datasets. Best in **bold**; second best underlined.

Datasets	Cora			Citeseer			Pubmed			Photo			Computers		
IR = 100	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1
Vanilla	74.80±0.27	72.11±1.69	68.92±0.37	55.51±0.22	54.13±0.75	44.31±0.42	63.79±1.32	62.87±0.81	60.91±1.36	86.81±0.22	84.12±0.44	76.03±0.54	81.87±0.29	80.30±0.56	80.12±1.23
Re-weight	73.29±1.66	65.13±1.62	65.72±2.59	60.97±0.98	57.13±1.03	57.01±1.47	70.23±1.61	69.00±1.45	69.49±1.92	86.17±0.43	85.01±0.66	82.78±0.57	87.06±0.77	86.18±0.72	83.11±1.16
PC-Softmax	68.54±3.91	62.96±1.32	66.55±3.00	58.22±2.27	52.19±1.12	54.81±2.21	70.84±0.49	64.61±0.78	58.92±0.39	90.34±0.33	88.56±0.43	87.73±0.42	87.38±0.58	84.39±0.81	85.19±0.31
ReNode	79.02±0.17	73.53±0.25	74.71±0.22	63.40±0.50	55.79±1.48	55.10±0.47	75.64±0.32	72.56±2.27	72.22±0.35	86.42±0.15	84.74±1.07	83.86±0.21	78.85±0.38	76.02±1.09	73.16±0.66
TAM	78.34±0.11	76.11±0.32	73.34±0.34	64.22±0.50	55.62±1.55	58.13±0.65	77.84±0.31	71.96±3.54	73.08±0.53	85.08±0.36	77.47±1.24	82.69±0.89	77.95±0.50	73.39±0.51	72.31±0.93
BAT	76.23±1.60	72.00±1.34	72.01±1.72	61.47±1.42	59.21±1.13	57.22±0.87	75.52±1.48	72.58±2.60	73.16±2.39	87.19±1.28	86.07±1.66	85.97±1.43	82.87±0.72	81.95±0.54	80.56±0.67
Upsample	76.48±1.64	70.76±2.81	71.24±2.90	62.33±1.01	59.52±0.98	58.43±1.23	73.07±1.01	70.26±0.98	71.34±0.90	88.27±0.78	86.14±1.56	87.11±1.72	84.17±1.00	82.76±0.97	83.19±1.76
GraphSMOTE	73.72±0.33	74.18±0.51	66.74±0.37	57.81±0.54	53.54±0.40	52.01±0.49	73.70±0.31	68.11±1.88	65.32±0.94	86.37±0.14	85.37±0.25	84.11±0.29	82.81±1.12	82.09±1.66	81.81±2.12
GraphENS	75.90±0.21	74.73±0.18	68.06±1.01	58.30±0.29	56.44±0.66	53.48±0.51	81.32±2.01	76.30±3.74	73.82±1.16	88.22±0.31	86.77±0.89	81.19±1.09	88.31±0.31	85.81±0.28	83.51±0.88
GraphSHA	79.28±0.13	77.12±0.21	75.10±0.15	63.98±0.43	55.84±1.45	57.08±0.37	80.24±0.43	77.37±3.96	74.24±0.89	86.12±0.17	87.43±1.09	86.75±0.33	87.27±0.98	83.07±1.65	83.26±0.90
NodeImport	80.03±0.50	77.80±0.64	75.01±0.72	64.43±0.55	**62.21**±0.73	58.61±0.57	84.57±0.23	83.18±0.19	83.16±0.31	90.09±0.23	89.03±0.36	87.47±0.40	88.07±0.12	85.98±0.20	86.01±0.17
SeSA	**80.21**±0.26	**78.03**±0.54	**76.89**±0.42	**65.39**±0.27	60.01±0.23	**58.82**±0.17	**89.83**±0.27	**86.50**±0.25	**87.82**±0.17	**91.17**±0.21	**90.87**±0.20	**89.23**±0.35	**88.53**±0.28	**86.93**±0.77	**87.56**±0.26

purchase networks (Photo and Computers) despite severe structural skewness, where our model maintains high stability, achieving a 91.17% BAcc and an 89.23% Macro-F1 on Photo. Compared to specialized baselines like GraphSHA (Li et al., 2023a), BAT (Liu et al., 2024), and NodeImport (Chen et al., 2025), whose efficacy often diminishes on larger graphs, our framework delivers superior and robust results. These empirical findings underscore that leveraging LLM-derived semantic guidance to address class scarcity is fundamentally more effective than conventional loss calibration or raw feature space resampling, as higher-order semantic cues successfully compensate for and transcend the limitations of skewed graph topologies.

3.2.2 Comparison with LLM-Enhanced Graph Learning Baselines

To explicitly isolate the performance gains of our imbalance-aware design from the raw representational capacity of LLMs, we introduce a comparative study against classic LLM-enhanced graph learning frameworks. To ensure a fair comparison, all methods utilize the same lightweight SLM (DeBERTa) for final text encoding and are evaluated using GCN as the fixed downstream backbone. Although powerful LLM-derived semantics yield performance advantages under standard settings, their efficacy remains sub-optimal in highly imbalanced scenarios. As shown in Table 2, SeSA consistently outperforms pure LLM-enhanced methods, achieving the most significant improvements in BAcc across the evaluated imbalanced settings. This demonstrates that merely injecting stronger semantic features, without dedicated imbalance-handling mechanisms, is insufficient for robust classification in skewed distributions; sole LLM enhancement fails to rescue minority nodes suffering from severe local topological degradation. In contrast, our method directly targets this bottleneck. This asymmetric improvement firmly establishes that our structural compensation mechanism is indispensable and cannot be substituted by simply employing more powerful language models.

3.3 Ablation Study (RQ2)

As demonstrated in Table 3, omitting the dual-stage LM extraction module (w/o DLM) leads to an approximate 8% degradation in Cora accuracy and a 14% drop in the Pubmed F1-score. This confirms that relying solely on raw text or vanilla GNN aggregation fails to capture deeply discriminative features, underscoring the role of LLM-extracted semantics in mitigating distribution disparities. Furthermore, excluding the graph structure augmentation module (w/o GSA) causes the Pubmed Balanced Accuracy (BAcc) to decline sharply from 86.50% to 76.60%. This pivotal drop indicates that by reconstructing topological connectivity, the GSA module establishes vital communication pathways for marginalized minority nodes, thereby rectifying structural imbalance. Overall, fine-grained analysis reveals that each constituent of SeSA addresses distinct facets of multi-dimensional challenges, including semantic deficiency and structural isolation, while profound synergies among these modules collaboratively form a robust framework for complex TAG tasks.

3.4 Model Analysis (RQ3)

Hyperparameter Sensitivity Analysis. To evaluate the impact of topological augmentation inten-

Table 2: Performance comparison against LLMs-Enhanced GNN Baselines. Best in **bold**; second best underlined.

Datasets	Cora			Citeseer			Pubmed			Photo			Computers		
IR = 100	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1
Vanilla	74.80±0.27	72.11±1.69	68.92±0.37	55.51±0.22	54.13±0.75	44.31±0.42	63.79±1.32	62.87±0.81	60.91±1.36	86.81±0.22	84.12±0.44	76.03±0.54	81.87±0.29	80.30±0.56	80.12±1.23
GIANT	72.11±0.57	70.01±0.82	70.53±0.49	57.83±0.47	55.09±0.32	52.11±0.52	74.51±0.30	73.52±1.17	68.30±0.39	87.22±0.19	82.71±0.88	83.52±0.51	85.08±0.21	80.91±1.19	84.27±0.24
GLEM	73.29±1.66	65.13±1.62	65.72±2.59	60.97±0.98	57.13±1.03	57.01±1.47	70.23±1.61	69.00±1.45	69.49±1.92	86.17±0.43	83.01±0.66	82.78±0.57	86.06±0.77	82.18±0.72	83.11±1.16
LLM-GNN	77.01±0.43	75.13±1.39	72.41±0.67	60.56±0.60	56.22±0.43	52.74±0.67	76.41±0.33	71.12±1.09	72.67±0.34	87.19±0.21	82.31±0.38	84.66±0.41	85.61±0.38	82.43±1.17	81.56±0.40
TAPE	76.09±0.34	76.66±0.51	69.18±0.30	60.31±0.47	55.48±1.42	52.44±0.69	83.71±1.92	77.23±0.48	80.49±3.56	87.02±0.19	82.17±0.27	83.21±0.25	86.07±0.49	83.35±0.93	82.85±0.69
GraphBridge	78.46±0.37	75.13±0.42	75.72±0.59	59.47±0.64	55.13±0.43	56.21±0.47	78.23±0.61	76.04±0.45	74.49±0.92	85.47±0.67	83.01±0.89	81.48±0.76	84.16±0.57	82.10±0.52	83.41±0.73
SeSA	**80.21**±0.26	**78.03**±0.54	**76.89**±0.42	**65.39**±0.27	**60.01**±0.23	**58.82**±0.17	**89.83**±0.27	**86.50**±0.25	**87.82**±0.17	**91.17**±0.21	**90.87**±0.20	**89.23**±0.35	**88.53**±0.28	**86.93**±0.77	**87.56**±0.26

Table 3: Ablation study of our proposed method on Cora and Pubmed. Best in **bold**; second best underlined.

Dataset	**Cora**			**Pubmed**		
Ablation	Acc	BAcc	F1	Acc	BAcc	F1
w/o DLM	72.21	70.47	71.18	78.57	77.72	73.67
w/o Dis-LM	74.78	73.19	72.84	78.49	78.26	74.13
w/o GSA	76.21	74.78	75.11	81.95	76.60	77.98
w/o Global-Edge	77.14	75.82	75.93	86.30	82.01	83.32
w/o Homo-Edge	78.27	76.01	75.80	86.82	83.85	84.76
SeSA	**80.21**	**78.03**	**76.89**	**89.83**	**86.50**	**87.82**

Table 4: Parameter sensitivity and performance across different structural imbalance levels. Best in **bold**; second best underlined.

Imbalance	**Moderate**			**Advanced**			**Extreme**		
Parameters	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1
Vanilla	81.95	76.60	77.98	80.81	74.57	75.66	74.44	64.35	61.33
$\gamma = 0.9, k = 1$	87.32	85.96	84.44	88.47	7.44	86.82	87.82	85.02	85.79
$\gamma = 0.8, k = 3$	**89.83**	**86.50**	**87.82**	89.20	**86.91**	87.60	88.51	85.77	86.48
$\gamma = 0.6, k = 5$	88.94	86.46	86.89	**89.25**	86.87	**87.61**	**88.57**	**86.10**	**86.56**

Table 5: **F1 scores for each class on the Computers dataset.** The corresponding distribution and Structural Information Purity (SIP) are located below each class label. The best results are highlighted in **bold**.

Class	C_3	C_5	C_6	C_7	C_8	C_9
Distribution	5.20%	1.87%	3.12%	1.12%	0.40%	0.67%
SIP	0.728	0.477	0.576	0.768	−0.173	0.219
Vanilla	77.14	69.64	72.21	71.36	60.47	68.95
PC-Softmax	82.68	78.47	80.30	77.13	70.54	72.49
BAT	79.67	78.42	76.37	74.60	72.01	74.37
GraphSHA	81.29	80.11	79.72	81.36	78.64	78.51
SeSA	**86.69**	**85.27**	**85.19**	**84.95**	**84.20**	**84.78**
SIP after GSA	0.730	0.560	0.657	0.7923	0.032	0.346

sity, a hyperparameter sensitivity analysis on γ and k is conducted on the Pubmed dataset. Crucially, to assess adaptability across structural anomalies, three topological imbalance regimes, moderate, advanced, and extreme, are constructed via random edge masking. As summarized in Table 4, the default configuration ($\gamma = 0.8, k = 3$) achieves optimal performance under moderate imbalance, validating our parameter heuristic. Under more adverse conditions, the joint execution of both augmentation components consistently yields improvements. Specifically, under advanced imbalance, the standard setup exhibits comparable efficacy to the maximum intensity setting. Remarkably, under extreme structural imbalance, a relaxed threshold delivers superior results; admitting a larger volume of edges effectively reconstructs the highly fractured topology, facilitating robust message passing for isolated minority nodes. Collectively, these insights substantiate the efficacy and adaptability of our structure augmentation module across diverse imbalanced deployment scenarios.

Focus on Minority Classes. To evaluate our model's capacity to enhance tail-class performance, we introduce Structural Information Purity (SIP) to assess whether local graph topologies facilitate discriminative signal transmission. Focusing on six minority classes in the Computers dataset, Table 5 presents their population proportions, SIP metrics, and F1-scores across methods. Remarkably, SeSA achieves leading performance across nearly all categories, particularly for "Hard-Classes" characterized by extremely low population ratios and minimal SIP. Furthermore, the observed improvement in post-augmentation SIP is strictly synchronous with performance gains. For instance, on class C_8 (0.4% population with severe topological impairment), our framework successfully recovers essential structural information, yielding a substantial lead. This alignment validates that our integrated feature extraction and structural augmentation modules effectively capture minority-class discriminative characteristics while fostering a topological environment conducive to feature propagation within the GNN.

4 Related Work

4.1 LLMs for Text-Attributed Graphs

Recently, synergizing the semantic modeling prowess of LLMs with the structural encoding capabilities of GNNs has emerged as a dominant paradigm for TAG learning. This integration is fundamentally driven by their intrinsic complementarity: LLMs excel at capturing nuanced textual semantics and contextual logic within node attributes, whereas GNNs effectively aggregate high-order topological dependencies via MPNNs. Within the current research (Jin et al., 2024) landscape, LLM as Enhancer stands out as a particularly streamlined and effective paradigm. Pioneer works like GIANT (Chien et al., 2021) leverage language models to mine structural insights through self-supervised neighbor prediction, thereby refining node embeddings. TAPE (He et al., 2024) further broadens the scope of enhancement by utilizing LLMs to generate predictive explanations and ranked features as auxiliary inputs for GNNs, significantly bolstering downstream classifier robustness. Recent advancements of new frameworks (Sun et al., 2025; Huang et al., 2024; Zhu et al., 2025; Wang et al., 2024; Khoshraftar et al., 2025; Liu et al., 2025a) and systematic analyses (Yan et al., 2023; Dai et al., 2024; Guan et al., 2025; Gao et al., 2025; Wang et al., 2025a,b; Wu et al., 2025) continue to push these boundaries, collectively signaling a pivotal shift: moving beyond superficial semantic mapping toward the profound intervention and substantive reconstruction of graph data distributions.

4.2 Node Class Imbalance Problem

In the context of class-imbalanced node classification characterized by BoW-embeddings, early research predominantly focused on mitigating label imbalance. Within this scope, data augmentation has been a cornerstone paradigm. Specifically, GraphSMOTE (Zhao et al., 2021) performs interpolation in the embedding space and utilizes an edge predictor to synthesize connections; and NodeImport (Chen et al., 2025) distills informative supervision signals to generate synthetic nodes that bolster performance within an unbiased context. More recently, research interest has pivoted toward structural imbalance. GraphSANN (Liu et al., 2023) captures long-range structural similarities in heterophilous scenarios; and BAT (Liu et al., 2024) leverages node risk estimation to guide virtual topology augmentation. These methodologies collectively aim to compensate for performance degradation induced by skewed graph structures. Furthermore, several approaches have tackled the problem from higher-level perspectives, such as learning paradigms and training mechanisms. For instance, ReNode (Chen et al., 2021) reweights loss functions based on topological positioning; TAM (Song et al., 2022) expands decision boundaries via topology-aware margin calibration; and GraphMuGu (Zhao et al., 2025) transitions from simple numerical balancing to the balanced learning of supervisory signals. While these approaches have become increasingly systematic in their conceptual design, they still lack a holistic understanding and comprehensive treatment of the multi-dimensional imbalance phenomena in TAGs.

5 Conclusion

This paper deconstructs the multi-dimensional class imbalance problem on TAGs across label, structure, and text dimensions. A quantitative metric is introduced to evaluate structural imbalance, and an integrated framework is proposed to address these multifaceted challenges. Our approach leverages a dual-stage LM feature extraction module to generate high-fidelity embeddings and a semantic-aware augmentation module to optimize graph topology. Empirical evaluations across five datasets demonstrate that the proposed method consistently outperforms state-of-the-art baselines, confirming its robustness across varying imbalance ratios, computational efficiency, and seamless compatibility with diverse GNN backbones.

6 Limitations

While SeSA provides an effective paradigm for multi-dimensional imbalance learning on TAGs, certain limitations offer avenues for future work. Although computational overhead is mitigated via offline LLM feature extraction, the Semantic-aware Graph Structure Augmentation may incur a heavy memory footprint when deployed on extreme-scale industrial graphs. Integrating advanced system-level optimizations, such as approximate nearest neighbor scaling or distributed graph processing. The current framework assumes static TAG topologies and node semantics. Adapting our dual-gating structure augmentation mechanism to accommodate temporal dynamics and strictly inductive learning settings represents a valuable direction for future exploration.

References

Deli Chen, Yankai Lin, Guangxiang Zhao, Xuancheng Ren, Peng Li, Jie Zhou, and Xu Sun. 2021. Topology-imbalance learning for semi-supervised node classification. *Advances in Neural Information Processing Systems*, 34:29885–29897.

Jie Chen, Tengfei Ma, and Cao Xiao. 2018. Fastgcn: fast learning with graph convolutional networks via importance sampling. *arXiv preprint arXiv:1801.10247*.

Nan Chen, Zemin Liu, Bryan Hooi, Bingsheng He, Jun Hu, and Jia Chen. 2025. Nodeimport: Imbalanced node classification with node importance assessment. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pages 94–105.

Zhikai Chen, Haitao Mao, Hongzhi Wen, Haoyu Han, Wei Jin, Haiyang Zhang, Hui Liu, and Jiliang Tang. 2024. Label-free node classification on graphs with large language models (llms). In *International Conference on Learning Representations*, volume 2024, pages 31632–31655.

Eli Chien, Wei-Cheng Chang, Cho-Jui Hsieh, Hsiang-Fu Yu, Jiong Zhang, Olgica Milenkovic, and Inderjit S Dhillon. 2021. Node feature extraction by self-supervised multi-scale neighborhood prediction. *arXiv preprint arXiv:2111.00064*.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. 2019. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277.

Xinnan Dai, Haohao Qu, Yifen Shen, Bohang Zhang, Qihao Wen, Wenqi Fan, Dongsheng Li, Jiliang Tang, and Caihua Shan. 2024. How do large language models understand graph patterns? a benchmark for graph pattern comprehension. *arXiv preprint arXiv:2410.05298*.

Bahare Fatemi, Sami Abu-El-Haija, Anton Tsitsulin, Mehran Kazemi, Dustin Zelle, Neslihan Bulut, Jonathan Halcrow, and Bryan Perozzi. 2023. Ugsl: A unified framework for benchmarking graph structure learning. *arXiv preprint arXiv:2308.10737*.

Matthias Fey and Jan Eric Lenssen. 2019. Fast graph representation learning with pytorch geometric. *ArXiv*, abs/1903.02428.

Hang Gao, Wenxuan Huang, Fengge Wu, Junsuo Zhao, Changwen Zheng, and Huaping Liu. 2025. Llm enhancers for gnns: An analysis from the perspective of causal mechanism identification. *arXiv preprint arXiv:2505.08265*.

Zhong Guan, Likang Wu, Hongke Zhao, Ming He, and Jianpin Fan. 2025. Attention mechanisms perspective: Exploring llm processing of graph-structured data. *arXiv preprint arXiv:2505.02130*.

Jiayan Guo, Lun Du, Hengyu Liu, Mengyu Zhou, Xinyi He, and Shi Han. 2023. Gpt4graph: Can large language models understand graph structured data? an empirical evaluation and benchmarking. *arXiv preprint arXiv:2305.15066*.

Zirui Guo, Lianghao Xia, Yanhua Yu, Yuling Wang, Kangkang Lu, Zhiyong Huang, and Chao Huang. 2024. Graphedit: Large language models for graph structure learning. *arXiv preprint arXiv:2402.15183*.

Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Shen Han, Zhiyao Zhou, Jiawei Chen, Zhezheng Hao, Sheng Zhou, Gang Wang, Yan Feng, Chun Chen, and Can Wang. 2025. Uncertainty-aware graph structure learning. In *Proceedings of the ACM on Web Conference 2025*, pages 4863–4874.

Xiaoxin He, Xavier Bresson, Thomas Laurent, Adam Perold, Yann LeCun, and Bryan Hooi. 2024. Harnessing explanations: LLM-to-LM interpreter for enhanced text-attributed graph representation learning. In *The Twelfth International Conference on Learning Representations*.

Youngkyu Hong, Seungju Han, Kwanghee Choi, Seokjun Seo, Beomsu Kim, and Buru Chang. 2021. Disentangling label distribution for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6626–6636.

Xuanwen Huang, Kaiqiao Han, Yang Yang, Dezheng Bao, Quanjin Tao, Ziwei Chai, and Qi Zhu. 2024. Gnns as adapters for llms on text-attributed graphs. In *The Web Conference 2024*.

Nathalie Japkowicz and Shaju Stephen. 2002. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449.

Bowen Jin, Gang Liu, Chi Han, Meng Jiang, Heng Ji, and Jiawei Han. 2024. Large language models on graphs: A comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering*.

Shima Khoshraftar, Niaz Abedini, and Amir Hajian. 2025. Graphit: Efficient node classification on text-attributed graphs with prompt optimized llms. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 1824–1829.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Wen-Zhi Li, Chang-Dong Wang, Hui Xiong, and Jian-Huang Lai. 2023a. Graphsha: Synthesizing harder samples for class-imbalanced node classification. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1328–1340.

Yuhan Li, Zhixun Li, Peisong Wang, Jia Li, Xiangguo Sun, Hong Cheng, and Jeffrey Xu Yu. 2023b. A survey of graph meets large language model: Progress and future directions. *arXiv preprint arXiv:2311.12399*.

Zhixun Li, Dingshuo Chen, Tong Zhao, Daixin Wang, Hongrui Liu, Zhiqiang Zhang, Jun Zhou, and Jeffrey Xu Yu. 2025. Iceberg: Debiased self-training for class-imbalanced node classification. In *Proceedings of the ACM on Web Conference 2025*, pages 3160–3170.

Jie Liao, Jintang Li, Liang Chen, Bingzhe Wu, Yatao Bian, and Zibin Zheng. 2023. Sailor: Structural augmentation based tail node representation learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 1389–1399.

Jie Liu, Mengting He, Guangtao Wang, Nguyen Quoc Viet Hung, Xuequn Shang, and Hongzhi Yin. 2023. Imbalanced node classification beyond homophilic assumption. *arXiv preprint arXiv:2304.14635*.

Ruyue Liu, Rong Yin, Xiangzhen Bo, Xiaoshuai Hao, Yong Liu, Jinwen Zhong, Can Ma, and Weiping Wang. 2025a. Sstag: Structure-aware self-supervised learning method for text-attributed graphs. *arXiv preprint arXiv:2510.01248*.

Yongxu Liu, Zhi Zhang, Yan Liu, and Yao Zhu. 2022. Gatsmote: Improving imbalanced node classification on graphs via attention and homophily. *Mathematics*, 10(11):1799.

Zemin Liu, Yuan Li, Nan Chen, Qian Wang, Bryan Hooi, and Bingsheng He. 2025b. A survey of imbalanced learning on graphs: Problems, techniques, and future directions. *IEEE Transactions on Knowledge and Data Engineering*.

Zhining Liu, Ruizhong Qiu, Zhichen Zeng, Hyunsik Yoo, David Zhou, Zhe Xu, Yada Zhu, Kommy Weldemariam, Jingrui He, and Hanghang Tong. 2024. Class-imbalanced graph learning without class rebalancing. *Preprint*, arXiv:2308.14181.

Chao Ma, Jing An, Xiang-En Bai, and Han-Qiu Bao. 2022. Attention and cost-sensitive graph neural network for imbalanced node classification. In *2022 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pages 1–6. IEEE.

Mohammadreza Mohammadrezaei, Mohammad Ebrahim Shiri, and Amir Masoud Rahmani. 2018. Identifying fake accounts on social networks based on graph analysis and classification algorithms. *Security and Communication Networks*, 2018(1):5923156.

Joonhyung Park, Jaeyun Song, and Eunho Yang. 2021. Graphens: Neighbor-aware ego network synthesis for class-imbalanced node classification. In *International conference on learning representations*.

Jiawen Qin, Haonan Yuan, Qingyun Sun, Lyujin Xu, Jiaqi Yuan, Pengfeng Huang, Zhaonan Wang, Xingcheng Fu, Hao Peng, Jianxin Li, and Philip S. Yu. 2025. IGL-bench: Establishing the comprehensive benchmark for imbalanced graph learning. In *The Thirteenth International Conference on Learning Representations*.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. 2008. Collective classification in network data. *AI magazine*, 29(3):93–93.

Jaeyun Song, Joonhyung Park, and Eunho Yang. 2022. Tam: topology-aware margin loss for class-imbalanced node classification. In *International Conference on Machine Learning*, pages 20369–20383. PMLR.

Guangxin Su, Yifan Zhu, Wenjie Zhang, Hanchen Wang, and Ying Zhang. 2024. Bridging large language models and graph structure learning models for robust representation learning. *arXiv preprint arXiv:2410.12096*.

Yuanfu Sun, Zhengnan Ma, Yi Fang, Jing Ma, and Qiaoyu Tan. 2025. Graphicl: Unlocking graph learning potential in llms through structured prompt design. *arXiv preprint arXiv:2501.15755*.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. 2018. Deep graph infomax. *arXiv preprint arXiv:1809.10341*.

Haoyu Peter Wang, Shikun Liu, Rongzhe Wei, and Pan Li. 2025a. Generalization principles for inference over text-attributed graphs with large language models. In *Forty-second International Conference on Machine Learning*.

Yaoke Wang, Yun Zhu, Wenqiao Zhang, Yueting Zhuang, Liyunfei Liyunfei, and Siliang Tang. 2024. Bridging local details and global context in text-attributed graphs. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 14830–14841.

Zehong Wang, Sidney Liu, Zheyuan Zhang, Tianyi Ma, Chuxu Zhang, and Yanfang Ye. 2025b. Can llms convert graphs to text-attributed graphs? In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 1412–1432.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le

Scao, Sylvain Gugger, and 3 others. 2020. Huggingface's transformers: State-of-the-art natural language processing. *Preprint*, arXiv:1910.03771.

Lirong Wu, Jun Xia, Zhangyang Gao, Haitao Lin, Cheng Tan, and Stan Z Li. 2022. Graphmixup: Improving class-imbalanced node classification by reinforcement mixup and self-supervised context prediction. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 519–535. Springer.

Xixi Wu, Yifei Shen, Fangzhou Ge, Caihua Shan, Yizhu Jiao, Xiangguo Sun, and Hong Cheng. 2025. When do llms help with node classification? a comprehensive analysis. *arXiv preprint arXiv:2502.00829*.

Hao Yan, Chaozhuo Li, Ruosong Long, Chao Yan, Jianan Zhao, Wenwen Zhuang, Jun Yin, Peiyan Zhang, Weihao Han, Hao Sun, and 1 others. 2023. A comprehensive study on text-attributed graphs: Benchmarking and rethinking. *Advances in Neural Information Processing Systems*, 36:17238–17264.

Sukwon Yun, Kibum Kim, Kanghoon Yoon, and Chanyoung Park. 2022. Lte4g: Long-tail experts for graph neural networks. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 2434–2443.

Jianan Zhao, Meng Qu, Chaozhuo Li, Hao Yan, Qian Liu, Rui Li, Xing Xie, and Jian Tang. 2022. Learning on large-scale text-attributed graphs via variational inference. *arXiv preprint arXiv:2210.14709*.

Jianan Zhao, Le Zhuo, Yikang Shen, Meng Qu, Kai Liu, Michael Bronstein, Zhaocheng Zhu, and Jian Tang. 2023. Graphtext: Graph reasoning in text space. *arXiv preprint arXiv:2310.01089*.

Tianxiang Zhao, Xiang Zhang, and Suhang Wang. 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*, pages 833–841.

Zhongying Zhao, Gen Liu, Qi Meng, Chao Li, and Qingtian Zeng. 2025. Generate or re-weight? a mutual-guidance method for class-imbalanced graphs. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 3680–3688.

Zhiyao Zhou, Sheng Zhou, Bochao Mao, Xuanyi Zhou, Jiawei Chen, Qiaoyu Tan, Daochen Zha, Yan Feng, Chun Chen, and Can Wang. 2023. Opengsl: A comprehensive benchmark for graph structure learning. *Advances in Neural Information Processing Systems*, 36:17904–17928.

Yun Zhu, Haizhou Shi, Xiaotang Wang, Yongchao Liu, Yaoke Wang, Boci Peng, Chuntao Hong, and Siliang Tang. 2025. Graphclip: Enhancing transferability in graph foundation models for text-attributed graphs. In *Proceedings of the ACM on Web Conference 2025*, pages 2183–2197.

Dongcheng Zou, Hao Peng, Xiang Huang, Renyu Yang, Jianxin Li, Jia Wu, Chunyang Liu, and Philip S Yu. 2023. Se-gsl: A general and effective graph structure learning framework through structural entropy optimization. In *Proceedings of the ACM Web Conference 2023*, pages 499–510.

A Extended Discussion regarding Comprehensive Imbalance Ratio

In this section, we provide a comprehensive exploration of the foundational mechanics underlying the Comprehensive Imbalance Ratio (CIR), elucidating how it characterizes inter-class inequality by capturing higher-order topological features. Beyond its theoretical rigor, this metric serves as a robust indicator for quantifying the extent of structural optimization achieved by the model during empirical evaluations, a utility we consistently leverage throughout our experiments.

The conventional Imbalance Ratio (IR) serves merely as a simplistic numerical proxy based on class cardinality, failing to encapsulate the latent structural nuances between categories. Despite extensive research into the impact of structural imbalance on GNNs, formalized metrics to quantify its severity remain remarkably sparse in the existing literature. While ReNode (Chen et al., 2021) is recognized as a pioneer in this domain, and recent advancements such as BAT (Liu et al., 2024) and IGL-Bench (Qin et al., 2025) have introduced valuable indicators, these metrics remain predominantly node-centric. Consequently, they are insufficient for describing the deep-seated structural imbalances inherent in extensive node cohorts or complex subgraphs across different classes.

To provide a more rigorous characterization of the structural disparities between head and tail classes, we introduce the Comprehensive Imbalance Ratio (CIR) as a novel metric for quantifying topological divergence in imbalanced datasets. CIR is designed to evaluate the intensity of structural imbalance from the perspective of Message Passing Neural Networks (MPNNs) across various datasets and partition schemes. Given that the majority of state-of-the-art backbones rely on message-passing mechanisms, our CIR metric is specifically tailored to account for low-degree connectivity and high heterophily, the two phenomena where structural imbalance most severely impairs GNN performance.

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, the topological influence of a node $u \in \mathcal{V}$ is recalibrated by a structural

saliency function $S(u)$:

$$S(u) = \ln(1 + d_u), \tag{13}$$

where d_u denotes the degree of node u. Then for any node $v \in C_i$, its local structural purity $\rho(v)$ is defined as the weighted ratio of effective signals to topological noise within its neighborhood:

$$\rho(v) = \frac{\sum_{u \in N(v)} f_i(u) \cdot S(u)}{\sum_{u \in N(v)} S(u) + \epsilon}, \tag{14}$$

where ϵ is a small constant for numerical stability, and $f_i(u)$ is a class indicator function defined as:

$$f_i(u) = \begin{cases} 1, & \text{if } y_u = y_v \\ -1, & \text{otherwise} \end{cases}. \tag{15}$$

Within the framework of MPNNs, the quality of node representations is intrinsically governed by the Neighborhood Signal-to-Noise Ratio. We model $\rho(v)$ as a proxy for this ratio. Unlike traditional statistics based solely on node counts, $\rho(v)$ captures the topological inductive bias:

- As $\rho(v) \to 1$, the node resides within a highly homophilous central cluster, receiving reinforced semantic signals.
- As $\rho(v) \to -1$, the node is "eroded" by heterophilous neighbors. Its representation becomes severely diluted during propagation, leading to the phenomenon we define as topological marginalization.

The global topological potential of class C_i is defined as the expected Structural Information Purity (SIP) of its constituent nodes, denoted as

$$\Phi_i = \mathbb{E}_{v \in C_i}[\rho(v)]. \tag{16}$$

We interpret Φ_i as the topological cohesion of class C_i. In majority classes, nodes are not only numerically superior but typically occupy the central manifold of the graph, forming a positive potential field. Conversely, minority classes are often relegated to the graph periphery or encircled by majority nodes, resulting in a significantly lower Φ_i. By contrasting $\Phi_{\max}$ with $\Phi_{\min}$, CIR directly exposes the structural asymmetry induced by graph connectivity patterns. Finally, the Comprehensive Imbalance Ratio for the entire graph is formulated as a non-linear mapping of the potential range across all classes:

$$CIR = 1.0 + \eta \cdot \left| \max_i \Phi_i \cdot \Sigma_i - \min_j \Phi_j \cdot \Sigma_j \right|^{\theta}. \tag{17}$$

Σ_i is the intra-class textual semantic homogeneity of class C_i, which is computed by leveraging the LLM to uniformly sample an equal number of instances across different classes, thereby characterising the intra-class textual attribute consistency. The introduction of the power term $(\cdot)^{\theta}$ in the final formulation of CIR is motivated by the observation that structural imbalance does not impair model performance in a linear fashion; rather, it exhibits Cascading Degradation. When the structural disparity between classes exceeds a critical threshold, discriminative signals from minority classes are effectively masked by the topological noise generated by the majority. To ensure robust evaluation across heterogeneous datasets, we implement a Topological Contrast Normalization mechanism via a parameterized combination of η and θ. The selection of these parameters is inspired by the BAT (Liu et al., 2024) framework, which formalizes the exponential and polynomial scaling relationships between intertwined imbalance phenomena. Specifically, since datasets from diverse domains, such as citation networks and co-purchase graphs, exhibit distinct baseline homophily levels, this mechanism maps varied topological distributions into a unified observation space. Consequently, the CIR metric transcends dataset-specific biases, serving as a domain-invariant indicator that characterizes the structural integrity and "health" of the graph data.

For experimental control and manipulation, the Comprehensive Imbalance Ratio serves as an intrinsic metric for quantifying dataset-level structural disparities; unlike the Imbalance Ratio, which can be easily manipulated via node sampling, CIR is not directly controllable. Given our objective to provide a holistic measure of structural imbalance, CIR is influenced by a confluence of factors, including homophily and graph density. Consequently, precisely attaining a target CIR value is challenging; furthermore, identical CIR values across different datasets denote class-level structural variance rather than topological congruence. In our experimental design, even when the IR is fixed at a standard value, CIR values often manifest substantial divergence across datasets. While our primary comparative evaluations focus on benchmarking model performance under standard configurations, where we simply report the empirical CIR, our subsequent ablation studies involve more granular manipulation. Specifically, after fixing the data partition for a given IR, we modulate CIR through stochastic edge masking. To maintain graph connectivity, we

strictly preclude the formation of isolated nodes, which imposes an "intrinsic upper bound" on the attainable CIR for any specific partition. Our results across two sets of experiments demonstrate that our structural augmentation module effectively reduces elevated CIR levels, thereby mitigating structural disparities and driving significant gains in classification performance.

B Experimental Details and Additional Results

Our model is implemented using PyTorch Geometric (Fey and Lenssen, 2019) and the Hugging Face Transformers library (Wolf et al., 2020). Experiments were conducted on Ubuntu 20.04.5 LTS and an NVIDIA GeForce RTX 3090 GPU (24GB). The following three subsections provide a more detailed account of the experimental settings and results. The source code will be made publicly available upon acceptance of the paper.

B.1 Datasets

A brief overview of the datasets is provided, with their fundamental characteristics summarized in Table 6. Additionally, we specify the number of classes designated as Min. Class for each dataset.

Cora: Citation network dataset. Each node represents a paper, and edges represent citation relationships between papers. Cora contains 2,708 nodes and 10,556 edges, with 7 categories, such as Neural Networks, Reinforcement Learning, Genetic Algorithms, Rule Learning, etc.

Citeseer: Citation network dataset. Citeseer contains 3,327 papers and 9104 edges, with 6 categories, such as Artificial Intelligence, Databases, etc.

Pubmed: Citation network dataset. This dataset contains 19,717 papers on diabetes from the Pubmed database, with 88,648 edges. The node classification labels correspond to the type of diabetes discussed in the paper, with 3 categories.

Amazon-Photo: Amazon product purchase dataset. Nodes represent products, and edges connect products that are frequently co-purchased. The Photo dataset contains 48,362 products and 500,939 edges. The node classification labels represent the product category, with 12 categories.

Amazon-Computer: Amazon product purchase dataset. The Computer dataset contains 87,229 products and 721,081 edges. The node classification labels represent the specific domain or method of the product, with 10 categories.

Table 6: Statistics of the datasets.

Name	Nodes	Edges	Class	Min. Class
Cora	2,708	10,556	7	5
Citeseer	3,327	9,104	6	4
PubMed	19,717	88,648	3	2
Photo	48,362	500,939	12	8
Computers	87,229	721,081	10	6

B.2 Baselines

We selected a range of high-performing strategies, from traditional to recent approaches, and carefully chose representative methods at both the **Training Mechanisms** and **Data Augmentation** approaches to serve as baselines for comparison with our model.

Training Mechanism Motheds

Re-weight (Japkowicz and Stephen, 2002): This method adjusts loss weights to favor minority classes, improving their classification performance.

PC-Softmax (Hong et al., 2021): This method adjusts logits during testing to decouple model predictions from the prior probability, achieving effective calibration.

ReNode (Chen et al., 2021): This method addresses both topological and class imbalances by modifying the loss function to enhance classification performance for minority classes.

TAM (Song et al., 2022): TAM leverages local topological information of individual nodes to determine the strength of compensation for class imbalance, effectively mitigating node-level class imbalance.

BAT (Liu et al., 2024) (BAlanced Topological augmentation): BAT is a lightweight, plug-and-play augmentation technique for class-imbalanced node classification. It mitigates the class-imbalance bias introduced by ambivalent and distant message-passing on graph topology with PURE topological manipulation.

Data Augmentation Methods

Upsample : A conventional data augmentation technique that balances the class distribution by randomly duplicating minority class instances until they match the size of the majority class. No specific parameter tuning is required for this baseline.

GraphSMOTE (Zhao et al., 2021): For minority class nodes, GraphSMOTE identifies the node with the highest feature similarity within the same

class and generates new minority class nodes and employs an edge generator to assign neighbors to synthetic nodes based on node feature similarity, integrating them into the original graph.

GraphENS (Park et al., 2021): GraphENS applies the Mixup technique to synthesize the features of nodes of minority class in the original input space. It assigns neighbors to synthetic nodes based on the KL divergence of node features, connecting the generated nodes to the original graph.

GraphSHA (Li et al., 2023a): GraphSHA generates synthetic minority class nodes by focusing on hard samples, thus expanding the decision limits of minority class decisions and improving the generalization of the model.

NodeImport (Chen et al., 2025): NodeImport proposes using a balanced meta-set to select nodes valuable to model training for identifing high-value supervisory signals to synthesize performance, enhancing nodes within unbiased contexts.

Furthermore, to contrast our framework against conventional LLM-enhanced GNNs unequipped to handle imbalance phenomena, five representative LLM-based feature enhancement approaches are introduced as baselines. This comparison serves to verify the effectiveness of our specificity-aware feature capture mechanism. The profiles of these baselines are summarized below.

LLM as Enhancer Methods

GIANT (Chien et al., 2021): GIANT utilizes self-supervised neighborhood prediction to mine structural insights for embedding optimization.

GLEM (Zhao et al., 2022): GLEM facilitates iterative pseudo-label interaction and spatial alignment via an Expectation-Maximization framework.

LLM-GNN (Chen et al., 2024): LLM-GNN proposes a novel label-free framework, which leverages LLM-generated annotations as supervisory signals to guide GNN training.

TAPE (He et al., 2024): TAPE expands the enhancement horizon by leveraging LLMs to generate predictive explanations or ranked features as auxiliary GNN inputs, substantially boosting classifier robustness.

GraphBridge (Wang et al., 2024): GraphBridge addresses the integration of local details and global context in representation learning, leveraging textual information on nodes to enhance GNN comprehension of graph structures, thereby boosting performance in downstream tasks such as node classification.

B.3 Results of Different Backbone Models

We evaluate the influence of various GNN foundations on model performance by employing GCN, GAT, and GraphSAGE as primary MPNN backbones. These architectures integrate node features x_v with local topology into high-dimensional hidden spaces through distinct aggregation mechanisms. Table 7 summarizes the numerical results under an imbalance ratio (IR) of 100 on the Cora and Pubmed datasets and reveals these key empirical observations.

Consistent SOTA Performance: SeSA demonstrates decisive superiority across nearly all metrics. With a GCN backbone on Pubmed, SeSA achieves 89.83% Acc and an 87.82% F1-score, significantly outperforming the next-best competitor, NodeImport. This gap underscores the efficacy of our semantic-aware augmentation and dual-stage feature extraction in mitigating extreme skewness.

Limitation of Heuristic Baselines: While traditional re-weighting and sampling methods show marginal gains, their performance is volatile across different backbones. Conversely, graph-specific methods like GraphSHA are more competitive but still trail SeSA in BAcc and F1-score. This suggests that addressing structural or label disparity in isolation is insufficient for complex TAGs.

Architectural Generalizability: SeSA maintains a dominant lead whether integrated with the neighborhood sampling of GraphSAGE or the attention mechanism of GAT. This validates its robustness and generalizability across diverse message-passing paradigms.

C Complete LLM Input-Output Exemplars and Prompt Template

In this section, we provide a concrete example to illustrate how the Structural-Semantic Distillation stage operates in practice. We developed a task-specific prompt that goes beyond simple summarization. Unlike generic summary prompts, our design explicitly instructs the LLM to prioritize semantic cues that are critical for distinguishing between closely related categories, which is particularly vital for identifying minority class nodes. This "classification-oriented" perspective ensures that the distilled snippets serve as high-quality supervisory signals rather than just shortened text.We demonstrate this process using a representative publication from the PubMed dataset. The raw input like Figure 3 and the resulting distilled snippet like

Raw Text :
"TI": "Mitochondria respiration and susceptibility to ischemia-reperfusion injury in diabetic hearts.",
"AB": "Cardiovascular complications are the primary cause of death for diabetic patients. Clinical and experimental observation has showed the development of dysfunctional cardiomyopathy as one of the main complications of diabetes. Whether the cardiomyopathy results from an increased susceptibility of cardiac tissue to ischemic insult or from a specific functional defect of cardiac mitochondria is a controversial issue. The investigation of possible functional defect in cardiac mitochondria from diabetic rats indicates a decline in state 3 respiration only in animals presenting a marked decrease in body weight. Mitochondria from rats presenting a level of hyperglycemia similar to diabetic animals but not the marked weight loss typical of the latter group show no decline in state 3 respiration, the values being indistinguishable from those of control mitochondria. Mitochondria from hyperglycemic rats, however, show a 15-20% increase in state 4 oxygen consumption but only when glutamate is used as energetic substrate, as compared to a 40-50% increase in state 4 respiration in mitochondria from diabetic rats under similar experimental conditions. This phenomenon is unrelated to diabetes duration, as it is observed at 2 as well as 8 weeks after diabetes onset. Taken together, these data argue against hyperglycemia per se being a direct cause of the decline in state 3 oxygen consumption observed in cardiac mitochondria of type-I diabetic rats and indicate that differences exist in cardiac mitochondrial function in rats generically labeled as diabetic. These differences can contribute to explain discrepancies in experimental results reported by various groups in the field and provide an additional parameter to be taken into consideration in evaluating the varying sensitivity of diabetic hearts to ischemia-reperfusion injury.", (282 words)

Figure 3: An Example Input of LLM

Distilled Text
"This paper most likely addresses Type 1 diabetes. It is hardly possible for it to be in minority class. This paper addresses the impact of cardiovascular complications on diabetic patients and investigates whether diabetic cardiomyopathy stems from an enhanced susceptibility to ischemic insult or from specific functional defects in cardiac mitochondria. The findings reveal that while hyperglycemia does not directly precipitate a decline in cardiac mitochondrial state 3 respiration, it may lead to an elevation in state 4 oxygen consumption under specific conditions. This suggests that hyperglycemia per se is not the direct cause of the diminished state 3 oxygen consumption observed in type-I diabetic hearts. Consequently, the study proposes that distinct types of mitochondrial dysfunction may exist among diabetic individuals." (121 words)

Figure 4: An Example Output of LLM

Figure 4 are compared to show how the LLM filters out background noise while preserving core categorical identifiers. For reference, the complete Instructional Template used for citation networks is detailed in Table 8. This template illustrates our strategy of using domain-specific constraints to guide the LLM toward generating more robust and category-aware representations.

D Extended Related Work

D.1 Node Class Imbalance Problem.

In the context of class-imbalanced node classification where nodes are characterized by BoW embeddings, the majority of existing literature prioritizes addressing quantity-level imbalance in label distributions, leaving the critical issue of structural imbalance comparatively under-explored. We categorize existing methodologies into three streams based on their primary focal points:

Label-Centric Data Augmentation: Early research predominantly focuses on mitigating numerical disparities through various data augmentation paradigms. For instance, GraphSMOTE (Zhao et al., 2021) pioneered the interpolation of node embeddings complemented by an edge predictor for connectivity generation; GATSMOTE (Liu et al., 2022) incorporates neighbor-constrained attention to preserve local structural integrity; GraphMixup (Wu et al., 2022) leverages self-supervised and reinforcement learning to optimize sampling ratios; and NodeImport (Chen et al., 2025) identifies high-value supervisory signals to synthesize performance-enhancing nodes within unbiased contexts.

Structural Imbalance Mitigation: More recent studies have shifted focus toward the structural dimension of imbalance. SAILOR (Liao et al., 2023) refines the feature diffusion process for tail nodes via structural enhancement; GraphSANN (Liu et al., 2023) addresses "long-range structural similarity" in heterogeneous environments; and BAT (Liu et al., 2024) utilizes node risk estimation to guide virtual topological augmentation. Collectively, these methods aim to rectify the performance degradation inherent in skewed graph topologies.

Advanced Learning Paradigms and Training Mechanisms: A third category approaches the problem from a systemic level. ACS-GNN (Ma et al., 2022) introduces a class-aware, neighborhood-weighted attention mechanism; LTE4G (Yun et al., 2022) achieves distributional balance through knowledge distillation; and ReNode (Chen et al., 2021) re-weights the loss function based on the node's topological position. Furthermore, TAM (Song et al., 2022) expands decision boundaries using topology-aware calibration, while GraphMuGu (Zhao et al., 2025) transitions from simple numerical balancing to the balanced

learning of supervisory signals. Additionally, IceBerg (Li et al., 2025) proposes a de-biased self-training mechanism tailored for decoupled GNNs.

D.2 LLMs for Text-Attributed Graphs.

In recent years, synergizing the superior semantic modeling prowess of Large Language Models with the structural relationship modeling capabilities of Graph Neural Networks has emerged as the prevailing trend for Text-Attributed Graph tasks. The underlying rationale for this integration lies in their profound complementarity: GNNs efficiently capture higher-order topological associations via message passing but struggle with deep semantic parsing; conversely, LLMs excel at capturing intricate contextual logic and knowledge but lack a direct mechanism to perceive graph structures. Current research (Jin et al., 2024) generally categorizes the LLM-GNN coupling modalities into three distinct paradigms:

The LLM as Enhancer paradigm presents a more streamlined and efficient approach. Pioneering works like GIANT (Chien et al., 2021) utilize self-supervised neighborhood prediction to mine structural insights for embedding optimization, while TAPE (He et al., 2024) expands the enhancement horizon by leveraging LLMs to generate predictive explanations or ranked features as auxiliary GNN inputs, substantially boosting classifier robustness. Recent advancements and systematic analyses further signify a transition for this paradigm from simple semantic mapping to the deep intervention and reconstruction of graph data distributions.

The LLM as Predictor paradigm seeks to execute downstream tasks, such as node classification or reasoning, directly within a unified generative framework. The primary challenge involves linearizing non-sequential graph structures into LLM-understandable formats. For instance, GraphText (Zhao et al., 2023) employs graph grammar trees to convert structures into sequences for zero-shot inference, while GPT4Graph (Guo et al., 2023) explores standardized graph description languages for structural representation.

The GNN-LLM Alignment paradigm focuses on harmonizing the embedding spaces of both modalities to ensure collaborative encoding. Notable works include GLEM (Zhao et al., 2022) which facilitates iterative pseudo-label interaction and spatial alignment via an Expectation-Maximization framework, and GRAD (Mohammadrezaei et al., 2018), which employs knowledge distillation, treating the GNN as a teacher model, to imbue the LLM with structural aggregation capabilities, thereby decoupling inference from GNN dependency to enhance efficiency.

D.3 Graph Structure Learning.

Graph Structure Learning (GSL) is a burgeoning research field dedicated to uncovering latent patterns and underlying relationships within graph-structured data. While GNNs have demonstrated remarkable efficacy, they rely heavily on explicit graph structures as foundational supervisory signals. However, real-world graphs are frequently plagued by data noise and sparsity, which can compromise the reliability of the original topology. The primary objective of GSL is to reveal latent dependencies that are not directly observable in the raw data. By synthesizing optimized graph structures, GSL facilitates a more holistic understanding of the data, thereby enhancing performance in various downstream tasks, such as node classification.

Representative milestones in the GSL landscape include: UGSL (Fatemi et al., 2023), which systematically summarized early explorations and established the first holistic benchmark for GSL; SE-GSL (Zou et al., 2023), which introduced a generalized framework leveraging structural entropy and hierarchical coding trees; and UnGSL (Han et al., 2025), which proposed an uncertainty-aware GSL approach via a lightweight plug-in module.

Driven by the rapid evolution of LLMs, the fusion of LLMs with GSL has emerged as a powerful new paradigm. For instance, GraphEdit (Guo et al., 2024) explores how LLMs can reason about underlying graph topologies through instruction tuning to achieve structural enhancement. Similarly, Lang-GSL (Su et al., 2024) facilitates reciprocal learning between language models and GSL, seamlessly integrating node feature acquisition with structural optimization.

Furthermore, fundamental theoretical studies have sought to unravel the intrinsic mechanics of GSL and GNNs. OpenGSL (Zhou et al., 2023) scrutinized the pivotal homophily hypothesis and empirically revealed that the homophily level of a learned structure does not necessarily correlate with task performance, thereby challenging conventional wisdom. Nevertheless, research confirms that GSL-derived structures consistently exhibit superior generalization capability and robustness compared to their raw counterparts.

Table 7: Performance comparison on Cora and Pubmed datasets across three backbones. Results show the mean accuracy (%), best in **bold**;second on underlined.

	Datasets	Cora			Pubmed		
	IR=100	Acc	BAcc	F1	Acc	BAcc	F1
GCN	Vanilla	74.80	72.11	68.92	63.79	62.87	60.91
	Re-weight	73.29	65.13	65.72	70.23	69.00	69.49
	PC-Softmax	68.54	62.96	66.55	70.84	64.61	58.92
	ReNode	79.02	73.53	74.71	75.64	72.56	72.22
	TAM	78.34	76.11	73.34	77.84	71.96	73.08
	BAT	76.23	72.00	72.01	75.52	72.58	73.16
	Upsample	76.48	70.76	71.24	68.91	65.28	66.04
	GraphSMOTE	73.72	74.18	66.74	70.10	68.25	68.54
	GraphENS	75.90	74.73	68.06	70.62	69.66	69.34
	GraphSHA	79.28	77.12	75.10	75.47	74.16	74.19
	NodeImport	80.03	77.80	75.01	84.57	83.18	83.16
	Ours (SeSA)	**80.21**	**78.03**	**76.89**	**89.83**	**86.50**	**87.82**
GAT	Vanilla	69.73	60.60	60.83	69.99	64.85	65.83
	Re-weight	72.86	65.45	66.21	70.04	66.60	67.00
	PC-Softmax	71.46	63.80	64.59	70.51	66.05	67.14
	ReNode	74.29	69.86	69.70	72.79	70.12	70.58
	TAM	73.15	69.34	69.29	75.44	74.36	74.62
	BAT	76.23	72.00	72.01	75.23	72.81	73.62
	Upsample	70.68	61.75	61.69	73.07	70.26	71.34
	GraphSMOTE	73.04	65.99	66.59	73.70	68.11	65.32
	GraphENS	75.44	68.82	69.05	81.32	76.30	73.82
	GraphSHA	73.37	68.74	67.87	80.24	77.37	74.24
	NodeImport	79.51	76.43	74.71	84.57	83.18	83.16
	Ours (SeSA)	**80.02**	**77.69**	**76.14**	**88.96**	**85.92**	**86.44**
GraphSAGE	Vanilla	69.48	59.77	60.61	64.40	57.02	54.54
	Re-weight	70.52	61.22	62.30	65.67	58.33	56.07
	PC-Softmax	70.47	61.19	62.26	66.45	59.09	56.69
	ReNode	75.51	69.59	70.12	71.13	66.13	66.95
	TAM	76.30	70.85	71.34	79.52	77.97	78.41
	BAT	73.05	69.87	71.43	75.52	72.58	73.16
	Upsample	71.10	62.26	63.38	65.56	56.97	51.82
	GraphSMOTE	72.35	64.00	64.53	65.96	56.78	51.97
	GraphENS	73.76	68.66	68.51	75.25	73.46	74.05
	GraphSHA	73.68	68.55	68.44	74.34	72.33	72.42
	NodeImport	**79.63**	77.46	74.39	85.46	**84.93**	84.72
	Ours (SeSA)	79.51	**77.53**	**74.67**	**87.78**	84.51	**85.42**

Table 8: Prompt template for the text simplification module.

Role:	You are an LLM-based text simplification module for node classification on text-attributed graphs. Your goal is to extract class-distinctive and label-relevant content. The class label of this dataset include XX, XX, etc.
Context:	Simplified text addresses Class Imbalance. Prioritize minority node features. Nodes from minority classes usually have lower degree and centrality, and most of their neighbors belong to different classes. You will be provided with this information to assist your task, like target Node Info: Degree=`{degree_str}`, Homophily=`{homophily_str}`, Centrality=`{centrality_str}`
Requirements:	• Output ONLY simplified text. Use Definitely, Probably, Maybe, Possibly, and Hardly to show minority likelihood • First sentence: 'This paper is most likely in the field of XX. It is XX possible for it to be in minority class. '. • Total length: 5–8 sentences.
Input:	Original Text: `{text}`
Output:	Simplified Output: